

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

SONOS, INC.,
Petitioner,

v.

GOOGLE LLC,
Patent Owner.

Case No. TBD
U.S. Patent No. 9,632,748

**SONOS, INC.'S PETITION FOR *INTER PARTES*
REVIEW OF U.S. PATENT NO. 9,632,748**

TABLE OF CONTENTS

LISTING OF EXHIBITS	ii
I. Introduction	1
II. Mandatory Notices	1
A. Real Party-in-Interest (37 C.F.R. §42.8(b)(1))	1
B. Related Matters (37 C.F.R. §42.8(b)(2))	1
C. Counsel Information (37 C.F.R. §42.8(b)(3))	2
D. Service Information (37 C.F.R. §42.8(b)(4)) and Payment of Fees	3
III. Requirements for IPR	3
A. Grounds for Standing	3
B. Identification of Challenge	3
IV. Discretionary Factors Favor Institution	4
A. Denial Under §314(a) Would Be Inappropriate	4
B. Denial Under §325(d) Would Be Inappropriate	5
V. '748 Patent Overview	5
VI. Level of Ordinary Skill in the Art	8
VII. Claim Construction	8
A. 1330 ITC Investigation Construed '748 Terms	8
B. “A Spoken Audio Input... the Spoken Audio Input”	10
C. “A Command to Temporarily Refrain from Performing Speech Recognition”	11
VIII. Priority Date	14
IX. Prior Art Overview	15
A. Rosenberger	15
B. Jang	21
X. Ground 1—Rosenberger (§102)	24
A. Rosenberger Anticipates Independent Claims 1, 7, 11	24
1. Independent Claim 7	24

a.	<i>[7.0] [a] first computing device comprising: [7.0(a)] at least one processor; and [7.0(b)] at least one module operable by the at least one processor to:</i>	24
b.	<i>[7.1] receive a spoken audio input that is also received by one or more additional computing devices in physical proximity to the first computing device</i>	28
c.	<i>[7.2] determine, 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; and.....</i>	31
d.	<i>[7.3] responsive to determining that the first computing device should perform speech recognition on the spoken audio input:.....</i>	34
e.	<i>[7.3(a)] send, to at least a second computing device from a plurality of computing devices, a command to temporarily refrain from performing speech recognition</i>	36
f.	<i>[7.3(b)] perform speech recognition on the spoken audio input.....</i>	42
g.	<i>[7.3(c)] determine, based on performing speech recognition on the spoken audio input, whether the spoken audio input includes a predetermined spoken audio command; and.....</i>	43
h.	<i>[7.3(d)] responsive to determining that the spoken audio input includes the predetermined spoken audio command, execute an action associated with the predetermined spoken audio command.</i>	44
2.	Independent Claim 1	47
3.	Independent Claim 11	48
B.	Rosenberger Anticipates Dependent Claims 2-4, 9-10, 12, 14-15	49
1.	Dependent Claims 2, 9, 12.....	49

2.	Dependent Claims 3, 10	56
3.	Dependent Claims 4, 14, 15	57
XI.	Ground 2—Rosenberger + Jang (§103).....	59
A.	Rosenberger+Jang Render Obvious the Additional Elements of Claims 2/9/12 & 3/10	59
B.	Motivation to Combine.....	67
XII.	Lack of Secondary Considerations.....	71
XIII.	Conclusion	71

TABLE OF AUTHORITIES

Cases

<i>Darfon Elecs. Corp. v. Lite-on Singapore Pte. Ltd.</i> , IPR2018-01608, 2019 WL 994656 (PTAB Feb. 28, 2019).....	5
<i>InterDigital Commc'ns, LLC v. Int'l Trade Comm'n</i> , 690 F.3d 1318 (Fed. Cir. 2012)	50

LISTING OF EXHIBITS

Exhibit	Description
1001	U.S. Patent No. 9,632,748
1002	Declaration of Dr. Michael T. Johnson
1003	U.S. Patent No. 8,340,975 (“Rosenberger”)
1004	U.S. Patent Publication 2013/0073293 (“Jang”)
1005	<i>Certain Audio Players and Components Thereof</i> , Inv. No. 337-TA-1330 (“1330 Investigation”), Order No. 14
1006	1330 Investigation, Order No. 16
1007	Redacted excerpt of Google’s Suppl. Infringement Claim Chart for U.S. Patent No. 9,632,748 (Jan. 31, 2023) in the 1330 Investigation
1008	Redacted excerpt of Dr. Mark Mahon’s Evidentiary Hearing Testimony in the 1330 Investigation
1009	Redacted excerpt of Google’s 6 th Suppl. Obj. & Resp. to Sonos’s 1 st Set of Interrogatories (Nos. 1-44) in the 1330 Investigation
1010	Public version of Google’s Initial <i>Markman</i> Brief in the 1330 Investigation (Dec. 12, 2022)
1011	U.S. Patent Publication 2010/0088100 (“Lindahl”)
1012	Excerpts from <i>Techopedia.com</i> , “Microcontroller” (Sept. 15, 2014)
1013	Excerpts from <i>Merriam-Webster Dictionary</i> (Jan. 28, 2012)
1014	Excerpts from <i>Dictionary of Computer and Internet Terms</i> (10th ed. 2009)
1015	Excerpts from <i>Microsoft Computer Dictionary</i> (5th ed. 2002)
1016	NLP-5x Natural Language Processor Data Sheet (2009)

I. Introduction

Petitioner Sonos, Inc. (“Sonos”) respectfully requests institution of *Inter Partes* Review (“IPR”) of claims 1-4, 7, 9-12, and 14-15 (“Challenged Claims”) of U.S. Patent 9,632,748 (the “’748 Patent”; Ex.1001). The ’748 Patent issued on April 25, 2017 and is currently assigned to Google LLC (“Google” or “Patent Owner”).

This Petition demonstrates at least a reasonable likelihood that Sonos will prevail with respect to at least one Challenged Claim. 35 U.S.C. §314(a). Sonos asserts that the Challenged Claims are anticipated and/or rendered obvious by the asserted prior art.

Pursuant to 37 CFR §42.22, Petitioner respectfully requests the Board to review the asserted prior art and below analysis, institute a trial for IPR of the Challenged Claims, and cancel those claims as unpatentable.

II. Mandatory Notices

A. Real Party-in-Interest (37 C.F.R. §42.8(b)(1))

Sonos is the real party-in-interest.

B. Related Matters (37 C.F.R. §42.8(b)(2))

Google asserted, *inter alia*, the ’748 Patent against Sonos in the following ITC and district court proceedings:

- *Certain Audio Players and Components Thereof*, Inv. No. 337-TA-1330 (“1330 Investigation”);
- *Google LLC v. Sonos, Inc.*, Case No. 3:22-CV-04553 (N.D. Cal.)

(“NDCA-Action”).

Sonos was served with the NDCA-Action complaint on August 9, 2022.

Thus, this Petition is timely under 35 U.S.C. § 315(b).

C. Counsel Information (37 C.F.R. §42.8(b)(3))

Sonos designates the following counsel:

Lead Counsel:

Michael P. Boyea, Reg. No. 70,248
LEE SULLIVAN SHEA & SMITH LLP
656 West Randolph Street, Suite 5W
Chicago, IL 60661
(312) 754-9602
boyea@ls3ip.com

Back-up Counsel:

Jae Y. Pak, Reg. No. 72,428
LEE SULLIVAN SHEA & SMITH LLP
656 West Randolph Street, Suite 5W
Chicago, IL 60661
(312) 754-9602
pak@ls3ip.com

Cole B. Richter, Reg. No. 65,398
LEE SULLIVAN SHEA & SMITH LLP
656 West Randolph Street, Suite 5W
Chicago, IL 60661
(312) 754-9602
richter@ls3ip.com

Pursuant to 37 C.F.R. § 42.10(b), a Power of Attorney accompanies this Petition. The above identified lead and backup counsel are registered practitioners associated with Customer No. 130085 listed in that Power of Attorney.

D. Service Information (37 C.F.R. §42.8(b)(4)) and Payment of Fees

Sonos consents to service by electronic mail at boyea@ls3ip.com, pak@ls3ip.com, richter@ls3ip.com, and LS3_team@ls3ip.com.

The fee set forth in 37 C.F.R. § 42.15(a) for this Petition has been paid. Any additional fees due in connection with this Petition may be charged to Deposit Account 506632.

III. Requirements for IPR

A. Grounds for Standing

Sonos certifies that the '748 Patent is available for IPR and that Sonos is not barred or estopped from requesting IPR. Sonos has filed this Petition within one year of service of the NDCA-Action complaint on August 9, 2022. Sonos has not filed a civil action challenging the validity of any claim of the '748 Patent.

B. Identification of Challenge

Sonos requests IPR of Challenged Claims 1-4, 7, 9-12, and 14-15. The application underlying the '748 Patent was filed July 29, 2014 and claims priority to a June 24, 2014 provisional application filing. Ex.1001. For purposes of this Petition only, Sonos assumes that the '748 Patent is entitled to the benefit of this June 2014 date. The prior art relied upon in this IPR predates the earliest priority date Google can claim. Thus, it need not be determined whether the Challenged Claims are actually entitled to that priority date.

Review is requested in view of U.S. Patent No. 8,340,975

(“Rosenberger”;Ex.1003) and U.S. Publication No. 2013/0073293 (“Jang”;Ex.1004). These references qualify as prior art under 35 U.S.C. §§102(a)(1)-(2)(AIA). Sonos requests institution on the Challenged Claims based on the following grounds:

Ground 1	Claims 1-4, 7, 9-12, and 14-15 anticipated by Rosenberger
Ground 2	Claims 2-3, 9-10, 12 rendered obvious by Rosenberger and Jang

These grounds are supported by a Declaration from Dr. Michael T. Johnson. Ex.1002.

IV. Discretionary Factors Favor Institution

A. Denial Under §314(a) Would Be Inappropriate

Consistent with the goals of *Fintiv* and Director Vidal’s June 21, 2022 Interim Procedure (“I.P.”), Sonos requests that the Board grant this Petition and institute IPR. The I.P. states “the PTAB will not discretionarily deny petitions based on applying *Fintiv* to a parallel ITC proceeding.” I.P., 7. Thus, the parallel 1330 Investigation does not implicate *Fintiv*.

Moreover, the *Fintiv* factors in view of the parallel NDCA-Action weigh strongly in favor of institution. Specifically, *Fintiv* factors 1-5 favor institution because the court stayed the NDCA-Action pending the outcome of the 1330 Investigation (Target Date is January 16, 2024) and any appeals. NDCA-Action, Dkt.22. There has been minimal investment in the NDCA-Action by the court and parties, and the Board is likely to address the ’748 Patent via this Petition before the

district court. Thus, there are no concerns involving conflicting decisions or duplicative work by the Board and district court. Additionally, the compelling strength of the grounds presented herein (*Fintiv* factor 6) further favors institution.

B. Denial Under §325(d) Would Be Inappropriate

The Board should not exercise its discretion to deny institution under §325(d) because this Petition relies on prior art and arguments that were not previously presented to the USPTO and this Petition's prior art and arguments are materially different and not cumulative of what the USPTO did consider. Indeed, unlike the Rosenberger and Jang references discussed in this Petition, the primary reference and lead secondary reference considered by the USPTO during prosecution (US2003/0167335 and US2011/0257971, respectively) are silent on multiple voice-enabled devices performing an arbitration process to select a device to interact with a user. Thus, Google cannot credibly contend that *Advanced Bionics* part one is even satisfied. See *Darfon Elecs. Corp. v. Lite-on Singapore Pte. Ltd.*, IPR2018-01608, 2019 WL 994656, *11 (PTAB Feb. 28, 2019) (declining to exercise discretion under §325(d) because, apart from one reference, "none of the asserted prior art references... was considered previously by the Examiner during prosecution.").

V. '748 Patent Overview

The '748 Patent states "techniques of the present disclosure may enable one or more computing devices from among a plurality of computing devices to

designate a particular computing device to monitor for and/or act on audio input usable to initiate interaction with at least one of the plurality of computing devices.” Ex.1001, 2:36-41.

Much like the prior art, the ’748 Patent explains that “a plurality of computing devices... may each be operable to monitor for and/or act on audio input that includes one or more predetermined audio commands.” *Id.*, 2:42-45. The ’748 Patent further explains that “[o]ne or more of the computing devices may... designate a particular computing device... from among the plurality of computing devices as the best candidate for monitoring environmental audio for audio input that includes the one or more predetermined audio commands,” and “[r]esponsive to designating a particular computing device to monitor for the one or more predetermined audio inputs, other computing devices from the plurality of computing devices... may temporarily cease to monitor for audio input.” *Id.*, 2:45:57. The ’748 Patent generally refers to the process for making such designation as “arbitration.” *Id.*, 7:23-36, 22:51-56.

The ’748 Patent explains that the “best candidate” is determined during an initial phase based on various factors before a subsequent phase where only the “best candidate” receives a voice input and performs speech processing on that voice input to determine if a predetermined command was uttered. *See, e.g., id.*, FIGs. 4-5. In this initial phase, the ’748 Patent explains that the “best candidate” is determined

based on the following:

- “a ranked list of device types usable to monitor for and/or process audio input,” such as a “smartphone, smartwatch, headset, tablet, laptop, desktop, in-vehicle system, etc.” (*id.*, 7:37-64);
- “a level of audio input quality associated with one or more of [the] computing devices,” such as “volume, signal-to-noise ratio, and/or other parameters” (*id.*, 7:65-8:13);
- “the number of predetermined audio commands that each of [the] computing devices [] may be operable to act on” (*id.*, 8:14-25); or
- “a determined context of one or more of [the] computing devices” (*id.*, 8:26-62).

The '748 Patent also explains “the particular device may be the only one of the plurality of computing devices monitoring for audio input and, in some instances, upon receiving an indication of audio input that includes a predetermined audio command, the particular device may be the only one of the plurality of computing devices to perform one or more actions responsive thereto.” *Id.*, 2:57-63.

In regard to speech processing, the '748 Patent provides very little detail. For instance, the '748 Patent provides little detail on “speech recognition.” Indeed, the '748 Patent only generally describes two speech recognition techniques: (i) “performing one or more speech-to-text transformations to obtain textual data corresponding to the received audio input” (*id.*, 15:30-34; *see also id.*, 4:57-63) and

(ii) “perform[ing] other audio analysis to recognize and identify speech, such as performing frequency analysis to identify tonal characteristics or other sonic identifiers present in the audio input” (*id.*, 15:34-38). In this way, a POSITA would understand that the ’748 Patent is not focused on any specific speech processing techniques, and does not promote one type of speech recognition technique to the exclusion of other types of techniques. Ex.1002, ¶35.

VI. Level of Ordinary Skill in the Art

A POSITA in the technology field of the ’748 Patent as of the assumed priority date (i.e., June 24, 2014) 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 speech processing, or an equivalent level of skill, knowledge, and experience. *See* Ex.1002, ¶36-39.

VII. Claim Construction

A. 1330 ITC Investigation Construed ’748 Terms

The ALJ in the 1330 Investigation issued a *Markman* Order construing various terms of the ’748 Patent, including the following:

Claim Term(s)	ALJ's Claim Construction
<i>predetermined spoken audio command</i> ¹	Plain and ordinary meaning, which is not “any audio input usable to initiate interaction with a computing device, such as a spoken hotword”
<i>perform/performing speech recognition</i>	Plain and ordinary meaning: “the process of transforming spoken audio into text (i.e., to transcribe) or some other form readable by the claimed computing device” <i>*ALJ asked for further evidentiary development</i>
<i>a command to temporarily refrain from performing speech recognition</i>	“a command to temporarily refrain from performing any and all speech recognition”

Ex.1005, pp.27-30, 34.

Sonos subsequently filed a motion for reconsideration/clarification with regard to *perform/performing speech recognition*, and the ALJ stated that the “[r]esolution of this dispute requires further evidentiary development regarding the term’s plain and ordinary meaning.” Ex.1006, p.1. The evidentiary record in the 1330 Investigation is now complete, and the parties are awaiting the ALJ’s Initial Determination, along with his conclusion regarding the term’s plain and ordinary meaning, which is expected September 15, 2023.

Sonos applied the ALJ’s constructions above for purposes of this Petition, subject to Google’s broad interpretation of *a command to temporarily refrain from*

¹ Claim language is italicized throughout this Petition for clarity. All other emphasis herein has been added unless noted otherwise.

performing speech recognition discussed below. *Infra* VII.C.

B. “A Spoken Audio Input... the Spoken Audio Input”

In the 1330 Investigation, Google interpreted the phrase *a spoken audio input... the spoken audio input* broadly such that (i) the initial recitation of *a* *spoken audio input* would be satisfied by a first user utterance containing at least a hotword/wakeword (e.g., “Hey Sonos”) and perhaps an initial user request/command (e.g., “Play Some Music”) and (ii) subsequent recitations of *the* *spoken audio input* would be satisfied by a second user utterance containing a subsequent or “follow-up” request that does not contain the hotword/wakeword (e.g., “Louder”). *See, e.g.*, Ex.1007, p.5 (“The spoken audio input may consist of a wakeword (i.e., a hotword) and one or more spoken commands, including follow-up requests from the user after the device has responded to the user’s initial request.”), p.13:

Sonos Voice Control also supports use cases in which *the spoken audio input includes the “Hey Sonos” wakeword and an initial command, but also one or more follow-up requests*. For example, if the user utters, “Hey Sonos, play some music,” and then, without repeating the “Hey Sonos” wakeword, the user says “louder,” the Accused Products will increase the volume notwithstanding that the user did not repeat the wakeword.

Id., p.55:

After the ’748 Accused Product responds to the user’s query, the ’748 Accused Product serving as the first computing device will then

monitor for *additional follow-on commands in the spoken audio input* without requiring a new hotword utterance (e.g., in response to the device turning the volume louder, the user may utter “louder” without repeating the hotword) and continue to perform speech recognition on the “continued conversation” follow-on speech.

See also, e.g., Ex.1008, pp.32-34.

While Sonos disagrees with Google’s broad interpretation of *a spoken audio input... the spoken audio input*, Sonos assumes it is correct for purposes of this Petition.

C. “A Command to Temporarily Refrain from Performing Speech Recognition”

As discussed, the ALJ in the 1330 Investigation construed *a command to temporarily refrain from performing speech recognition* as “a command to temporarily refrain from performing *any and all* speech recognition.” Despite this construction, Google interpreted this term broadly in several respects in the 1330 Investigation.

For example, Google contended that when evaluating whether a message amounts to *a command to temporarily refrain from performing speech recognition* it is *not relevant* to consider what the recipient of the message does in response to the message. Ex.1009, p.12 (“To determine whether a first computing device practices [the ‘command to temporarily refrain’ and thus,] the claim, the *behavior of other* computing devices of the plurality of computing devices *need not be*

assessed.”), p.13 (“Whether the first computing device sends a ‘command to temporarily refrain from performing speech recognition’ should be determined by the command itself, ***not by the effect it has on some other device.***”).

As another example, with regard to *temporarily refrain*, Google initially contended this only refers to not “perform[ing] speech recognition for a limited amount of time.” Ex.1010, p.22. But, Google’s interpretation evolved to cover situations where *temporarily* is essentially read out of the claim such that (i) refraining for practically ***zero*** seconds, (ii) refraining for only the remainder of the spoken audio input (whatever that is, such as ***nothing***), and (iii) refraining from performing speech recognition on the ***current*** spoken audio input but performing speech recognition on the ***next*** spoken audio input would nevertheless satisfy the claim term. This is evident from the testimony of Google’s own technical expert, Dr. Mahon, during the 1330 Investigation evidentiary hearing:

- “Q. Right. So under your view, a message that results in a device refraining from performing speech recognition, for a very, very short time that is practically zero, would still be a command to temporarily refrain from performing speech recognition; correct?

A. ***There is no time limit*** in terms of temporarily refraining.”
Ex.1008, p.45.

- “Q. So you were asked at your deposition, Dr. Mahon, ‘And so, in the example, ***temporarily is zero seconds; correct?***’ (As read.) You said, ‘Again, it depends on the input stream, what has been

identified, what else is in the room or what other words might be in context, *but certainly, it's not without possibility that time is very, very short*, which could be *practically zero*.'

A. Yes, I gave that answer....” *Id.*, p.46.

- “Q. And during that deposition, you were asked the question, ‘What is your interpretation of ‘temporarily’ in the context of the '748 Patent?’ And your response was, ‘That the device will refrain until the winning device -- it would *refrain* from performing speech recognition *on the remainder of the spoken audio input, whatever that is*.’ That’s the answer you gave; correct?

A. Yes” *Id.*, pp.43-44; *see also id.*, p.45.

- “Q. Okay. Now I’d look to look at the question that begins at line 25; in other words, the very next question and answer. The question, ‘And what happens if a device has *already performed speech recognition on the entirety of the spoken audio input*?’

Your answer, ‘The devices will *refrain until the instance of that first spoken audio input is complete*, and then they would be -- well, they -- they just -- they’re not performing speech recognition, I should say, on that spoken audio input. That’s the result in the case that you just indicated. Meaning, the case you proposed was that if the user gave a short command that didn’t require a follow-on request, that *the refraining would basically be for that duration of that process*, so meaning there’s no follow-on request. It will -- but *it will not perform speech recognition on anything that remains*, which in this case would be *nothing*.’” *Id.*, p.55.

- “A. Well, because in that – it’s the exact, same scenario. The exact, same things are going on. If there’s *no follow-on request*, the other devices are *still refraining from performing speech recognition*. They don’t know that there’s not going to be another follow-on request, and they’re waiting for another hotword. They go back to hotword detection. So *they’re refraining by going back to hotword detection*.” *Id.*, p.56.
- “Q. Assume that a device is performing hotword detection without interruption, and assume that hotword detection is speech recognition. Are you with me so far?

A. Yes.

Q. So in that scenario, you would agree with me that that device that is continuing to perform hotword detection is not refraining from performing speech recognition; correct?

A. *No*, I’d still disagree because *you’re* going back and *performing speech recognition for the next spoken audio input*. So *it’s not performing speech recognition on the current spoken audio input*.” *Id.*, p.41.

While Sonos disagrees with this broad interpretation of *a command to temporarily refrain from performing speech recognition*, Sonos assumes it is correct for purposes of this Petition.

VIII. Priority Date

The ’748 Patent issued on April 25, 2017 from an application filed on July 29, 2014, which claims priority to a provisional application filed on June 24, 2014. For

purposes of this Petition only, Sonos assumes that the '748 Patent is entitled to the benefit of this June 2014 date.

IX. Prior Art Overview

A. Rosenberg

Rosenberger published on December 25, 2012 (Ex.1003) and thus, qualifies as prior art to the '748 Patent, which has an earliest possible priority date of June 24, 2014.

Rosenberger generally “relates to improvements in speech-recognizing control devices that respond to oral commands from a system user to produce a desired effect or result.” Ex.1003, 1:8-13. In this regard, Rosenberger explains, “to provide complete hands-free speech recognition coverage and comfortable user interaction from anywhere in a building using speech recognition,” it is often necessary to arrange many “speech recognition modules” “in reasonable proximity to one another, sometimes several per room in larger rooms.” *Id.*, 1:53-2:1, 2:41-47.

In such an arrangement, Rosenberger describes what is referred to as the “duplicate response” problem where multiple “speech recognition modules” simultaneously recognize the same “speech command” and both “audibly respond” to and/or “process” the speech command causing “user confusion” and/or “undesirable or adverse effects.” *Id.*, 2:42-59; *see also id.*, 9:18-10:4.

While Rosenberger acknowledges several approaches to address this “duplicate response” problem, Rosenberger finds these approaches to be deficient (*id.*, 2:59-3:26) and provides a system where “[s]everal devices may be installed in close proximity to one another to ensure hands-free coverage throughout [a] space” and “[w]hen two or more devices are triggered simultaneously by the same speech command, real time coordination ensures that only one device will respond to the command.” *Id.*, Abstract.

Specifically, Rosenberger discloses provisioning a “home or commercial building having controllable automation systems, appliances, electronic equipment or machines (i.e. ‘automated mechanisms’)” with a plurality of “speech recognition control devices” (or simply, “control devices”)—a phrase Rosenberger uses synonymously with the phrase “speech recognition module” that was used in describing the “duplicate response” problem. *Id.*, 6:14-21. Each “control device” (also referred to as “device D”) includes “embedded speech recognition software and circuitry capable of quickly recognizing spoken speech commands from an expected vocabulary or ‘flexible grammar’.” *Id.*, 6:21-25.

Addressing this problem through “coordination” (or “arbitration” in the ’748 Patent’s parlance) of multiple “control devices” is a central aspect of Rosenberger:

Another object of this invention is to provide a hands-free speech-recognition system that automatically determines which one of a plurality of interactive speech recognition control devices that have

been simultaneously triggered by a speech command will further interact with the user as needed and control the system response.

Id., 3:57-62.

Rosenberger explains that provisioning a space with a plurality of “control devices” “add[s] useful ‘hands-free’ speech control capabilities to” “existing controllers of home automation systems, building control systems, and various machines,” such that “thermostats, light switches, motorized shading systems, audio-visual equipment, security systems, kitchen equipment, pumps, motors and actuators” (among other examples of “automated mechanisms” (AMs)) can be controlled directly by a “control device” via voice. *Id.*, 6:40-61.

Rosenberger illustrates an example of such a provisioned space (reproduced below) that includes (i) several “table-top control devices” (T1-T6), (ii) several “wall-box mounted control devices” (W1-W9), and (iii) numerous “[a]utomated mechanisms such as Door lock DL1, security panel A1, television TV1, thermostat Tstat1, light dimmers L1, L2, L3, L4 and L5 and motorized shades S1, S2 and S3” *Id.*, 9:38-49, FIG. 3.

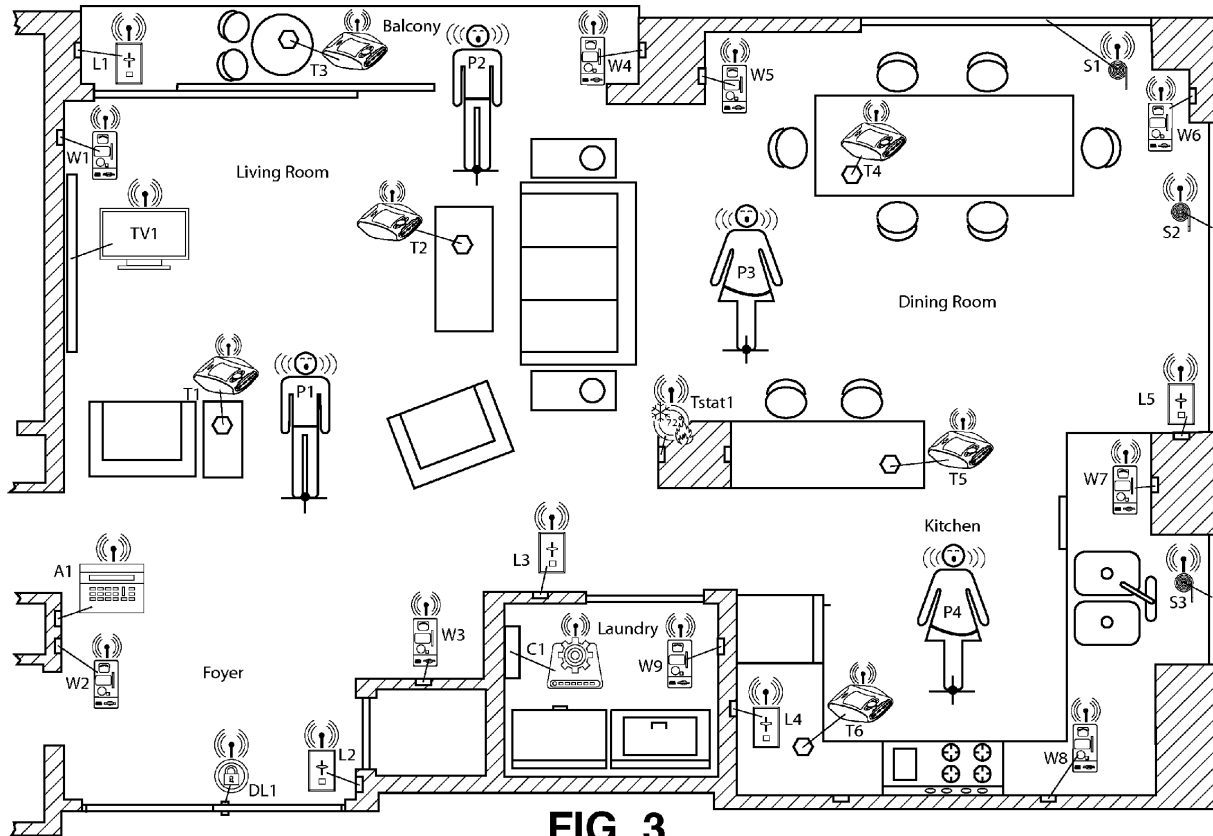


FIG. 3

With reference to FIG. 3, Rosenberger explains that the “automated mechanisms” “may all be controlled ‘hands-free’ by persons P1, P2, P3 and P4 using only speech commands” and that applicable “[c]ontrol signals may be sent directly from any speech recognition device to system controller C1, which in turn, sends appropriate control signals to the target automated mechanisms to implement the speech command” or “any speech recognition device can control the automated mechanisms directly using standard wireless communications and automation protocols such as Zigbee.” *Id.*, 9:45-56.

Rosenberger in turn discloses that “the control device of the invention uses one of several different coordination methods to achieve the design goal of avoiding duplicate responses to the same speech command while selecting a single control device best able to successfully interact with the system user,” and “[a]ll of the 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.” *Id.*, 10:34-41.

Rosenberger describes four methods that could be employed, “[o]nce each device in a space the recognizes a speech command has calculated its weighted signal,” for the “control devices” “to be coordinated so that only the single device with the highest weighted signal proceeds to interact with the user and/or process the speech command.” *Id.*, 11:27-14:9. These four methods are categorized as either involving a “Coordinating Controller” or not involving a “Coordinating Controller.”

In the non-Coordinating-Controller methods, “each control device is programmed to carry out the decision-making” (i.e., make the arbitration determination), as opposed to one “coordinating controller” carrying out the decision-making on behalf of other “control devices.” *Id.*, 11:37-40. In these methods, each “control device” exchanges WS values with one or more other “control devices,” and each given “control device” then compares its computed WS value to any received WS values to determine whether it has the highest WS value

and therefore, should handle the voice input. *Id.*, 11:40-12:27 (describing “Broadcast and Monitor” method), 13:30-14:9 (describing “Direct Neighbor Pairing” method).

In the “Coordinating Controller” methods, “a coordinating controller provides coordination services for all interactive speech recognition control devices in a given space.” *Id.*, 12:35-44. Notably, the “coordinating controller” itself could be implemented on an “interactive speech recognition control device” that also hears a given user utterance and computes a corresponding WS value. *Id.*, 12:42-44, 4:60-5:6; 11:27-29. In the “Coordinating Controller” methods, each “control device” that recognizes a speech command sends its calculated WS value to the “coordinating controller” that in turn decides which “control device” should handle the voice input. *Id.*, 12:28-56.

The “Coordinating Controller” methods can follow what is known as an “optimistic model.” In the “optimistic model,” after making its arbitration decision, the “coordinating controller” transmits a “stop” message to any “control device” that the “coordinating controller” did not determine should handle the voice input. *Id.*, 12:49-52, 13:3-7.

Rosenberger discloses embodiments where its “coordination” methods can be initiated based on a “trigger phrase” (e.g., “Hello Voice Control”) contained in an utterance, which a POSITA would understand was also known in the art as a

“hotword” or “wakeword.” Ex.1002, ¶71 (*citing* Ex.1003, 2:1-9, 3:30-35, 8:16-34, FIGs. 4, 5A, 6). Rosenberger also discloses embodiments where its “coordination” methods can be initiated based on a more general “speech command” contained in an utterance (e.g., “**Watch** Cable Channel CNN” or “**Call** the Police”). *See id.*, Abstract, 2:9-11, 3:30-35, 7:19-28, 8:44-49, 10:5-17, FIGs. 4 (“Speech Trigger or Command Recognized”), 5A (same), FIG. 6 (same).

B. Jang

Jang published on March 21, 2013 (Ex.1004) and thus, qualifies as prior art to the ’748 Patent, which has an earliest possible priority date of June 24, 2014.

Jang discloses a system of “electronic devices” (*see, e.g.*, Ex.1004, ¶30, 53, FIGs. 1-2), where each “electronic device” includes a “voice recognition unit” that “can carry out voice recognition upon voice signals input through the microphone 122” of the electronic device. *Id.*, ¶110. Jang explains that the “voice recognition unit” enables each electronic device to recognize a user’s voice command, and there can be scenarios where these devices recognize the same voice command and “generate[] the same output in response to the user’s voice command.” *Id.*, ¶120.

To address this duplicate response problem, Jang provides arbitration techniques such that, “[i]n an environment involving a plurality of devices, it can be determined by communication between the devices or by a third device managing the plurality of devices which of the plurality of devices is to carry out a user’s voice

command.” *Id.*, ¶122. In this way, Jang’s embodiments generally either involve each electronic device performing arbitration independently (*see, e.g., id.*, ¶148-52, FIG. 9) or a “managing” device performing arbitration on behalf of the electronic devices in the system.

For instance, much like Rosenberger, Jang discloses that multiple “electronic devices” detect a voice input. *See, e.g., id.*, ¶110, 127. Each “electronic device” may then compute a respective “voice recognition result” (also sometimes called a “voice command result”), share the “voice recognition results” amongst one another via a communication network, and then use the “voice recognition results” to determine which “electronic device” should handle the user’s request in the voice input. *Id.*, ¶130-132, 135-140, 149-151.

Jang discloses that a “voice recognition result” (or “voice command result”) may take various forms, including “a magnitude (gain value) of the recognized voice signal, voice recognition ratio of each device, type of content or application in execution by each device upon voice recognition, and remaining power.” *Id.*, ¶135; *see also, e.g., id.*, ¶160-63.

As one specific example, Jang discloses “select[ing] an electronic device that may perform a user input voice command most efficiently in an environment involving a plurality of electronic devices,” such as by “identif[ying] an application in execution in each electronic device” and “identif[ying]... an electronic device

executing an application corresponding to an input voice command....” *Id.*, ¶175-181. In other words, Jang’s “electronic device” can select an “electronic device” to respond to a voice input having a command of a type that is relevant or applicable to the selected “electronic device.”

Jang provides a concrete example where “a voice command saying ‘transfer a picture to Chulsu’ enables a predetermined picture to be transferred to an electronic device through emailing or MMS mailing” and “when there is any electronic device executing an application relating to messaging or emailing among a plurality of electronic devices, it is most efficient for the corresponding electronic device to perform the voice command.” *Id.*, ¶178. In this way, Jang discloses that an “electronic device” may recognize a voice command and select itself (or another “electronic device”) to perform the voice command in accordance with determining that a type of the voice command is related to the “electronic device,” such as because the voice command calls for a particular application that the “electronic device” is currently executing.

Jang also explains that various, different types of “voice recognition results” can be considered for purposes of selecting the arbitration winner. *See, e.g., id.*, ¶203-206. In this respect, embodiments of the arbitration process are described as involving assigning “priority values” for each electronic device based on their “voice recognition results” where the “priority values” are “assigned in an order of

electronic devices satisfying better conditions to perform the input voice command among a plurality of electronic devices.” *Id.*, ¶205.

Jang provides the following example of assigning priority based on different types of “voice recognition results”:

For example, *at least one factor* of the user-device distance, voice recognition rate, relevancy between the executing program and a program to be executed through the input voice command, and remaining power in each device may be considered to determine the order of the priority values.

Id., ¶206; *see also id.*, ¶207.

X. Ground 1—Rosenberger (§102)

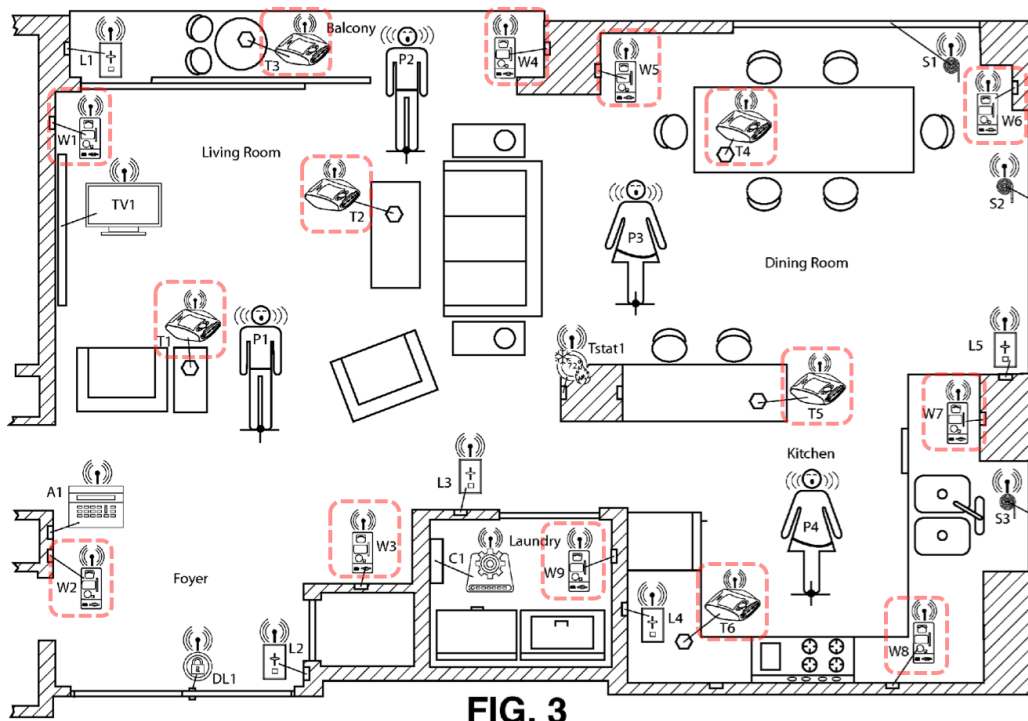
A. Rosenberger Anticipates Independent Claims 1, 7, 11

1. Independent Claim 7

- a. [7.0] [a] first *computing device* comprising: [7.0(a)] *at least one processor*; and [7.0(b)] *at least one module operable by the at least one processor to:*

To the extent that the preamble is limiting, Rosenberger discloses it. Ex.1002, ¶88-98. For instance, Rosenberger discloses “an automated building control system (e.g., a home automation system)” that includes “a plurality of interactive, self-contained, *speech recognition control devices*....” Ex.1003, 4:54-60. An example of such a system is illustrated in Rosenberger’s FIG. 3 below, which shows “several of the [described] table-top *control devices* (T1 through T6) and electrical wall-box mounted *control devices* (W1 through W9) of the invention are strategically

positioned to provide hands-free operation throughout the illustrated space” *Id.*, 9:37-49.

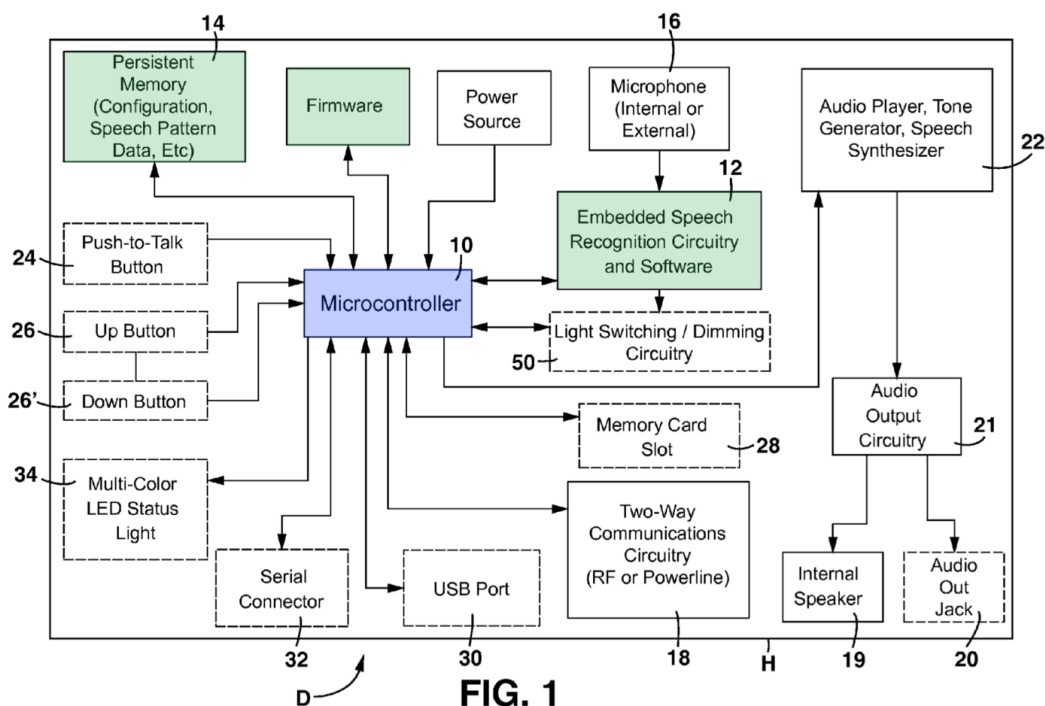


Id., FIG. 3 (annotated); *see also, e.g., id.*, 3:63-4:5.

As explained, Rosenberger discloses “Coordinating Controller” methods where “a coordinating controller provides coordination services for all interactive speech recognition control devices in a given space,” and Rosenberger explains that the “coordinating controller” itself could be implemented on a “control device” that also receives user utterances. *Supra* §IX.A. A “control device” serving as the “coordinating controller” amounts to the *first computing device*. *See* Ex.1002, ¶97.

With reference to FIG. 1 (below), Rosenberger explains that each “**control device**” (*first computing device*)—which Rosenberger sometimes refers to as

“**device D**”—includes “a uniquely-programmed **microcontroller 10** having embedded speech recognition circuitry and software 12, a persistent internal memory 14 (e.g. non-volatile flash memory) for storing speech pattern data (phoneme expressions, etc)” and “**firmware.**” Ex.1003, 6:40-43, 6:62-7:4, 8:8-12.



See also, *id.*, 4:6-32 (“[T]he **control device** of the invention includes... (b) a **programmable microcontroller**, operatively coupled to the microphone and comprising **embedded software and circuitry serving to....**”), 12:28-31 (indicating to POSITA that flow diagrams “illustrate[] a [control] device’s programming logic”).

With respect to the “programmable microcontroller” with “embedded speech recognition circuitry and software,” Rosenberger explains that (i) its “embedded speech recognition circuitry and software 12 is commercially available from a

variety of manufacturers (see, e.g., IC Route’s Model LD3320, and Sensory, Inc.’s Model NLP-5x),” (ii) “[t]hese circuitry and software packages are preferred due, in part, to their ability to perform both speaker independent and speaker dependent speech recognition on a single chip,” and (iii) “[t]he embedded speech recognition circuitry and software 12 and persistent memory 14 in each device is configured to recognize a fixed vocabulary of spoken commands, whether spoken discretely by the end user or as part of a flexible grammar containing one or more commands spoken in a variety of ways.” *Id.*, 7:5-24.

Before the priority date of the ’748 Patent, it was common knowledge to a POSITA that a “microcontroller” (e.g., Rosenberger’s “microcontroller”) amounted to or otherwise included *at least one processor*. Ex.1002, ¶92. Indeed, references cited on the face of the ’748 Patent and technical dictionary definitions routinely equate a “microcontroller” to a “processor,” and the datasheet for the “Sensory NLP-5x” module referenced by Rosenberger (*see, e.g.*, Ex.1003, 10:47-52) confirms that it was a “natural language *processor*” based on a 16-bit “general purpose” digital signal *processor* (DSP). Ex.1002, ¶92-95 (*citing* Exs.1011-1016).

Consequently, a POSITA would have understood that a control device’s “microcontroller,” individually and collectively with the control device’s “persistent memory,” stores software programs/modules/logic in memory that are executable by

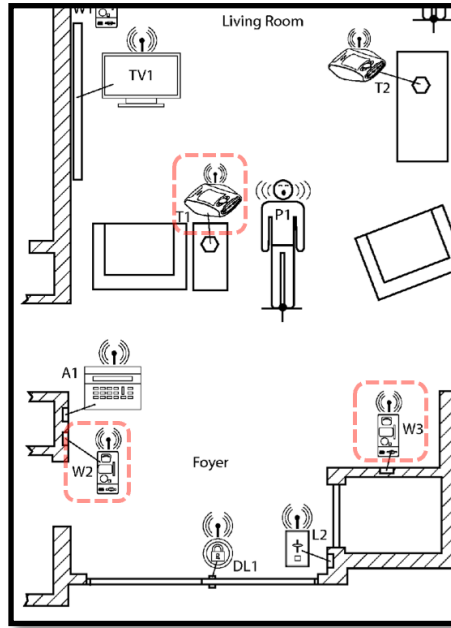
the “microcontroller” to cause the “control device” to perform the functions disclosed in Rosenberg. *Id.*, ¶96.

- b. [7.1] *receive a spoken audio input that is also received by one or more additional computing devices in physical proximity to the first computing device*

Rosenberger discloses this element. Ex.1002, ¶99-106. In this regard, Rosenberg’s “control device” serving as the “coordinating controller” (*first computing device*) is operable to receive *a spoken audio input* that is also received by one or more additional “control devices” (*one or more additional computing devices in physical proximity to the first computing device*).

For instance, Rosenberg discloses that “each control device D is programmed so that multiple devices can be placed *in close proximity* to one another in the same space.” Ex.1003, 10:5-14; *see also id.*, 3:63-4:5. Rosenberg explains “[b]y liberally positioning the devices within a given space (*sometimes several per room*), ‘comfortable’ interaction is achieved....” *Id.*, 6:34-39.

Rosenberger also discloses that two or more of the plurality devices may simultaneously receive a *spoken audio input*. For example, with reference to FIG. 3 (below), Rosenberg discloses, “if person P1 issues a triggering speech command while television TV1 is on, [control] *devices T1, W2 and W3 may all be able to recognize the speech command.*” *Id.*, 12:14-19.



Id., FIG. 3 (annotated; cropped).

Rosenberger further discloses receiving a *spoken audio input* under Google’s broad interpretation. For instance, Rosenberger discloses embodiments of a “control device” receiving a first utterance from a user containing a “speech trigger,” such as “Hello Voice Control” (*id.*, 9:57-63, 10:19-23, 14:26-29), or a more general “speech command” contained in a first utterance (e.g., “**Watch** Cable Channel CNN” or “**Call** the Police”). *See id.*, 7:19-28, 8:44-49. Rosenberger discloses a “control device” also receiving a second utterance from a user containing a more general “speech command from an expected vocabulary or grammar” that can follow the first utterance. *Id.*, 4:6-14; *see also id.*, 3:63-4:5, 6:21-39, 8:20-39:

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... plays a pre-recorded or synthesized audio message (e.g. “How may I help you?”) through speaker 19 and/ or 20 to *prompt the user to say one of the speech commands* from a known vocabulary or grammar stored in the speech pattern memory 14. *Upon recognizing a particular speech command* (and optionally confirming with the user), the device produces and transmits... a corresponding system control signal designed to instruct the appropriate automated mechanisms to *carry out the subject of the speech command*.

Id., 12:23-27 (“When T2’s waiting period expires, it will not have received any weighted signals from T3 or W4 that were higher than its own. Hence, only T2 will interact with person P2 and *handle a subsequent speech command*.”).

As one example, after receiving an initial utterance containing a “speech trigger,” Rosenberger’s “control device” may prompt the user (e.g., “How may I help you?”), and then receive a more general “speech command” contained in a subsequent utterance (e.g., “**Lower** thermostat two degrees”). *See id.*, 14:26-45. As another example, after receiving a more general “speech command” in an initial utterance (e.g., “**Call** the Police”), Rosenberger’s “control device” may prompt the user (e.g., “Are you sure you want to call the police?”), and then again receive a “speech command” contained in a subsequent utterance (e.g., “**Call** the Police”) to confirm the user’s request before executing the call. *See id.*, 3:3-5, 8:44-46.

- c. [7.2] *determine, 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; and*

Rosenberger discloses this element. Ex.1002, ¶107-17. For instance, 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*.

In this respect, Rosenberger generally discloses that, when a plurality of “control devices” receive the same utterance of a speech command, only one of them—the one that produces the highest “weighted signal”—will then handle the utterance or otherwise, interact with the user, which can involve performing speech recognition on subsequent utterances. Specifically, Rosenberger explains that “the control device of the invention uses one of several different ***coordination methods*** to achieve the design goal of avoiding duplicate responses to the same speech command while selecting a single control device best able to successfully interact with the system user.” Ex.1003, 10:34-38; *see also id.*, FIG. 4.

Rosenberger generally describes “coordination” to involve the following steps for the “Coordinating Controller” methods. Ex.1002, ¶109-16.

First, Rosenberger discloses that each “control device” (including one serving as the “coordinating controller”) that detects a speech command (e.g., a “trigger phrase”) is configured to calculate a “weighted signal” (WS) value associated with the spoken audio input received by the given “control device.” Ex.1003, 4:6-16, 11:42-51, FIG. 4, 12:28-48, FIG. 5A. 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.” *Id.*, 10:34-41.

Second, Rosenberger discloses that each “control device” *not* serving as the “coordinating controller” “sends its weighted signal over a network directly to an independent coordinating controller via a unicast message” and “[i]mmediately after sending its weighted signal to the coordinating controller, a device waits a predetermined amount of time for a return unicast message from the coordinating controller.” *Id.*, 12:28-48, FIG. 5A.

Third, Rosenberger explains that the “control device” serving as the “coordinating controller” “operates to compare the value of [its own] 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....” *Id.*, 4:33-44, FIG. 4, 12:57-13:10, FIG. 5B. The “coordinating controller” “waits a predetermined amount of time to receive... weighted signals from other control

devices that may have been triggered by the same speech command” and then identifies the “control device” (such as itself) with the “highest weighted signal....” *Id.*, 12:57-13:10, FIG. 5B.

Fourth, Rosenberger explains that the “control device” with the highest “weighted signal” exchanged as part of “coordination” (such as the one serving as the “coordinating controller”) “proceeds to handle the speech command by interacting with the user as needed and subsequently transmitting the appropriate system control signals if the user does not cancel the command during user interaction.” *Id.*, 11:55-12:3, FIG. 4, 12:28-13:10, FIGs. 5A-5B. In other words, Rosenberger’s “coordination” methods select “a single control device best able to successfully interact with the system user.” *Id.*, 10:34-38.

By “handling the speech command” and “interact[ing] with the system user,” Rosenberger means that the selected “control device” may not only output a visual and/or audible response to the first user utterance but may also perform speech recognition on subsequent user utterances. *See id.*, 8:20-39:

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... plays a pre-recorded or synthesized audio message (e.g. “How may I help you?”) through speaker 19 and/ or 20 to *prompt the user to say one of the speech commands* from a known

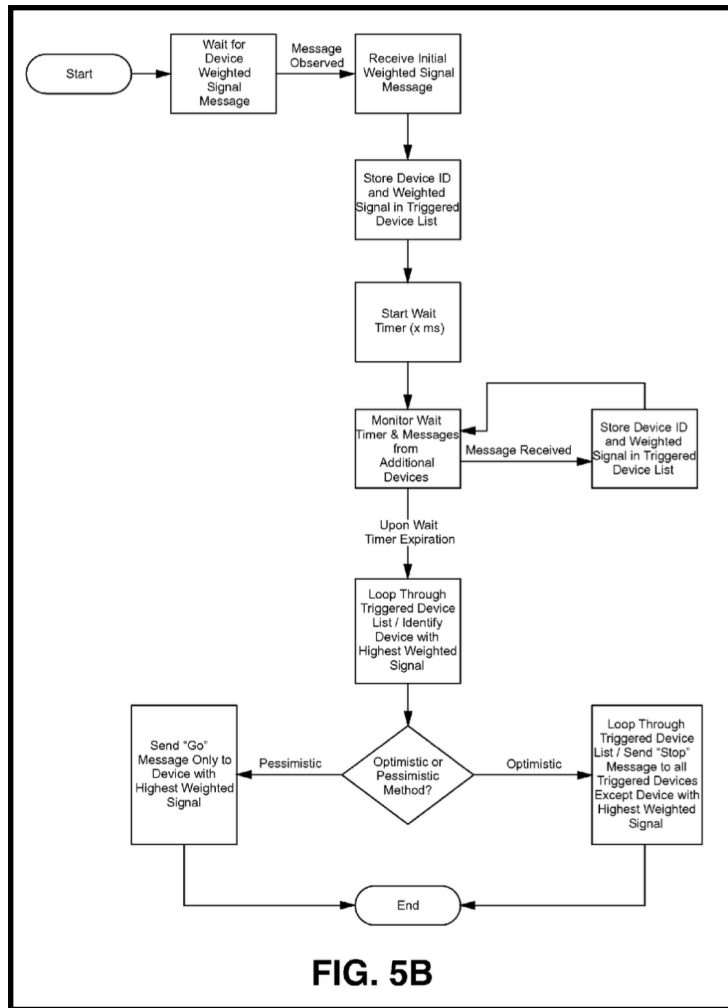
vocabulary or grammar stored in the speech pattern memory 14. ***Upon recognizing a particular speech command*** (and optionally confirming with the user), the device produces and transmits... a corresponding system control signal designed to instruct the appropriate automated mechanisms to carry out the subject of the speech command.

Id., 12:19-26 (“Assuming [control device] T2 receives the highest quality speech command... ***only T2 will*** interact with person P2 and ***handle a subsequent speech command.***”). In this way, Rosenberger’s “coordination” decision is a determination *whether the first computing device should perform speech recognition on the spoken audio input.*

- d. [7.3] *responsive to determining that the first computing device should perform speech recognition on the spoken audio input:*

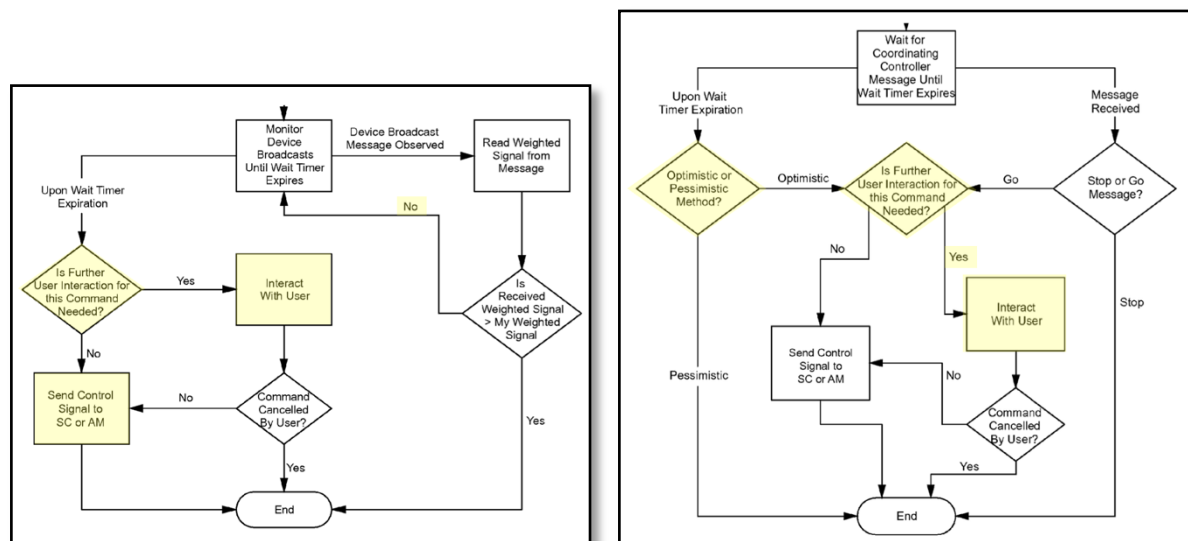
Rosenberger discloses this element. Ex.1002, ¶118-24. For instance, Rosenberger discloses that, responsive to the “control device” serving as the “coordinating controller” (*first computing device*) determining that it won “coordination” and thus, should perform speech recognition on *the spoken audio input*, the “coordinating controller” is operable to perform the functions of elements [7.3(a)-(d)].

Specifically, FIG. 5B illustrates the logic of the “coordinating controller” for performing “coordination”:



As shown, upon the expiration of the “wait timer,” the “coordinating controller” identifies the “control device” with the “highest weighted signal” (which could be itself) and causes that identified “control device” to interact with the user. *See, e.g.*, Ex.1003, 12:57-13:10, FIG. 5B; *see also id.*, 11:55-12:3 (“control device” deemed coordination winner “proceeds to handle the speech command by interacting with the user as needed and subsequently transmitting the appropriate system control signals if the user does not cancel the command during user interaction”).

A POSITA would appreciate that the following portions of Rosenberger’s FIG. 4 and FIG. 5A would apply to the “coordinating controller” when it is the “coordination” winner (Ex.1002, ¶122):



Ex.1003, FIG. 4 (annotated; cropped), FIG. 5B (annotated; cropped).

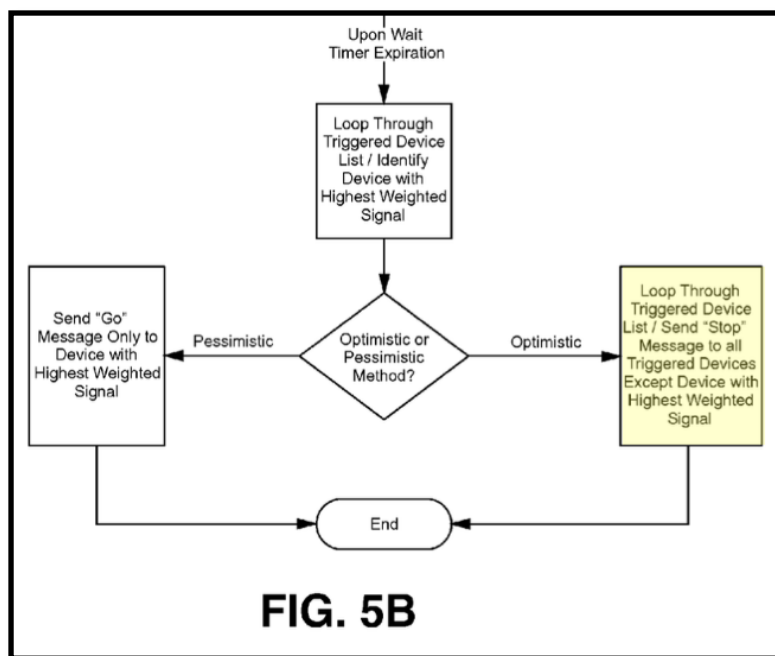
By “handl[ing] the speech command by interacting with the user as needed,” Rosenberger means that the selected “control device” may not only output a visual and/or audible response to the first user utterance but also perform speech recognition on subsequent user utterances. *Supra* §X.A.1.c.

- e. [7.3(a)] send, to at least a second computing device from a plurality of computing devices, a command to temporarily refrain from performing speech recognition

Rosenberger discloses this element, especially under Google’s broad interpretation (*supra* §VII.C.) Ex.1002, ¶125-36. For instance, Rosenberger’s “control device” serving as the “coordinating controller” (*first computing device*) is

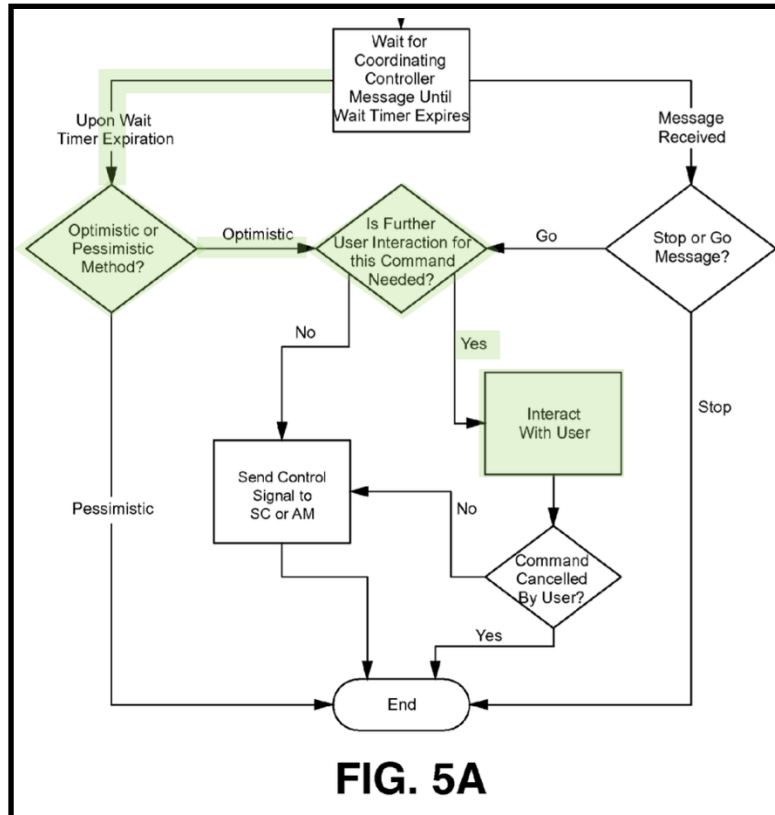
operable such that, responsive to it determining that it should perform speech recognition on *the spoken audio input*, it sends, to at least one other “control device” in the system (*at least a second computing device from a plurality of computing devices*), a “stop” message (*a command to temporarily refrain from performing speech recognition*).

In this regard, Rosenberger discloses a “Coordinating Controller” method that follows an “optimistic model.” Ex.1003, 12:49-52, 13:3-7; *see also id.*, 11:27-36, 12:28-52. For Rosenberger’s “optimistic model,” “when the waiting period expires, the coordinating controller *sends ‘stop’ unicast messages* to all triggered control devices except for the device which sent the highest weighted signal during the waiting period after the initial message was received.” *Id.*, 13:3-10; *see also, id.* 4:60-5:6, 11:27-36, 12:28-13:29, 14:10-45. This is illustrated in FIG. 5B:



Id., FIG. 5B (annotated; cropped). A POSITA would appreciate that, when the “control device” with the highest “weighted signal” is the “coordinating controller” itself, the “stop” message would be sent to all other triggered control devices. Ex.1002, ¶130.

In turn, FIG. 5A below illustrates a control device’s logic for interacting with a user in accordance with Rosenberger’s “Coordinating Controller” methods. Specifically, FIG. 5A illustrates the “optimistic model” where each “control device” assumes it is the coordination winner and continues to process the user’s initial utterance and interact with the user as needed (e.g., by performing speech recognition on subsequent user utterances) after its “wait timer” expires:



Id., ¶131; Ex.1003, FIG. 5A (annotated; cropped); *see also, e.g., id.*, 5:51-57, 12:28-52, 14:10-45.

FIG. 5A further illustrates what happens at a “control device” upon receiving a “stop” message from the “coordinating controller.” Specifically, when a “control device” receives the “stop” message, it understands that it lost “coordination” and should immediately terminate processing the user’s initial utterance and refrain from further interacting with user (e.g., refraining from performing speech recognition on subsequent user utterances). Ex.1002, ¶132.

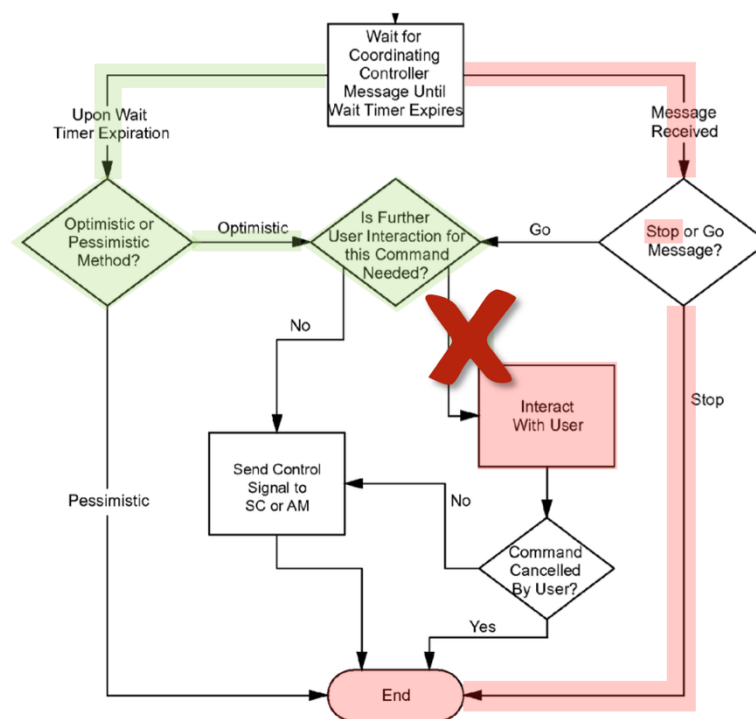


FIG. 5A

Ex.1003, FIG. 5A (annotated; cropped); *see also, e.g., id.*, 12:19-23 (losing devices “each will immediately terminate further processing.”), 12:28-52, 14:10-45.

As a first illustrative example, the “coordination” may be initiated based on a “trigger phrase” (e.g., “Hello Voice Control”) contained in an initial utterance. *Id.*, 2:1-9, 3:30-35, 8:16-34. After receiving the “stop” message from the “coordinating controller,” a “control device” would revert back to listening for the “trigger phrase” and temporarily refrain from performing speech recognition (e.g., on subsequent utterances like “Lower thermostat two degrees”). Ex.1002, ¶133. Google’s own technical expert confirmed this understanding during the 1330 Investigation evidentiary hearing. *See, e.g.*, Ex.1008, p.56 (“If there’s no follow-on request, the other devices are still refraining from performing speech recognition. They don’t know that there’s not going to be another follow-on request, and they’re waiting for another hotword. ***They go back to hotword detection. So they’re refraining by going back to hotword detection.***”), p.41:

Q. Assume that a device is performing hotword detection without interruption, and assume that hotword detection is speech recognition. Are you with me so far?

A. Yes.

Q. So in that scenario, you would agree with me that that device that is continuing to perform hotword detection is not refraining from performing speech recognition; correct?

A. *No*, I’d still disagree because *you’re* going back and ***performing speech recognition for the next spoken audio input***. So *it’s not performing speech recognition on the current spoken audio input*.

As a second illustrative example, the “coordination” may be initiated based on a more general “speech command” (e.g., “Watch”) contained in an initial utterance (e.g., “**Watch** Cable Channel CNN”). Ex.1003, 2:9-11, 3:30-35, 7:19-28, 8:44-49, 10:5-17. After receiving the “stop” message from the “coordinating controller,” a “control device” would refrain from performing speech recognition on the remainder of the initial utterance (e.g., “Cable Channel CNN”). Ex.1002, ¶134. Google’s technical expert confirmed this understanding as well. *See, e.g.*, Ex.1008, pp.66-67, pp.68-70 (conceding Rosenberger’s losing “control device” will arbitrate based on speech command “watch” and then refrain from performing speech recognition on remainder of initial utterance); *supra* §VII.C. (Google’s ITC expert contending that refraining “on the remainder of the spoken audio input, whatever that is” such as “nothing” and for “time” that is “very, very short, which could be practically zero” still constitutes *temporarily refrain[ing]*).

In this way, at least under Google’s broad interpretation (*id.*), Rosenberger’s “stop” message constitutes *a command to temporarily refrain from performing speech recognition* because its purpose is to cause the recipient to temporarily refrain from performing speech recognition on (i) the remainder of the initial utterance that initiated the “coordination” (e.g., in embodiments where “coordination” is initiated based on a speech command rather than “trigger phrase”) and/or (ii) a subsequent

utterance that is part of further user interaction handled by the “coordinating controller” that won coordination. Ex.1002, ¶135.

f. [7.3(b)] *perform speech recognition on the spoken audio input*

Rosenberger discloses this element. Ex.1002, ¶137-41. For instance, Rosenberger’s “control device” serving as the “coordinating controller” (*first computing device*) is operable such that, responsive to it determining that it should perform speech recognition on *the spoken audio input*, the “coordinating controller” performs speech recognition on *the spoken audio input*.

In this regard, as explained, Rosenberger’s “control device” selected as the coordination winner (e.g., the “control device” serving as the “coordinating controller”) may not only output a visual and/or audible response to the first user utterance (e.g., containing the “trigger phrase”) but may also perform speech recognition on subsequent user utterances. *See* Ex.1003, 8:20-39:

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... plays a pre-recorded or synthesized audio message (e.g. “How may I help you?”) ... to ***prompt the user to say one of the speech commands.... Upon recognizing a particular speech command*** (and optionally confirming with the user), the device ...

instruct[s] the appropriate automated mechanisms to carry out the subject of the speech command.

Id., 12:19-26 (“[O]nly T2 will interact with person P2 and **handle a subsequent speech command**”).

Moreover, Rosenberger is replete with disclosure of each “control device” being configured to perform speech recognition on spoken audio input. Ex.1002, ¶140. For instance, Rosenberger discloses that “[the] control device includes **embedded speech recognition** software and circuitry capable of quickly **recognizing spoken speech commands** from an expected vocabulary or ‘flexible grammar.’” Ex.1003, 6:21-25; *see also id.*, 2:1-15, 4:6-24, 4:33-44, 7:5-31, 10:47-52.

- g. [7.3(c)] *determine, based on performing speech recognition on the spoken audio input, whether the spoken audio input includes a predetermined spoken audio command; and*

Rosenberger discloses this element. Ex.1002, ¶142-46. For instance, Rosenberger’s “control device” serving as the “coordinating controller” (*first computing device*) is operable such that, responsive to it determining that it should perform speech recognition on *the spoken audio input*, the “coordinating controller” determines, based on performing speech recognition on *the spoken audio input*, whether *the spoken audio input* includes a particular speech command.

As just discussed, the selected “control device” (e.g., the “control device” serving as the “coordinating controller”) may perform speech recognition on a

subsequent user utterance and determine whether such an utterance includes a particular speech command. *Supra* §X.A.1.f.

Moreover, Rosenberger is replete with disclosure of each “control device” being configured to determine whether a user utterance includes a particular speech command based on performing speech recognition. *See* Ex.1003, 6:21-25 (“[C]ontrol device[s] include[] embedded speech recognition software and circuitry capable of *quickly recognizing spoken speech commands* from an expected vocabulary or ‘flexible grammar.’”), 10:47-52 (“Embedded speech recognition IC modules and software... calculate and return [] measures each time they successfully *match (i.e. recognize) a spoken word or phrase* with their *active vocabulary or grammar.*”), 8:32-39, 14:32-45; *see also, e.g., id.*, 2:6-15, 4:6-24, 7:5-31, 16:21-23.

- h. [7.3(d)] *responsive to determining that the spoken audio input includes the predetermined spoken audio command, execute an action associated with the predetermined spoken audio command.*

Rosenberger discloses this element. Ex.1002, ¶147-53. For instance, Rosenberger’s “control device” serving as the “coordinating controller” (*first computing device*) is operable such that, responsive to it determining that it should perform speech recognition on *the spoken audio input*, the “coordinating controller,” responsive to determining that *the spoken audio input* includes a particular speech command, executes an action associated with the particular speech command.

As discussed, the selected “control device” (e.g., the “control device” serving as the “coordinating controller”) may perform speech recognition on a subsequent user utterance, determine that such an utterance includes a particular speech command, and then execute an action associated with that particular speech command. *Supra* §X.A.1.f.; *see also, e.g.*, Ex.1003, 8:20-39:

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... plays a pre-recorded or synthesized audio message (e.g. “How may I help you?”) ... to prompt the user to say one of the speech commands.... ***Upon recognizing a particular speech command*** (and optionally confirming with the user), the device ***produces and transmits... a corresponding system control signal*** designed to instruct the appropriate automated mechanisms ***to carry out the subject of the speech command.***

Moreover, Rosenberger is replete with disclosure of each “control device” being configured to execute an action associated with a particular speech command that it determined to be included in a user utterance. *See, e.g.*, Ex.1003, 4:6-33 (“[R]ecognize from the electrical signals when a speech command from an expected vocabulary or grammar has likely been spoken... produce... predetermined system control signals by which a desired effect, ***reflective of the speech commands***, is achieved within the building space”), 4:54-5:6 (“[E]ach of the respective

microcontrollers of the control devices operates to... determine whether said microcontroller will produce the audible user interactions and system control signals ***necessary to carry out a given speech command*** within said building space.”), 8:20-42, 8:39-54 (describing “mapping of speech commands” to corresponding actions); *see also, id.*, 2:1-15, FIG. 4 (Send Control Signal to [system controller] SC or [automated mechanism] AM), FIG. 5A (same), FIGs. 7-10 (illustrating “control devices” sending control signals directly or indirectly to AMs).

As one concrete example, Rosenberger discloses a “control device” determining that a user utterance includes a particular speech command related to lowering a thermostat’s temperature setting (e.g., “[l]ower thermostat two degrees”) and executing an action associated with the particular speech command (e.g., “[t]he system controller will then send compatible ***instructions to thermostat Tstat1 telling it to lower its set point by two degrees***”). *Id.*, 13:32-45.

Rosenberger further explains that a “control device” (“device D”) can “be used to directly control the operation of wirelessly-controllable electronic devices and appliances (i.e. ‘automated mechanisms’),” such that “thermostats, light switches, motorized shading systems, audio-visual equipment, security systems, kitchen equipment, pumps, motors and actuators may all be controlled directly by device D.” *Id.*, 6:54-61; *see also, e.g., id.*, 16:48-60. Rosenberger also contemplates

that a “control device” could be embedded or integrated into another such device/appliance. *See, e.g., id.*, 15:62-16:36.

2. Independent Claim 1

Independent claim 1 recites *a method* corresponding to the recited functions of independent claim 7 that the *first computing device* is operable to perform. Specifically, independent claim 1 recites:

[1.0] *A method comprising:*

[1.1] *receiving, by a first computing device of a plurality of computing devices in physical proximity to one another, a spoken audio input that is also received by one or more additional computing devices of the plurality of computing devices;*

[1.2] *determining, 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; and*

[1.3] *responsive to determining that the first computing device should perform speech recognition on the spoken audio input:*

[1.3(a)] *sending, by the first computing device and to at least a second computing device from the plurality of computing devices, a command to temporarily refrain from performing speech recognition;*

[1.3(b)] *performing, by the first computing device, speech recognition on the spoken audio input;*

[1.3(c)] *determining, by the first computing device, based on performing speech recognition on the spoken audio input,*

whether the spoken audio input includes a predetermined spoken audio command; and

[1.3(d)] responsive to determining that the spoken audio input includes the predetermined spoken audio command, executing, by the first computing device, an action associated with the predetermined spoken audio command.

As is evident, elements [1.0], [1.1], [1.2], [1.3], [1.3(a)], [1.3(b)], [1.3(c)] and [1.3(d)] map to elements [7.0], [7.1], [7.2], [7.3], [7.3(a)], [7.3(b)], [7.3(c)] and [7.3(d)], respectively. Thus, Rosenberger discloses claim 1 for the same reasons it discloses claim 7. Ex.1002, ¶154-57.

3. Independent Claim 11

Independent claim 11 recites *a non-transitory computer-readable storage medium encoded with instructions that... cause at least one processor of a first computing device to perform the functions recited in independent claim 7.* Specifically, independent claim 11 recites:

[11.0] A non-transitory computer-readable storage medium encoded with [11.0(b)] instructions that, when executed, cause [11.0(a)] at least one processor of a first computing device to:

[11.1] receive a spoken audio input that is also received by one or more additional computing devices in physical proximity to the first computing device;

[11.2] determine, 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; and

[11.3] responsive to determining that the first computing device should perform speech recognition on the spoken audio input:

[11.3(a)] send, to at least a second computing device from a plurality of computing devices, a command to temporarily refrain from performing speech recognition;

[11.3(b)] perform speech recognition on the spoken audio input;

[11.3(c)] determine, based on performing speech recognition on the spoken audio input, whether the first audio input includes a predetermined spoken audio command; and

[11.3(d)] responsive to determining that the spoken audio input includes the predetermined spoken audio command, execute an action associated with the predetermined spoken audio command.

As is evident, elements [11.0], [11.1], [11.2], [11.3], [11.3(a)], [11.3(b)], [11.3(c)] and [11.3(d)] map to elements [7.0], [7.1], [7.2], [7.3], [7.3(a)], [7.3(b)], [7.3(c)] and [7.3(d)], respectively. Thus, Rosenberger discloses claim 11 for the same reasons it discloses claim 7. Ex.1002, ¶154-57.

B. Rosenberger Anticipates Dependent Claims 2-4, 9-10, 12, 14-15

1. Dependent Claims 2, 9, 12

Claim 9 depends on claim 7 and further recites:

[9.1] *wherein the at least one module is further operable by the at least one processor to determine a current context of the first computing device, and*

[9.2] *wherein the at least one module is operable by the at least one processor to determine whether the first computing device should perform speech recognition on the spoken audio input further based on the current context of the first computing device.*

Rosenberger discloses these additional elements. Ex.1002, ¶160-73. And, for the same reasons that Rosenberger discloses these additional elements, Rosenberger discloses each of the elements of dependent claims 2 and 12. *Id.*, ¶174-75.

First, Rosenberger’s “control device” serving as the “coordinating controller” (*first computing device*) is operable to determine a *current context* of the “coordinating controller,” thereby satisfying [9.1].

As an initial matter, the ’748 Patent broadly defines a device’s *current context*, as evident from dependent claim 10’s list of *current contexts*.² The ’748 specification describes a device’s “context” even more broadly:

² Claim 10 depends from claim 9. Thus, claim 10 must be **narrower** than claim 9, such that claim 9 covers other types of *current contexts* not expressly recited in dependent claim 10. *See InterDigital Commc'ns, LLC v. Int'l Trade Comm'n*, 690 F.3d 1318, 1324 (Fed. Cir. 2012) (“The clear implication of narrowing the term ‘code’ in dependent claim 5 by limiting the claim scope to cases in which the claimed code ‘is a spreading code’ is that the term ‘code’ in the independent claim is not limited to a spreading code.”).

- “The context of a computing device may be made up of ***any number of items of contextual information***. Example items of contextual information may include ... ***data obtained from one or more sensors³ of the computing device*** (e.g., temperature data, ambient noise level data, light level data, acceleration/ movement data, image/video data, ***and other data***), or ***any other data about the status or current state of the computing device***.” Ex.1001, 8:30-52.
- “In some examples, an item of contextual information may additionally or alternatively be ***data representing information about a user of the computing device***, such as a name of the user, a user identification (UID) of the user, information from one or more social media network service accounts associated with the user, information from one or more calendars or scheduling applications associated with the user, information indicating one or more social or professional relationships of the user (e.g., user contacts), or ***any other information about the user***.” *Id.*, 8:53-62.
- “[D]evice context module 36 may be operable to obtain information from other components of computing device 2A (e.g., application modules 34, communications units 4A, sensors 26, etc.) and/or from other computing devices in order to determine ***a current context that describes computing device 2A, the environment*** in which computing device 2A is situated, ***the activities*** in which ***computing device 2A is current engaged, a user*** of computing device 2A, ***the***

³The '748 Patent explains that the computing device's “one or more sensors 26” include “a microphone or other audio sensor.” *Id.*, 11:60-12:11.

activities in which the user of computing device 2A is currently engaged, and/or other information.” Id., 13:1-12.

- “Device context module 36 may 20 communicate with one or more of sensors 26 to obtain *information about the physical environment* in which computing device 2A exists (e.g., a light level, an ambient noise level, a temperature level, a humidity level, a location, etc.).” *Id., 13:20-24.*

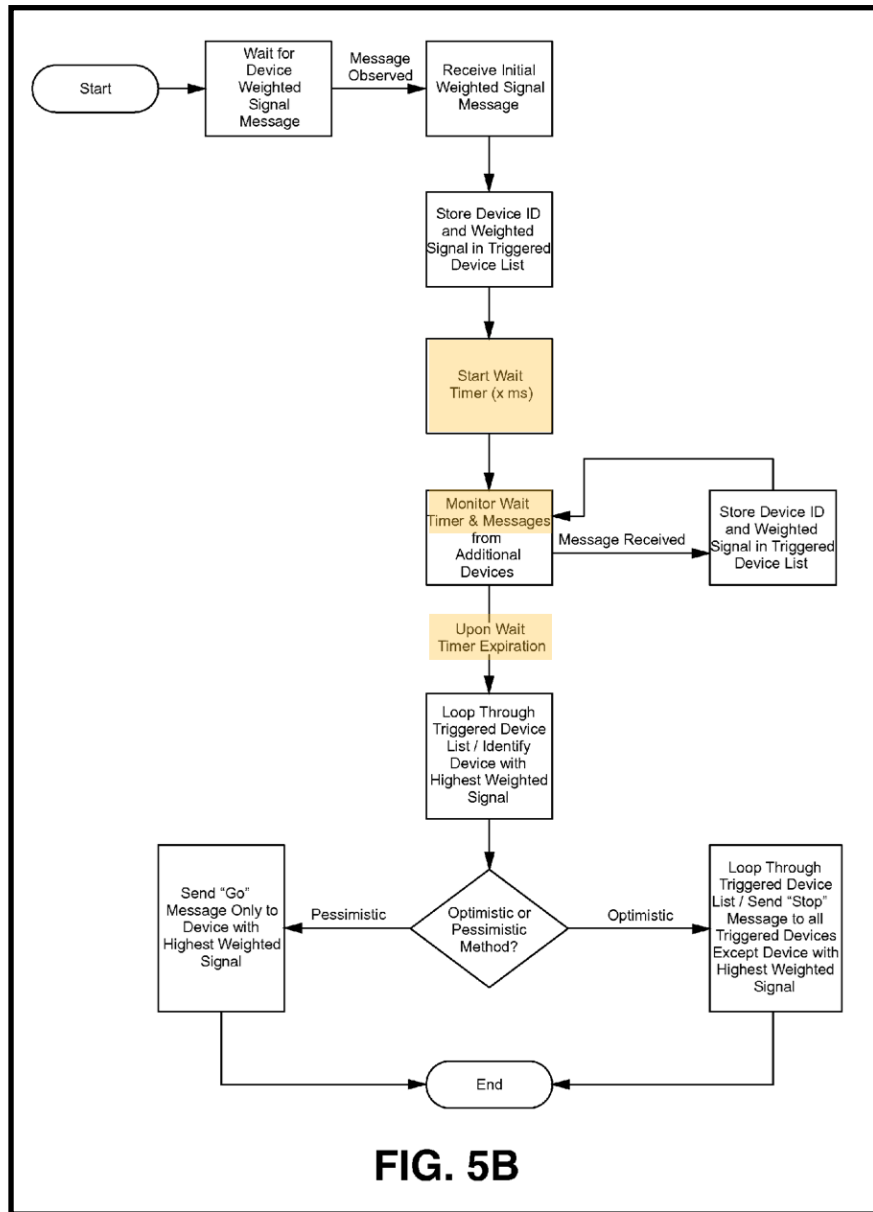
As further discussed below, Rosenberger discloses determining multiple *current contexts* broadly described in the ’748 Patent, including (i) the coordinating controller’s “wait timer,” which constitutes *a current time as defined by the first computing device*, as recited in claim 10 and (ii) the coordinating controller’s own “weighted signal” comprising data obtained from its microphone, which constitutes “data obtained from one or more sensors of the computing device” that describes the environment in which the coordinating controller is situated, as described in the ’748 specification (*id.*, 8:47-50, 13:1-12).⁴ Ex.1002, ¶166.

Second, Rosenberger’s “control device” serving as the “coordinating controller” (*first computing device*) is operable to determine whether it should perform speech recognition on *the spoken audio input* further based on the *current*

⁴ The ’748 Patent explains that, “[i]n some examples, device context module 36 may not communicate directly with sensors 26. In some examples, device context module 36 may obtain data from modules of computing device 2A (not shown) that process sensor data received from sensors 26.” ’748 Patent, 13:25-29.

context determined for [9.1], thereby satisfying element [9.2].

As one example, Rosenberger’s “control device” serving as the “coordinating controller” (*first computing device*) determines whether itself or another “control device” is the “coordination winner” (and thus, should perform speech recognition) based on *a current time as defined by the first computing device* (“coordinating coordinator”) and a WS message received from another “control device.” In this regard, the “coordinating controller” initiates a “wait timer” that dictates when other control devices’ transmitted WS messages can affect the coordinating controller’s “coordination” determination, which is illustrated in FIG. 5B:



Ex.1003, FIG. 5B (annotated); *see also id.*, 4:33-53, 12:57-66 (“[T]he coordinating controller... ***waits a predetermined amount of time*** to receive... weighted signals from other controls devices that may have been triggered by the same speech command.”), 13:3-7 (“[W]hen the ***waiting period*** expires, the coordinating controller sends ‘stop’ unicast messages to all triggered control devices except for

the device which sent the highest weighted signal during the waiting period after the initial message was received.”), 13:20-29.

As another example, Rosenberger’s “control device” serving as the “coordinating controller” (*first computing device*) determines whether itself or another “control device” is the “coordination winner” (and thus, should perform speech recognition) based on comparing the coordinating controller’s *own* “weighted signal” that is derived from data obtained from its microphone to any WS received from any other “control device.” *See id.*, 4:6-5:6, 10:38-47:

All of the 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*. Preferably, the value of the weighted signal is based on at least two variables, specifically, the *microphone signal strength* (SS), which is *directly related to the proximity of the user to the device*, and *the recognition confidence (RC)* which quantifies the probability that a speech command was recognized correctly.

See also id., 10:5-11:26, 11:55-12:3 (describing how a “control device” compares its own WS to WS from another “control device”), FIG. 4, 12:57-13:10; *supra* §X.A.1.c. In this way, the “coordination” decision is made based on *both* the coordinating controller’s *own* WS and any WS received from any other “control device.” *Compare with* Ex.1001, 17:10-40.

2. Dependent Claims 3, 10

Claim 10 depends on claim 9 and further recites:

[10.1] *wherein the at least one module operable to determine the current context of the first computing device is operable by the at least one processor to determine one or more of: a location of the first computing device, a current time as defined by the first computing device, one or more applications installed at the first computing device, one or more applications currently executing at the first computing device, one or more networks available to the first computing device, one or more other computing devices in proximity to the first computing device, an operating mode of the first computing device, an ambient temperature of the location of the first computing device, an ambient noise level of the location of the first computing device, an ambient light level of the location of the first computing device, a movement of the first computing device, a name of a user of the first computing device, a user identification (UID) of the user of the first computing device, a social media network service account associated with the user of the first computing device, one or more calendars associated with the user of the first computing device, or one or more social relationships of the user of the first computing device.*

Rosenberger discloses this additional element. Ex.1002, ¶177-78. And, for the same reasons that Rosenberger discloses this additional element, Rosenberger discloses

the additional element of dependent claim 3.⁵ *Id.*, ¶179-81.

For instance, as explained in connection with dependent claim 9, one of the current contexts that the “coordinating controller” determines is *a current time as defined by the first computing device* (i.e., the “coordinating controller” itself). *Supra* §X.B.1.

3. Dependent Claims 4, 14, 15

Claim 15 depends on claim 7 and further recites:

[15.1] wherein the information received from the one or more additional computing devices includes an indication of an audio input quality associated with a second computing device from the one or more additional computing devices, and

[15.2] wherein the at least one module is operable by the at least one processor to determine whether the first computing device should perform speech recognition on the spoken audio input further based at least in part on the audio input quality associated with the second computing device and an audio input quality associated with the first computing device.

Rosenberger discloses these additional elements. Ex.1002, ¶182-91. And, for the same reasons that Rosenberger discloses these additional elements, Rosenberger discloses the additional elements of dependent claims 4 and 14. *Id.*, ¶192-95.

⁵ For purposes of this IPR only, Sonos has assumed that the phrase *the sec-end first computing device* in claim 3 instead reads as “the first computing device.”

First, Rosenberger discloses that each “weighted signal” (*information*) received from another “control device” (*one or more additional computing devices*) includes an indication of an audio input quality associated with the other “control device” (*second computing device*), thereby satisfying element [15.1].

In this regard, Rosenberger discloses that “[a]ll of the [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.” Ex.1003, 10:38-41; *see also id.*, 4:14-16 (“a weighted signal representing ***the quality*** of a just-recognized speech command”), 5:51-56. Rosenberger states, “[p]referably, the value of the weighted signal is based on at least two variables, specifically, the microphone signal strength (SS), which is directly related to the proximity of the user to the device, and the recognition confidence (RC) which quantifies the probability that a speech command was recognized correctly.” *Id.*, 10:41-47.

With reference to Figs. 4, 5A, 5B, and 6, Rosenberger further discloses that the “control devices” may execute “a number of methods for the devices to be coordinated so that ***only the single device with the highest weighted signal*** proceeds to interact with the user and/or process the speech command.” *Id.*, 11:27-14:9; *see also id.*, 4:33-53.

Second, Rosenberger’s “control device” serving as the “coordinating controller” (*first computing device*) is operable to determine whether it should

perform speech recognition on *the spoken audio input* further based at least in part on the “weighted signal” from the other “control device” (*the audio input quality associated with the second computing device*) and the coordinating controller’s own “weighted signal” (*an audio input quality associated with the first computing device*), thereby satisfying element [15.2].

For instance, as discussed with [15.1], each “weighted signal” includes an indication of an audio input quality associated with a given “control device.” Moreover, as explained in connection with [7.2] and [9.2], the “coordinating controller” determines whether it or another “control device” is the “coordination winner” (and thus, should perform speech recognition) based on comparing the coordinating controller’s *own* “weighted signal” to any WS received from any other “control device.” *Supra* §§X.A.1.c., X.B.1.

XI. Ground 2—Rosenberger + Jang (§103)

Ground 1B is presented to the extent Google argues that Rosenberger does not adequately disclose the additional elements of claims 2/9/12 and 3/10. Other than Jang making up for this hypothetical deficiency, Rosenberger and Jang invalidate independent claims 1/7/11 and dependent claims 4/14/15 for the same reasons as Ground 1.

A. Rosenberger+Jang Render Obvious the Additional Elements of Claims 2/9/12 & 3/10

As explained above, Rosenberger discloses the independent claims. While Rosenberger itself discloses the additional elements of claims 2/9/12 and 3/10 (*supra* §§X.B.1.-X.B.2.), Rosenberger in combination with Jang renders obvious these additional elements as well. Ex.1002, ¶196-214.

First, Jang discloses that an “electronic device” (*first computing device*) is operable to determine several *current contexts* of itself that are found in the long list of dependent claim 10, thereby satisfying elements [9.1] and [10.1].⁶ *Id.*, ¶200-202.

For instance, Jang discloses determining multiple *current contexts* as described in the ’748 Patent (*supra* §X.B.1.), including at least (i) “time that the user’s voice command was entered” at the “electronic device” (Ex.1004, ¶133-34), which constitutes *a current time as defined by the first computing device*, as recited in claim 10, (ii) an amount of power remaining at the “electronic device” (*id.*, ¶20, 135, 183-87, 206, FIGs. 17-18), which constitutes *an operating mode of the first computing device*, as recited in claim 10,⁷ and (iii) whether the “electronic device”

⁶ As explained before, dependent claims 2/12 and 3 recite similar subject matter as claims 9 and 10, respectively, and thus, are invalidated for the same reasons as claims 9 and 10. Ex.1002, ¶160, 173-76, 179-81, 214.

⁷ See Ex.1001, 8:26-47 (“Example items of contextual information may include... data that represents an operating mode (e.g., silent mode, airplane mode, driving mode, standby mode, **low battery mode**, or any other mode of operation) of the computing device...”), 17:4-9 (“[D]esignation of computing device 2B to monitor for and/or process audio input may reflect the fact that computing device 2B is likely not **a battery-powered** device. That is, computing device 2B is less likely than computing devices 2A and 2C to have **limited available power**.”).

is “executing an application relating to” a received voice command (Ex.1004, ¶174-81, FIGs. 15-16), which constitutes *one or more applications currently executing at the first computing device*, as recited in claim 10. These examples are further discussed with reference to [9.2] below.

Second, Rosenberger discloses that a “control device” serving as the “coordinating controller” (*first computing device*) is operable to perform arbitration (or “coordination” in Rosenberger’s parlance) to select, based on the coordinating controller’s *current context* and a “weighted signal” (*information*) received from another “control device,” a single “control device” responsible for handling a user’s speech command contained in an initial utterance and interact with the user, which can involve performing speech recognition on subsequent user utterances. *Supra* §§X.A.1.c., X.B.1.

Similarly, Jang’s “electronic device” is operable to perform arbitration to identify, based on the given electronic device’s *current context* and “voice recognition results”/“voice command results” (*information*) received from another “electronic device,” a single “electronic device” responsible for handling a user’s speech command and interact with the user. Ex.1004, ¶130-132, 135-140, 149-151; *supra* §IX.B.

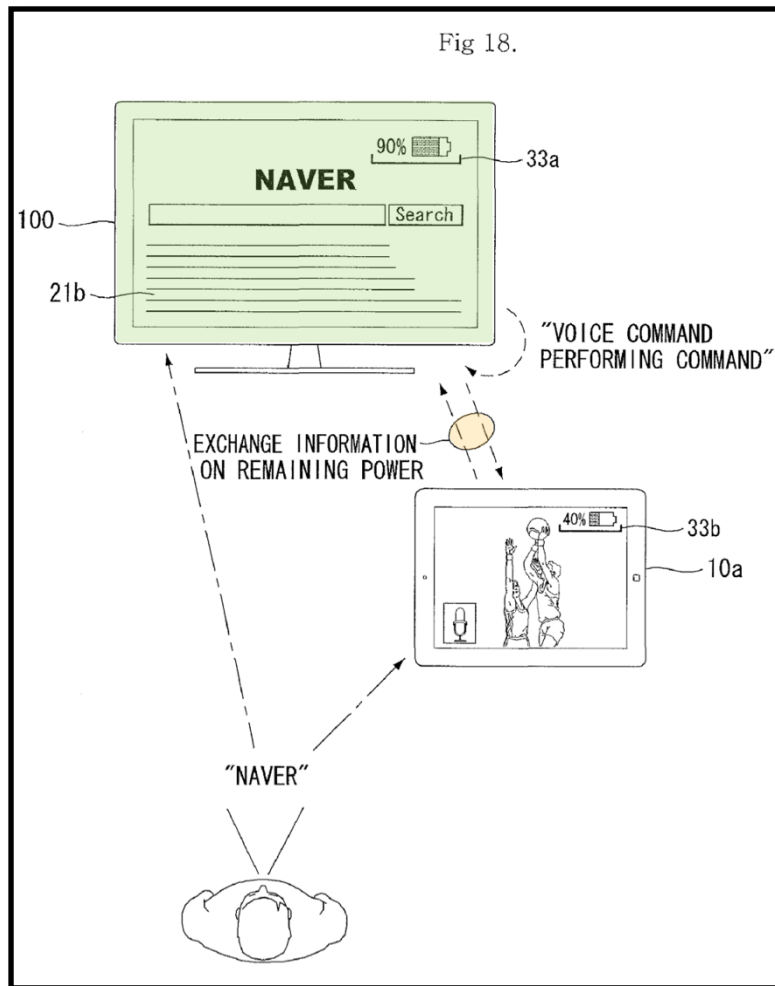
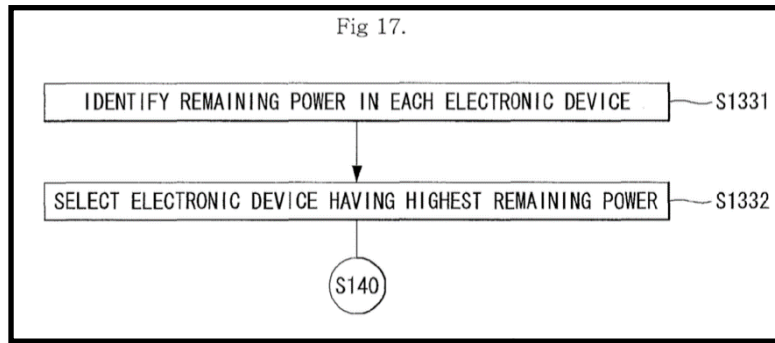
Jang discloses embodiments where the arbitration determination is based on multiple, different types of “voice recognition results” (or “voice command results”).

Ex.1004, ¶135 (“The voice command result received from the other electronic devices may include a magnitude (gain value) of the recognized voice signal, voice recognition ratio of each device, type of content or application in execution by each device upon voice recognition, and remaining power.”), ¶160-63; *see also, id.*, ¶205-206 (“For example, ***at least one factor*** of the user-device distance, voice recognition rate, relevancy between the executing program and a program to be executed through the input voice command, ***and*** remaining power in each device may be considered to determine the order of the priority values.”).

With respect to *a current time as defined by the first computing device*, Jang discloses that each “electronic device” is operable to make an arbitration determination based on a current time of the given “electronic device” (e.g., “first electronic device 100”) and “voice recognition results” containing “***time information***” from one or more other “electronic devices” (e.g., “second electronic device 10”). *Id.*, ¶133-134. In particular, Jang explains that the given “electronic device” compares the difference between its own “input time” and the “input time” received from another “electronic device” to a “predetermined interval” and continues with the arbitration process only “when [the] difference in input time... is within [the] predetermined interval.” *Id.*

With respect to *an operating mode of the first computing device*, Jang discloses that each “electronic device” is operable to make an arbitration

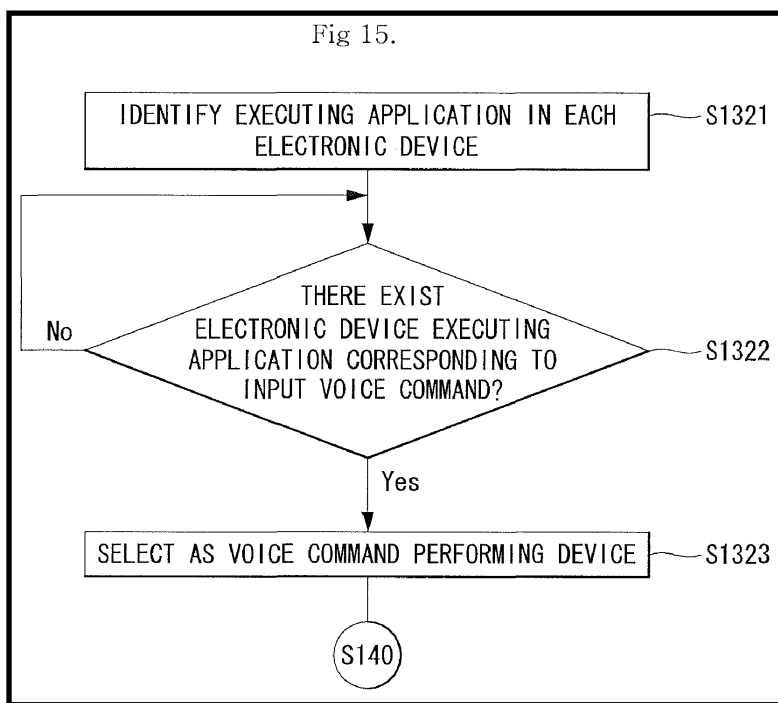
determination based on the amount of power remaining in the given “electronic device” (e.g., “first electronic device 100”) and “voice recognition results” containing the amount of power remaining in one or more other “electronic devices” (e.g., “second electronic device 10”). *Id.*, ¶182-86. Specifically, Jang explains that the given “electronic device” compares remaining power levels and selects the electronic device “having more remaining power....” *Id.* This is illustrated in the figures below:



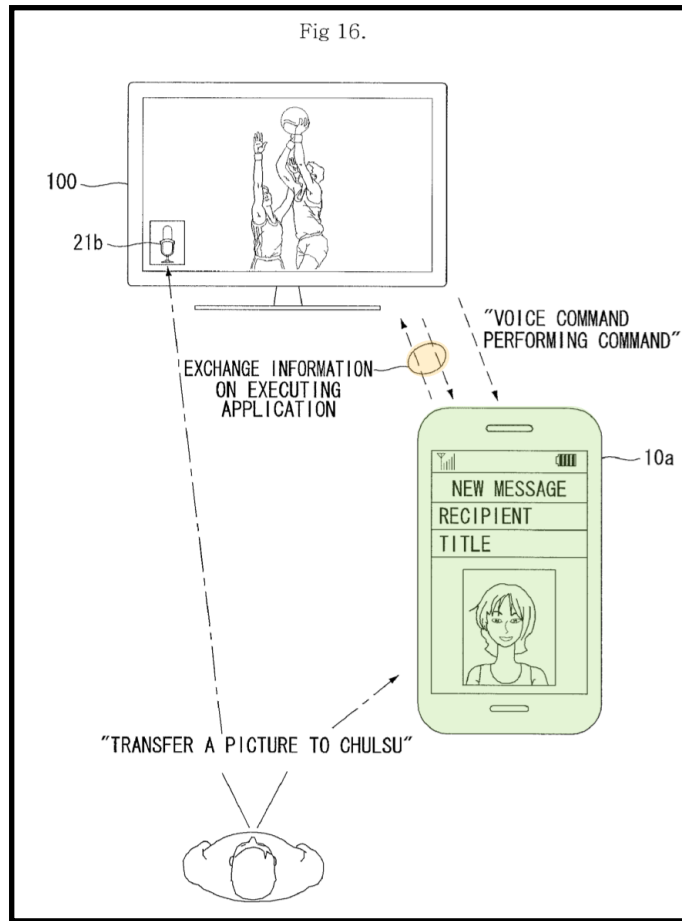
Id., FIGs. 17, 18 (annotated). As shown, "first electronic device 100" selects itself as the arbitration winner based on its determined remaining power level being

greater than the remaining power level of the “second electronic device 10.” *Id.*, ¶182-86.

With respect to *one or more applications currently executing at the first computing device*, Jang discloses that each “electronic device” is operable to make an arbitration determination based on an application executing at the given “electronic device” (e.g., “first electronic device 100”) and “voice recognition results” indicating an application executing at one or more other “electronic devices” (e.g., “second electronic device 10”). Specifically, Jang explains that the given “electronic device” selects an “an *electronic device executing an application corresponding* to an input *voice command* among a plurality of electronic devices....” *Id.*, ¶175-76. This is illustrated in FIG. 15 below:



As one specific example, Jang discloses a scenario where “a voice command saying ‘*transfer a picture to Chulsu*’” leads to the “first electronic device 100” selecting the “second electronic device 10” as the arbitration winner based on the first electronic device’s current context (e.g., executing a “*broadcast program*”) being deemed not related to the voice command and the second electronic device’s current context (e.g., executing an “*email application*”) deemed related to the voice command. *Id.*, ¶177-181. This is illustrated in FIG. 16 below, where “[t]he first electronic device 100 determines that the *second electronic device* 10a may *efficiently perform the newly input voice command* through the program executed by the second electronic device 10a, and selects the second electronic device 10a as the voice command performing device” (*id.*):



Id., FIG. 16 (annotated).

Thus, Rosenberger combined with Jang renders obvious [9.2]. Ex.1002, ¶212.

B. Motivation to Combine

For the reasons below, a POSITA would have found it obvious to, and would have been motivated to, combine Rosenberger with Jang before the earliest possible priority date of the '748 Patent (June 24, 2014) such that Rosenberger's "control device" serving as the "coordinating controller" would select the "coordination" winner (and thus, the "control device" responsible for performing speech recognition) further based on the coordinating controller's determined *current*

context, such as one or more of Jang’s *current contexts* (e.g., current time, remaining power, and/or application executing). Ex.1002, ¶215-25.

First, Rosenberger and Jang are in the same field of endeavor, relate to voice-enabled devices in similar environments, and seek to address the same problem—namely, the duplicate response problem where multiple voice-enabled devices detect the same user utterance. *See* Ex.1003, 2:42-59, 9:18-10:4 (describing “duplicate response” problem); Ex.1004, ¶119-22 (“[T]hat the plurality of devices simultaneously process the use’s voice command may ***cause a multi process to be unnecessarily performed***.... In an environment involving a plurality of devices, it can be determined by communication between the devices or by a third device managing the plurality of devices ***which of the plurality of devices*** is to carry out a user’s voice command.”).

Second, Rosenberger and Jang have interrelated teachings, both disclosing similar arbitration processes to solve the same duplicate response problem. *See* Ex.1003, 11:27-14:9, FIGs. 4, 5A-5B; Ex.1004, FIGs. 11-12.

Indeed, Rosenberger and Jang both disclose arbitration techniques where each device in the system makes the arbitration determination individually and where one device (in Rosenberger, the “coordinating controller”; in Jang, the “managing” device) makes the arbitration determination on behalf of the devices in the system. *Supra* §§IX.A.-IX.B. Further, Rosenberger and Jang’s disclosed arbitration

techniques involve electronic devices configured to determine, receive, and compare values associated with a detected voice input (e.g., loudness/sound pressure level in Rosenberger or magnitude (gain value) or volume of voice signal in Jang) and select the electronic device with the highest value to respond to the user. *See, e.g.*, Ex.1003, 10:34-67, 11:27-31; Ex.1004, ¶122, 127-136, FIG. 11. A POSITA studying Rosenberger would have been motivated to seek out references like Jang that disclose similar arbitration techniques to obtain additional details regarding such techniques. Ex.1002, ¶218.

Third, Jang itself provided a POSITA motivation to combine Jang’s teachings regarding *current context* with Rosenberger.

For example, Jang expressly explains that Jang’s *current contexts* (e.g., current time, remaining power, and/or application executing) would supplement and be incorporated into a “coordination” technique like Rosenberger’s, which involved exchanging audio quality values (e.g., “weighted signals”). *See* Ex.1004, ¶135 (explaining that “voice recognition result” used to select a device could include multiple, different types of information beyond “a magnitude (gain value) of the recognized voice signal,” including a “voice recognition ratio of each device, type of content or application in execution by each device upon voice recognition, and remaining power.”), ¶203-207.

As another example, Jang expressly touts the benefits of utilizing Jang’s

application-executing *current context*, stating that it promotes “select[ing] an electronic device that may perform a user input voice command ***most efficiently*** in an environment involving a plurality of electronic devices ***to effectively conduct*** the voice command.” *Id.*, ¶177; *see also id.*, ¶178 (explaining it is “most efficient” for device “executing an application relating to” the user’s voice command to “perform the voice command.”), ¶180.

A POSITA would have especially been motivated to incorporate Jang’s application-executing *current context* when implementing Rosenberger’s “control devices” that were “incorporate[d] or embed[ded]” into devices/appliances that execute particular applications, so that user commands related to a specific application were most efficiently handled by the system. *See* Ex.1002, ¶222 (*citing* Ex.1003, 15:62-16:36).

Likewise, a POSITA would have been motivated to incorporate Jang’s remaining-power *current context* when implementing Rosenberger’s “control device” as a “[t]abletop device[.]... ideally run on batteries” *Id.*, ¶223 (*citing* Ex.1003, 8:66-67), so that a given “control device” with a greater amount of remaining power would interact with the user (e.g., by performing an initial voice command and/or performing speech recognition on subsequent utterances), since such activity consumes a relatively high amount of power. *See* Ex.1004, ¶183-86.

Fourth, a POSITA would have been motivated to incorporate Jang’s use of

current context information into Rosenberger’s “control device” because it would have simply involved applying a known technique (i.e., Jang’s use of *current context* to make an arbitration decision) to an existing, similar device (i.e., Rosenberger’s “control device”) to achieve predicable results. Ex.1002, ¶224. A POSITA would have had at least a reasonable expectation of success, especially because Jang itself already discloses using such techniques in performing arbitration in a similar way as Rosenberger. *Id.*

XII. Lack of Secondary Considerations

Sonos is unaware of any secondary considerations of non-obviousness. Even if they existed, they would not overcome the compelling *prima facie* showing of unpatentability. *See* Ex.1002, ¶226-27.

XIII. Conclusion

For the foregoing reasons, Petitioner respectfully requests institution of IPR and cancellation of Challenged Claims 1-4, 7, 9-12, and 14-15.

Respectfully submitted,

LEE, SULLIVAN, SHEA & SMITH LLP

Dated: August 8, 2023

/Michael P. Boyea/

Michael P. Boyea, Reg. No. 70,248
LEE SULLIVAN SHEA & SMITH LLP
656 West Randolph Street, Suite 5W
Chicago, IL 60661
(312) 754-9602
boyea@ls3ip.com

Counsel for Petitioner Sonos, Inc.

CERTIFICATE OF WORD COUNT

The undersigned certifies that the foregoing PETITION FOR *INTER PARTES* REVIEW complies with the type volume limitation in 37 C.F.R. § 42.24(c)(1). According to the utilized word-processing system's word count, the petition—excluding the caption, table of contents, table of exhibits, mandatory notices, certificate of word count, and certificate of service—contains 12,980 words.

Dated: August 8, 2023

By: /Michael P. Boyea/

Michael P. Boyea, Reg. No. 70,248
LEE SULLIVAN SHEA & SMITH LLP
656 West Randolph Street, Suite 5W
Chicago, IL 60661
(312) 754-9602
boyea@ls3ip.com

Counsel for Petitioner Sonos, Inc.

CERTIFICATE OF SERVICE

The undersigned hereby confirms that the foregoing Petition for *Inter Partes* Review and associated exhibits were caused to be served on August 8, 2023 via overnight courier upon the following counsel of record for Patent Owner:

Shumaker & Sieffert, P.A.
1625 Radio Drive, Suite 300
Woodbury, MN 55125

Courtesy copies were also served via electronic mail on Patent Owner's counsel of record in a related district court litigation:

David A Nelson (davenelson@quinnemanuel.com)
Patrick D Curran (patrickcurran@quinnemanuel.com)
S. Alex Lasher (alexlasher@quinnemanuel.com)
Jeffrey S Gerchick (jeffgerchick@quinnemanuel.com)
Nina S. Tallon (ninatallon@quinnemanuel.com)

QUINN EMANUEL URQUHART & SULLIVAN, LLP

/Michael P. Boyea/

Michael P. Boyea, Reg. No. 70,248
LEE SULLIVAN SHEA & SMITH LLP
656 West Randolph Street, Suite 5W
Chicago, IL 60661
(312) 754-9602
boyea@ls3ip.com