

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

**DECLARATION OF DR. MICHAEL T. JOHNSON
IN SUPPORT OF PETITION FOR *INTER PARTES* REVIEW
OF U.S. PATENT NO. 9,632,748**

TABLE OF CONTENTS

I. INTRODUCTION	1
II. BACKGROUND AND QUALIFICATIONS	1
III. MATERIALS CONSIDERED	4
IV. LEGAL STANDARDS	5
A. Patent Invalidity	6
B. Anticipation	6
C. Obviousness	7
D. Claim Construction	10
V. OVERVIEW OF THE '748 PATENT	10
VI. PERSON OF ORDINARY SKILL IN THE ART	13
VII. CLAIM CONSTRUCTION	14
A. Relevant 1330 ITC Investigation Construed Terms	15
B. "A Spoken Audio Input... the Spoken Audio Input"	16
C. "A Command to Temporarily Refrain from Performing Speech Recognition"	18
VIII. PRIORITY DATE	21
IX. OVERVIEW OF THE PRIOR ART	21
A. Overview of Rosenberger	21
B. Overview of Jang	28
X. OVERVIEW OF THE GROUNDS	32

XI. THE CHALLENGED CLAIMS ARE UNPATENTABLE----- 32

A.	Ground 1: Rosenberger Anticipates Independent Claims 1, 7, & 11 -----	32
1.	Rosenberger Anticipates Claim 7 -----	32
a.	[7.0] Preamble -----	33
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-----	39
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-----	43
d.	[7.3] responsive to determining that the first computing device should perform speech recognition on the spoken audio input: -----	47
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-----	51
f.	[7.3(b)] perform speech recognition on the spoken audio input -----	57
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-----	59

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. -----	62
2.	Independent Claims 1 & 11 -----	65
B.	Ground 1: Rosenberger Anticipates the Challenged Dependent Claims (2-4, 9-10, 12, 14, 15) -----	67
1.	Dependent Claims 2, 9, 12 -----	68
2.	Dependent Claims 3, 10 -----	76
3.	Dependent Claims 4, 14, 15-----	78
C.	Ground 2: Rosenberger+Jang Renders Dependent Claims 2-3, 9-10, 12 Obvious -----	81
1.	Rosenberger+Jang Render Obvious [9.1]-[9.2] & [10.1]-----	82
2.	Motivation to Combine Rosenberger+Jang-----	92
XII. LACK OF SECONDARY CONSIDERATIONS-----		96
XIII. CONCLUSION-----		96

I, Dr. Michael T. Johnson, do hereby declare:

I. INTRODUCTION

1. I have been retained by Petitioner Sonos, Inc. (“Petitioner” or “Sonos”) as an independent expert in this matter before the United States Patent and Trademark Office (“USPTO”). I am not an employee of Sonos or any affiliate or subsidiary of Sonos.

2. I have been asked to provide my opinions in connection with this IPR, including my opinions concerning the patentability of claims 1-4, 7, 9-12, and 14-15 (the “Challenged Claims”) of U.S. Patent No. 9,632,748 (the “’748 Patent”), purportedly owned by Google LLC (“Patent Owner” or “Google”).

3. My opinions and the bases for my opinions are set forth below.

4. I am being compensated at my ordinary and customary rate of \$600 per hour for my work, plus reimbursement for any reasonable expenses. My compensation is based solely on the amount of time that I devote to activity related to this case and is in no way contingent on the nature of my findings, the presentation of my findings in testimony, or the outcome of this or any other proceeding. I have no other financial interest in this proceeding.

II. BACKGROUND AND QUALIFICATIONS

5. I am currently the Associate Dean of Undergraduate Education and Student Success in the College of Engineering at the University of Kentucky (UK),

a position which I have held since July 2023. In addition, I serve as the Department Chair and a Full Professor in the Department of Electrical and Computer Engineering at UK, having held that position since 2016. Prior to that, I was a faculty member at Marquette University in Milwaukee, Wisconsin from 2000 to 2016, during which time I also spent two separate sabbatical years as a Senior Visiting Professor at Tsinghua University in Beijing, China.

6. I received my PhD from Purdue University in West Lafayette, Indiana in 2000, with a doctoral dissertation focused on speech processing and automatic speech recognition. My undergraduate degrees are in Computer Science Engineering and in Electrical Engineering from LeTourneau University in Longview, TX, in 1989 and 1990, respectively, and I received a Master's degree in Electrical Engineering from the University of Texas at San Antonio in 1994. Prior to receiving my PhD, I worked in industry as a design engineer and engineering manager for more than 5 years, primarily in the telecommunications industry.

7. The focus of my research work is Speech and Audio Signal Processing, and I regularly conduct and publish original research articles within this field, with a publication record including 52 peer-reviewed journal articles and more than 150 total peer-reviewed publications over more than two decades. My specific areas of expertise include digital signal processing, speech enhancement and noise reduction, automatic speech recognition, speech production and kinematics, bioacoustics, and

machine learning. I have received multiple research grant awards for my work in these areas, including from the National Science Foundation and the National Institutes of Health.

8. As a faculty member, I regularly teach both undergraduate and graduate courses in electrical and computer engineering. This includes undergraduate level courses, such as Circuits, Signals and Systems, and Digital Signal Processing, as well as graduate level courses, such as Optimal and Adaptive Signal Processing and Speech Processing.

9. My professional service includes active membership in the Institute for Electrical and Electronic Engineers (IEEE), the IEEE Signal Processing Society, the Acoustical Society of America (ASA), and the Electrical and Computer Engineering Department Heads Association (ECEDHA). I served as the President of the Southeast ECE Department Heads Association (SECEDHA) in 2022-2023, and served as a member of the IEEE Signal Processing Society Speech and Language Technical Committee from 2014 to 2017. I also frequently serve as a reviewer for journals, conferences, and workshops in the area of speech and signal processing, including IEEE Transactions on Acoustics, Speech, and Signal Processing, IEEE Transactions on Signal Processing, the Journal of the Acoustical Society of America, Speech Communication, the International Conference on Acoustics, Speech, and Signal Processing, the IEEE Workshop on Automatic Speech Recognition, and

Interspeech. I have reviewed funding proposals for a number of public and private agencies and foundations, including for multiple programs and directorates at the National Science Foundation. In addition to these manuscript and grant proposal reviewing activities, I also currently serve as Area Editor for the journal Speech Communication in the area of speech enhancement.

10. During my academic career, I have been honored to receive a number of awards for my teaching and research work. Notable awards have included the College of Engineering Researcher of the Year at Marquette University in 2007, Milwaukee Young Engineer of the Year in 2006, and Electrical and Computer Engineering Eta Kappa Nu Teacher of the Year in 2005 and again in 2014. In 2008, my work applying speech recognition technology to animal vocalizations was featured in the Milwaukee Journal-Sentinel series “Groundbreaking Thinkers in Wisconsin.”

11. A detailed record of my professional qualifications, including a list of publications, awards, and professional activities, is set forth in my curriculum vitae attached as Appendix A.

III. MATERIALS CONSIDERED

12. In forming my opinions set forth herein, I considered the following materials, as well as my own experiences:

EXHIBIT	DESCRIPTION
1001	U.S. Patent No. 9,632,748 (“’748 Patent”)

EXHIBIT	DESCRIPTION
1003	U.S. Patent No. 8,340,975 (“Rosenberger”)
1004	U.S. Patent Publication 2013/0073293 (“Jang”)
1005	Order No. 14: Construing Terms of the Asserted Claims, International Trade Commission (ITC) Investigation No. 337-TA-1330
1006	Order No. 16, ITC Investigation No. 337-TA-1330
1007	Exhibit C, Infringement Claim Chart for U.S. Patent No. 9,632,748, ITC Investigation No. 337-TA-1330 (excerpted; redacted)
1008	Trial Testimony of Mark P. Mahon, ITC Investigation No. 337-TA-1330 (excerpted; redacted)
1009	Complainant’s Sixth Supplemental Objections and Responses to Respondent’s First Set of Interrogatories (Nos. 1-44), ITC Investigation No. 337-TA-1330 (excerpted; redacted)
1010	Complainant’s Initial Markman Brief, ITC Investigation No. 337-TA-1330
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 <i>Product Brief</i> 2009

IV. LEGAL STANDARDS

13. I have been instructed to follow the appropriate legal standards in providing my opinions. I am not a lawyer, but I have tried to apply the legal concepts as they have been explained to me.

A. Patent Invalidity

14. I understand that in an IPR, the petitioner has the burden of “proving a proposition of unpatentability by a preponderance of the evidence.” I understand that the “preponderance of the evidence” standard requires invalidating a challenged claim if, in view of the prior art, it is “more likely than not” that the challenged claim is unpatentable. I also understand that an independent claim can be found invalid, even though a dependent claim to which it refers is valid. I have considered these principles in forming my opinions contained in this declaration.

B. Anticipation

15. I understand that a prior art reference “anticipates” a challenged claim, and thus renders the claim invalid, if all limitations of the claim are disclosed in that prior art reference, either explicitly or inherently.

16. I understand that patents or printed publications that were available to the public before the effective filing date of the claimed invention can be used to invalidate a challenged claim as anticipated.

17. I understand that, once a challenged claim has been properly construed, the step in determining whether the challenged claim is anticipated by the prior art requires a comparison of the properly construed claim language to the prior art on a limitation-by-limitation basis.

C. Obviousness

18. I understand that even if a challenged claim is not anticipated, it can still be found invalid as obvious.

19. I understand that an obviousness analysis involves a number of considerations. I understand that the following factors must be evaluated to determine whether any challenged claim of a patent would have been obvious before the effective filing date of the claimed invention: (i) the scope and content of the prior art; (ii) the differences, if any, between the challenged claim and the prior art; (iii) the level of ordinary skill in the art before the effective filing date of the claimed invention; and (iv) additional considerations, if any, that indicate that the claimed invention was obvious or not obvious. I understand that these “additional considerations” are often referred to as “secondary considerations” or “objective indicia” of non-obviousness or obviousness.

20. I understand that the frame of reference when evaluating obviousness is what a hypothetical person of ordinary skill in the art (“POSITA”) pertinent to the claimed invention would have known before the effective filing date of the claimed invention. I understand that the hypothetical POSITA is presumed to have knowledge of all pertinent prior art references.

21. I understand that, when assessing the level of ordinary skill in the art, one is to consider factors such as (i) the type of problems encountered in the art, (ii)

the prior solutions to those problems, (iii) the rapidity at which innovations are made, (iv) the sophistication of the technology, and (v) the level of education of active workers in the relevant field.

22. I understand that a prior art reference may be a pertinent prior art reference (or “analogous art”) if it is in the same field of endeavor as the patent or if it is pertinent to the problem that the inventors were trying to solve.

23. I understand that an obviousness analysis can be based on either a single prior art reference or a combination of multiple prior art references. I understand that a single prior art reference can invalidate a challenged claim as obvious if some teaching, suggestion, or motivation would have led a POSITA to modify the single prior art reference to arrive at the challenged claim.

24. I understand that the following conditions may also indicate that the subject matter of a challenged claim was obvious at the time of the alleged invention:

- combining prior art elements according to known methods to yield predictable results;
- simple substitution of one known element for another to obtain predictable results;
- use of known techniques to improve similar devices (methods, or products) in the same way;
- applying a known technique to a known device (method, or product) ready for improvement to yield predictable results;
- “obvious to try”—choosing from a finite number of identified, predictable solutions, with a reasonable expectation of success;

- known work in one field of endeavor may prompt variations of it for use in either the same field or a different one based on design incentives or other market forces if the variations would have been predictable to one of ordinary skill in the art; and
- some teaching, suggestion, or motivation in the prior art that would have led one of ordinary skill to modify the prior art reference or to combine prior art reference teachings to arrive at the claimed invention.

25. I understand that “secondary considerations” must be considered as part of the obviousness analysis when present. I further understand that the secondary considerations may include: (i) a long-felt but unmet need in the prior art that was satisfied by the claimed invention; (ii) the failure of others; (iii) skepticism by experts; (iv) commercial success of a product covered by the patent, including licensing; (v) unexpected results achieved by the claimed invention; (vi) industry praise of the claimed invention; (vii) deliberate copying of the invention; and (viii) teaching away by others. I also understand that evidence of the independent and nearly simultaneous “invention” of the claimed subject matter by others is a secondary consideration supporting an obviousness determination and may support a conclusion that a claimed invention was within the knowledge of a POSITA before the effective filing date of the claimed invention.

D. Claim Construction

26. I understand that claim limitations must be viewed from the perspective of a POSITA to which the patent pertains before the effective filing date of the claimed invention.

27. I understand that a claim limitation is generally given the plain and ordinary meaning that a POSITA would ascribe to it when viewed in the context of the patent's claims, specification, and prosecution history.

28. I understand that a patent and its prosecution history are considered "intrinsic evidence" and are the most important sources for interpreting claim language in a patent. The prosecution history of related patents and applications can also be relevant.

29. I understand that sources extrinsic to a patent and its prosecution history (such as dictionary definitions and technical publications) may also be used to help interpret the claim language, but that such extrinsic sources cannot be used to contradict the unambiguous meaning of the claim language that is evident from the intrinsic evidence.

V. OVERVIEW OF THE '748 PATENT

30. The '748 Patent discloses that, "[i]n general, 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.” ’748 Patent, 2:36-41.

31. Specifically, 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 “arbitration.” *Id.*, 7:23-36, 22:51-56.

32. The ’748 Patent explains that the “best candidate” is determined during an initial phase based on a variety of factors prior to a second phase in which 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.

33. The '748 Patent explains that, in this initial phase, the “best candidate” is determined based on:

- “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,” such as “a location of the computing device (e.g., a GPS location),” “applications installed at the computing device,” “networks (e.g., wireless networks) available to the computing device,” “devices in proximity to the computing device (e.g., within 10 feet, within 100 feet, or other distance),” “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,” “data obtained from one or more sensors 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” (*id.*, 8:26-62).

34. The '748 Patent also explains that “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.

35. In terms of speech processing, the '748 Patent provides very little detail. For instance, the '748 Patent makes no mention of “voice activity detection” (VAD), makes no mention of “hotword detection,” and otherwise, provides little detail on other forms of “speech recognition.” As to the latter, 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 from the '748 Patent that the disclosure is not focused on any specific speech processing techniques, much less promotes one type of speech recognition technique to the exclusion of other types of techniques.

VI. PERSON OF ORDINARY SKILL IN THE ART

36. For purposes of this IPR only, I have been asked to assume that the Challenged Claims of the '748 Patent are entitled to a priority date of June 24, 2014.

37. I have been informed that, to assess the level of ordinary skill of a POSITA, I am to consider factors such as (i) the type of problems encountered in the art, (ii) the prior solutions to those problems, (iii) the rapidity at which innovations are made, (iv) the sophistication of the technology, and (v) the level of education of

active workers in the relevant field. To assess the appropriate level of ordinary skill of a POSITA for the '748 Patent, I considered these factors during my analysis of the '748 Patent and its file history, and I also considered my own experiences working and teaching in the field of speech processing.

38. Based on the foregoing, it is my opinion that a POSITA for 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.

39. Based on my education and experience, I am familiar with the level of knowledge and abilities of those meeting the above definition. I applied the above level of ordinary skill in forming my opinions set forth in this Declaration. It is worth noting, however, that my opinions would remain the same even if the assumed priority date, field of invention, and/or level of ordinary skill were slightly different.

VII. CLAIM CONSTRUCTION

40. I understand that a claim or claim element should only be construed when necessary to resolve a controversy over its interpretation. Aside from the terms discussed below, it is my opinion that the '748 Patent claim terms do not require

construction and can be afforded their plain and ordinary meaning for purposes of this IPR.

A. Relevant 1330 ITC Investigation Construed Terms

41. Google filed a complaint asserting the '748 Patent against Sonos at the International Trade Commission (ITC), which was assigned Investigation Number 337-TA-1330 (herein after "1330 ITC Investigation"). I was engaged by Sonos to offer expert opinions and testimony regarding, among other things, the subject matter of the '748 Patent.

42. In the 1330 ITC Investigation, the ALJ issued a *Markman* Order providing constructions for various terms of the '748 Patent, including the following constructions:

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"

¹ For clarity, I have italicized claim language throughout this Declaration.

Ex.1005, pp. 27-30, 32-34.

43. With regard to *perform/performing speech recognition*, after his *Markman* Order, I understand that the ALJ instructed the parties as follows in response to Sonos's motion for reconsideration/clarification:

As for "[perform/performing] speech recognition," in particular, this dispute appears to be over whether the term limits speech recognition to generating "text" (as Google contends) or may include generating "symbols" or other signals more broadly (as Sonos contends). Resolution of this dispute requires further evidentiary development regarding the term's plain and ordinary meaning.

Ex.1006, p. 1 (internal citation omitted).

44. I understand that the evidentiary record in the 1330 ITC Investigation is 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.

45. For purposes of this IPR, I have been asked to apply the ALJ's constructions from his *Markman* Order, subject to Google's broad interpretation of *a command to temporarily refrain from performing speech recognition* discussed below.

B. "A Spoken Audio Input... the Spoken Audio Input"

46. In the 1330 ITC 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 (transcript pages 383:3-385:9).

47. For purposes of this IPR, I have been asked to assume that Google’s interpretation of *a spoken audio input... the spoken audio input* is correct.

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

48. As noted above, in the 1330 ITC Investigation, the ALJ construed *a command to temporarily refrain from performing speech recognition* as “a command to temporarily refrain from performing **any and all** speech recognition.”

49. Despite this construction, in the 1330 ITC Investigation, Google interpreted *a command to temporarily refrain from performing speech recognition* broadly in several respects.

50. 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, pp.11-12 (document pages 376-77) (“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.**”), pp.12-13 (document pages 377-78) (“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.**”).

51. 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 (document page 16).

52. However, Google's infringement assertions involved interpreting *temporarily refrain* 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:

- “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 (transcript page 396:12-19).

- “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 (transcript page 397:8-20).

- “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 (transcript pages 394:17-395:7); *see also id.*, pp.44-45 (transcript pages 395:21-396:12).

- “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 (transcript page 414:3-23).

- “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 (transcript page 415:17-24).
- “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 (transcript page 392:13-25).

53. For purposes of this IPR, I have been asked to assume that Google’s interpretation of *a command to temporarily refrain from performing speech recognition* is correct.

VIII. PRIORITY DATE

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

55. For purposes of this IPR only, I have been asked to assume that the Challenged Claims of the '748 Patent are entitled to a priority date of June 24, 2014.

IX. OVERVIEW OF THE PRIOR ART²

A. Overview of Rosenberger

56. US Patent 8,340,975 (“Rosenberger”) issued on December 25, 2012 from a patent application filed on October 4, 2011. Ex.1003. Thus, Rosenberger qualifies as prior art to the '748 Patent, which has an earliest possible priority date of June 24, 2014.

57. 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.” Rosenberger, 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

² All emphasis set forth herein has been added unless noted otherwise.

to one another, sometimes several per room in larger rooms.” *Id.*, 1:53-2:1, 2:41-47.

58. 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”:

[I]t is inevitable that a given hands-free speech command would frequently be recognized simultaneously by two or more neighboring speech recognition modules. If multiple modules were to be triggered and audibly respond simultaneously, user confusion could occur. Worse, if two or more modules were to simultaneously process the same speech command, the target automated mechanism could be instructed by each triggered device to perform the same instruction with potentially undesirable or adverse effects. For example, “Lower Room Temperature Two Degrees” could be executed separately by several modules causing a thermostat to lower by multiples of the requested two degrees.

Id., 2:42-59; *see also id.*, 9:18-10:4.

59. While Rosenberger acknowledges several approaches to address the “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.

60. More 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” (often referred to simply as a “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.

61. Indeed, addressing this problem through “coordination” (or “arbitration” in the ’748 Patent’s parlance) by multiple “control devices” is a central aspect of Rosenberger:

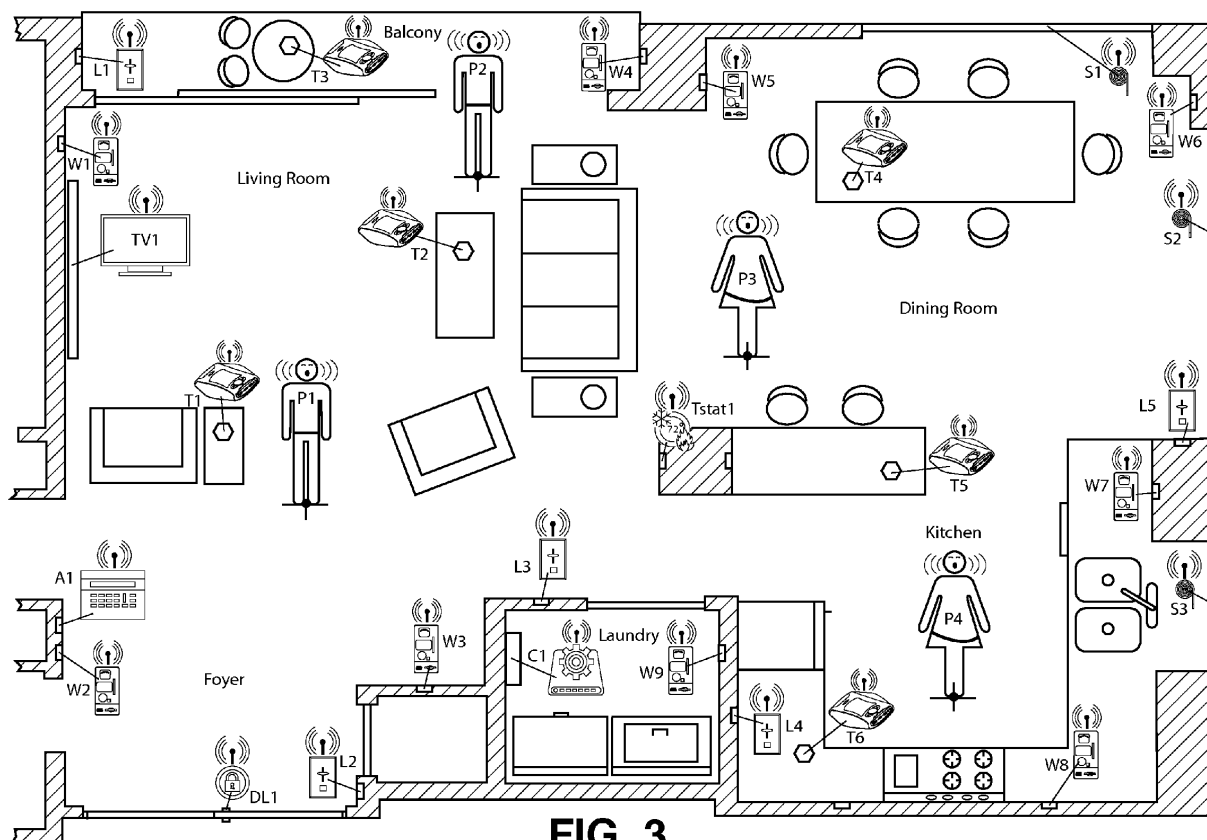
Another object of this invention is to provide a hands-free speech-recognition system that automatically determines which one of a plurality of interactive speech recognition control devices that have been simultaneously triggered by a speech command will further interact with the user as needed and control the system response.

Id., 3:57-62.

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.

63. 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.



64. 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.

65. Still with reference to FIG. 3, Rosenberger reiterates the “duplicate response” problem by providing the following example scenario:

[I]f person P3 stands in the dining room near the kitchen island and speaks a trigger phrase such as “Hello Voice Control”, table top devices T2, T4, T5 and T6 and in-wall devices W5 and W9 may all successfully “hear” and recognize this speech trigger. Thus, each of these devices, even those devices in different rooms (e.g. T2, T6 and W9) will want to respond to the same speech command. If they all were to respond with “How may I help you?” person P3 could become confused. Worse, if person P3 would proceed to say the speech command “Toggle All Shades”, intending to close currently-open window shades [S]1, S2 and S3, control devices T2, T4, T5, T6, W5 and W9 would all instruct system controller C1 to toggle the shades, resulting in the shades going up and down 6 times and ultimately remaining open.

Id., 9:57-10:4.

66. 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.

67. 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.”

68. 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).

69. 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.

70. The “Coordinating Controller” methods follow what is known as an “optimistic model” or what is known as a “pessimistic model.” In an “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. In a “pessimistic model,” after making such a decision, the “coordinating controller” may transmit a single “proceed” (or “go”) message to the “control device” that the “coordinating controller” determined should handle the voice input (assuming such “control device” was not the “coordinating controller” itself). *Id.*, 12:53-56, 13:7-10, 14:18-21.

71. 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.” *See, e.g., id.*, 2:1-9, 3:30-35, 8:16-34, FIGs. 4 (“Speech Trigger or Command Recognized”), 5A (same), FIG. 6 (same). And Rosenberger 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, e.g., 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. Overview of Jang

72. U.S. Patent Publication 2013/0073293 (“Jang”) was published on March 21, 2013 from a patent application filed on September 20, 2011. Ex.1004. Thus, Jang qualifies as prior art to the ’748 Patent, which has an earliest possible priority date of June 24, 2014.

73. Jang discloses a system of “electronic devices” (*see, e.g., Jang*, ¶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 given electronic device. *Id.*, ¶110.

74. Jang explains that the “voice recognition unit” enables each electronic device to recognize a user’s voice command and there can be scenarios where multiple electronic devices recognize the same voice command and “generate[] the same output in response to the user’s voice command.” *Id.*, ¶120 (“The TV 100 and the tablet PC 10a are driven under the same operating system (OS) and have the same voice recognition module for recognizing the user’s voice commands. Accordingly, the TV 100 and the tablet PC 10a ***generates the same output in response to the user’s voice command.***”).

75. 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 in the system making an arbitration determination individually (*see, e.g., id.*, ¶148-52, FIG. 9) or a “managing” device making an arbitration determination on behalf of the electronic devices in the system.

76. 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 and interact with the user. *Id.*, ¶¶130-132, 135-140, 149-151.

77. 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.

78. 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.”

79. 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.

80. 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.

81. 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. OVERVIEW OF THE GROUNDS

82. As explained below, it is my opinion that each of the Challenged Claims –claims 1-4, 7, 9-12, and 14-15– is anticipated by Rosenberger. This is **Ground 1**.

83. As also explained below, it is my opinion that Rosenberger in view of Jang renders dependent claims 2-3, 9-10, and 12 obvious. This is **Ground 2**.

XI. THE CHALLENGED CLAIMS ARE UNPATENTABLE

84. For the reasons I explain in greater detail below, it is my opinion that, as of the assumed priority date of the '748 Patent (i.e., June 24, 2014), a POSITA would have considered the Challenged Claims to be unpatentable as anticipated and obvious over the prior art discussed in this Declaration.

A. Ground 1: Rosenberger Anticipates Independent Claims 1, 7, & 11

85. As explained in detail below, it is my opinion that Rosenberger anticipates independent claims 1, 7, and 11. Below, I provide my analysis of independent claim 7 first and then discuss independent claims 1 and 11.

1. Rosenberger Anticipates Claim 7

86. As explained below, it is my opinion that Rosenberger discloses each element of claim 7 and thus, anticipates claim 7.

87. Claim 7 recites in full (with bracketed numerals inserted for ease of reference):

[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:*

[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;*

[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*

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

[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;*

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

[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*

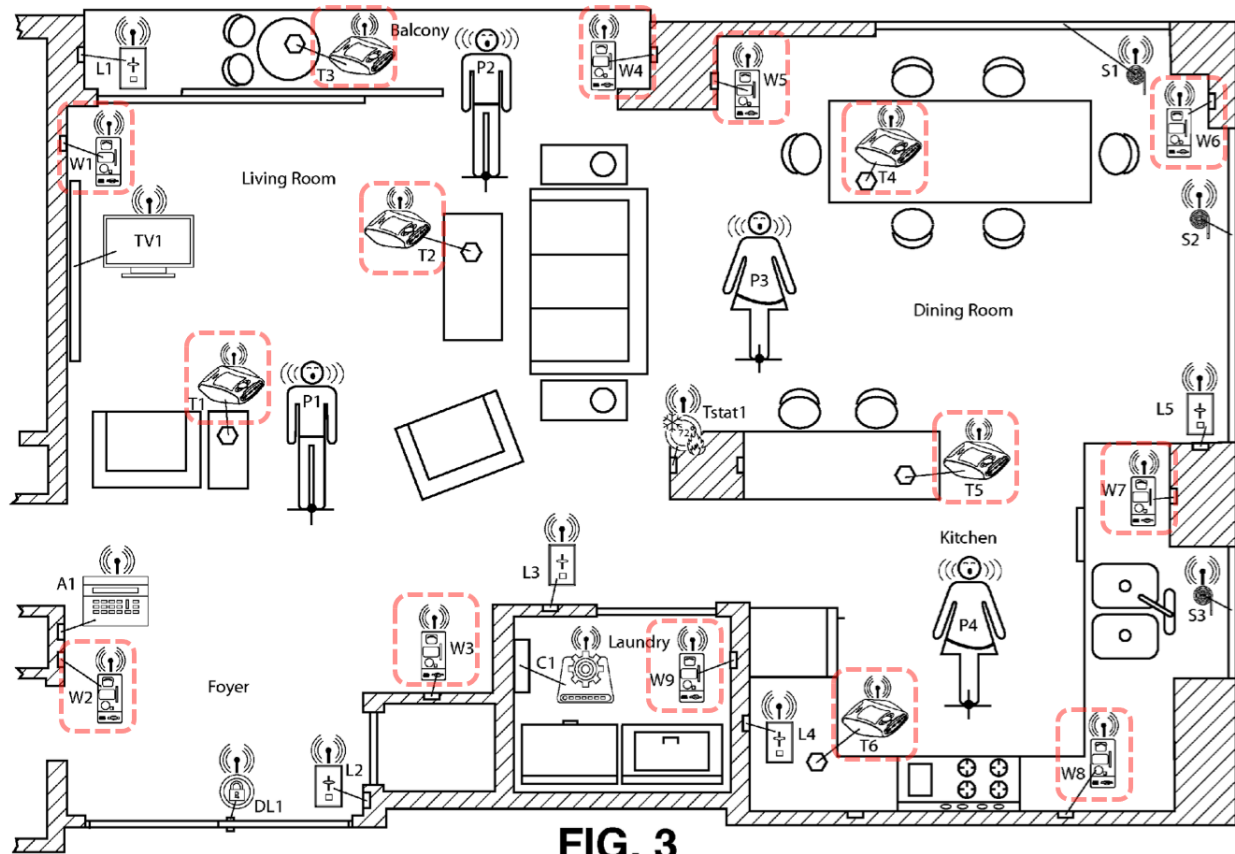
[7.4(d)] *responsive to determining that the spoken audio input includes the predetermined spoken audio command, execute an action associated with the predetermined spoken audio command.*

a. [7.0] Preamble

88. The preamble of claim 7 recites **[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 perform the recited functions.*** To the extent

that the preamble is limiting, it is my opinion that Rosenberger discloses the elements of the preamble.

89. In this regard, 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**....” Rosenberger, 4:54-60. An example of such a system is illustrated in Rosenberger’s FIG. 3 (reproduced below), which shows “a residential floor-plan... containing a home-automation system in which 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.

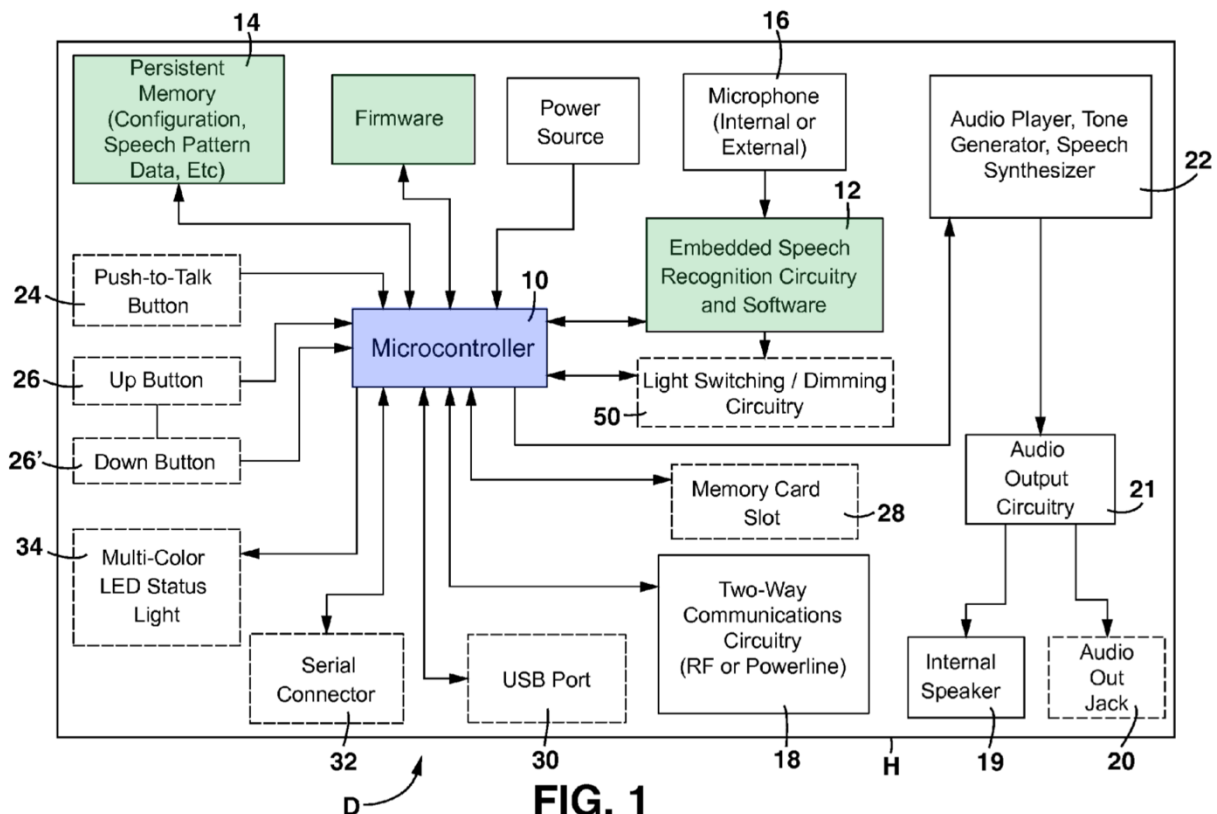


Id., FIG. 3 (annotations added); *see also, e.g., id.*, 3:63-4:5:

In accordance with one aspect of the invention, there is provided an interactive, self-contained, **speech recognition control device** that is adapted for use in an automated building control system in which a plurality of such devices are positioned at different locations throughout a building space to enable a system user to control the actions of the system, via speech commands only, from various locations within the building space. Such device also enables the system to audibly interact (e.g. communicate status information) with the system user.

90. With reference to FIG. 1 (reproduced 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 the like” and “firmware.” *Id.*, 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”).

91. 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.

92. In my opinion, a POSITA would have understood before the effective filing date of the ’748 Patent that Rosenberger’s “microcontroller” amounted to or otherwise included *at least one processor*. Indeed, this was common knowledge at the time.

93. For instance, my opinion that a POSITA would have understood that Rosenberger’s “microcontroller” included *at least one processor* is confirmed by prior art cited on the face of the ’748 Patent. For example, U.S. Patent Publication 2010/0088100 states “[p]rocessing circuitry 62 may be based on ***one or more processors such as*** microprocessors, ***microcontrollers***, digital signal processors, application specific integrated circuits, and other suitable integrated circuits.”

Ex.1011, ¶55. Because this reference is cited on the face of the '748 Patent, I understand that it constitutes part of the '748 Patent's intrinsic record.

94. Further, my opinion that a POSITA would have understood that Rosenberger's "microcontroller" amounted to or otherwise included *at least one processor* is confirmed by several technical dictionary definitions. *See, e.g.*, Ex.1012 (explaining that a "microcontroller" contains "a *processor*"); Ex.1013 (defining "microcontroller" as "a microprocessor that controls some or all of the functions of an electronic device (as a home appliance) or system"); Ex.1014 (defining "microcontroller" as "a microprocessor designed specifically for controlling equipment."); Ex.1015 (defining "microcontroller" as "[a] special-purpose, single-chip computer" that includes a "central *processing unit* (CPU)...").

95. Lastly, the "Sensory NLP-5x" module referenced by Rosenberger (*see, e.g.*, Ex.1003, 10:47-52) was a "natural language *processor*" based on a 16-bit "general purpose" digital signal *processor* (DSP). *See, e.g.*, Ex.1016, pp.1-2 (illustrating block diagram of NLP-5x module).

96. Consequently, it is my opinion that a POSITA would have understood that Rosenberger's 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 Rosenberger.

97. As I explained before, Rosenberger discloses “Coordinating Controller” methods for solving the “duplicate response” problem 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* ¶69. As explained below, a “control device” serving as the “coordinating controller” amounts to the *first computing device* of claim 7.

98. Thus, it is my opinion that Rosenberger discloses a *first computing device comprising: at least one processor; and at least one module operable by the at least one processor to perform the recited functions*, as discussed in further detail below.

- 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**

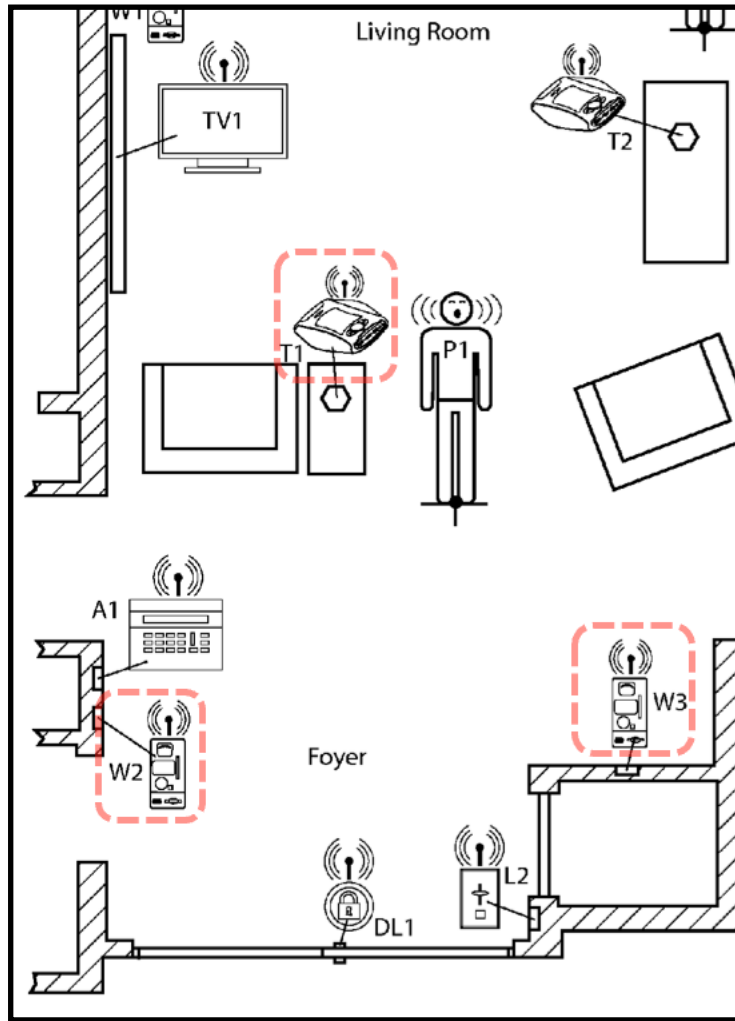
99. Claim 7 requires that the *first computing device* be operable to [7.1] *receive a spoken audio input that is also received by one or more additional computing devices in physical proximity to the first computing device....*

100. Rosenberger discloses this element. In this regard, Rosenberger discloses that a “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*).

101. For instance, Rosenberger 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.” Rosenberger, 10:5-14; *see also, e.g., id.*, 3:63-4:5. Rosenberger explains “[b]y liberally positioning the devices within a given space (*sometimes several per room*), ‘comfortable’ interaction is achieved—there is no need to raise one’s voice to have a speech command recognized by the control device, nor the need to strain to hear the audio responses from such device.” *Id.*, 6:34-39.

102. Rosenberger further discloses that two or more of the plurality devices may simultaneously receive a *spoken audio input*. For example, with reference to FIG. 3, Rosenberger 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 (annotations added & cropped).

103. Rosenberger expressly discloses receiving *a spoken audio input* under Google’s broad interpretation of the term. 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.” *See, e.g., id.*, 9:57-63, 10:19-23, 14:26-29. Rosenberger also discloses embodiments of a “control device” receiving

a more general “speech command” contained in a first utterance (e.g., “**Watch** Cable Channel CNN” or “**Call** the Police”). *See, e.g., id.*, 7:19-28, 8:44-49.

104. 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, e.g., 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*.”).

105. As one illustrative example, in Rosenberger, after receiving an initial utterance containing a “speech trigger,” a “control device” may prompt the user “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, e.g., id.*, 14:26-45. As another illustrative example, in Rosenberger, after receiving

a more general “speech command” in an initial utterance (e.g., “**Call** the Police”), a “control device” may prompt the user “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, e.g., id.*, 3:3-5, 8:44-46.

106. Thus, it is my opinion that Rosenberger discloses a *first computing device* operable to [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.*

- 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**

107. Claim 7 requires the *first computing device* to be operable to [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....*

108. Rosenberger discloses this element. For instance, Rosenberger discloses that a “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*.

109. 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 –specifically, 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. More 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.” Rosenberger, 10:34-38; *see also*, *e.g.*, *id.*, FIG. 4. Rosenberger generally describes “coordination” to involve the following steps for the “Coordinating Controller” methods.

110. **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.” *Id.*, 4:6-16:

[T]he control device of the invention includes... (b) a programmable microcontroller... comprising embedded software and circuitry serving to (i) recognize from the electrical signals when a speech command

from an expected vocabulary or grammar has likely been spoken, and (ii) to act thereon to first produce ***a weighted signal*** representing the quality of a just-recognized speech command....

Id., 11:42-51 (“In the event the microcontroller was triggered hands-free through its recognition of a spoken trigger or command phrase, the device calculates a weighted signal as described above.”), FIG. 4, 12:28-48, FIG. 5A.

111. In fact, Rosenberger explains that all of the coordination methods by which Rosenberger seeks to resolve the aforementioned “duplicate response” problem 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.

112. ***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.

113. ***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 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.

114. **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.

115. 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, e.g., 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 ***T2 receives the highest quality speech command***, both T3 and W4 will receive a weighted signal that is higher than their own respective weighted signals during the waiting period and each will immediately terminate further processing. 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***.

116. In this way, Rosenberger’s “coordination” decision is a determination *whether the first computing device should perform speech recognition on the spoken audio input.*

117. Thus, it is my opinion that Rosenberger discloses a *first computing device operable to [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.*

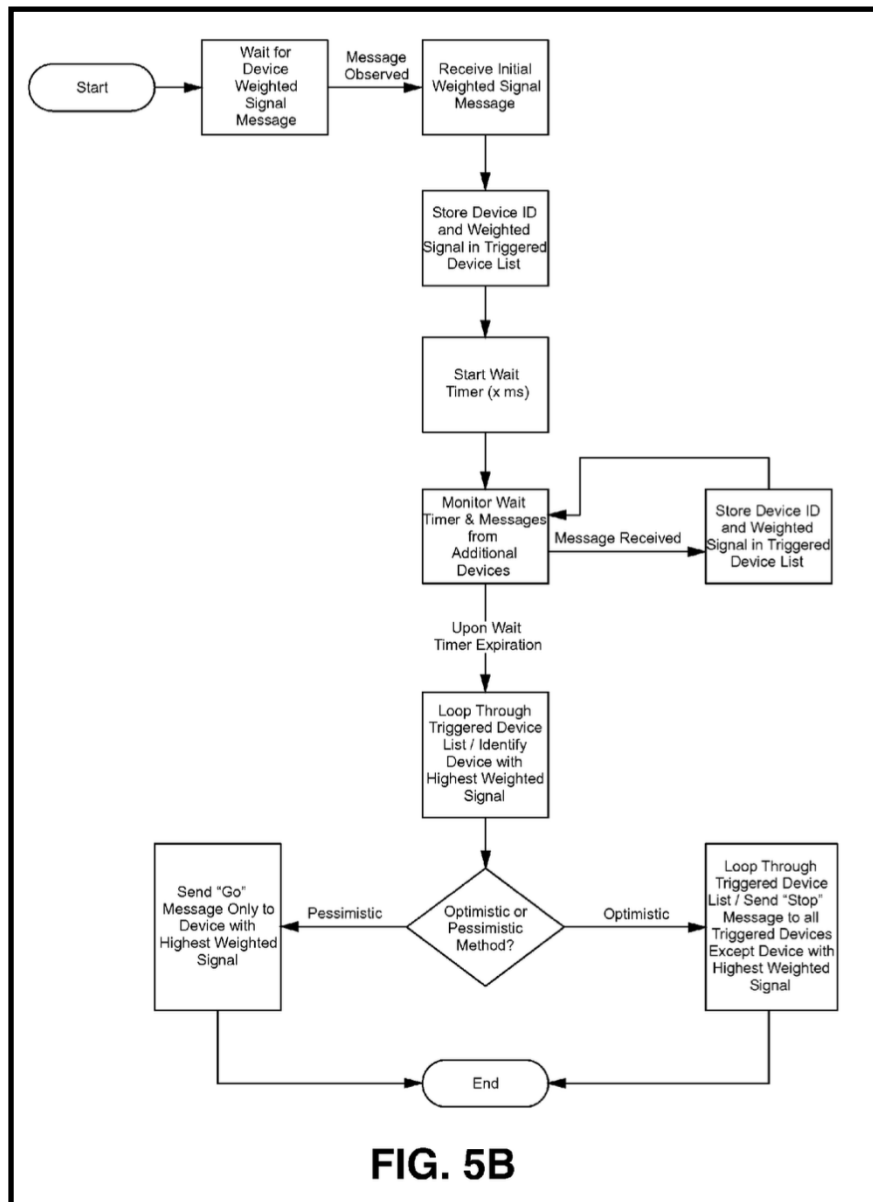
d. [7.3] responsive to determining that the first computing device should perform speech recognition on the spoken audio input:

118. Claim 7 requires the *first computing device* to be operable such that, ***[7.3] responsive to determining that the first computing device should perform***

speech recognition on the spoken audio input, it can perform the functions of elements [7.3(a)-(d)].

119. Rosenberger discloses this element. 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)].

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



121. 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.,* Rosenberger, 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”).

122. 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:

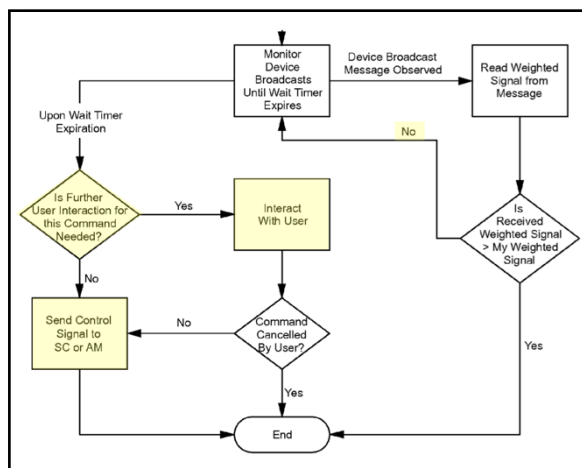


FIG. 4 (annotations added & cropped)

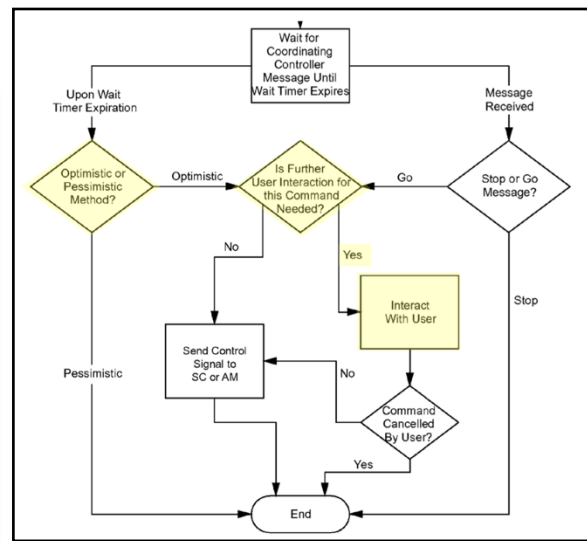


FIG. 5B (annotations added & cropped)

123. As explained, 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* ¶¶114-16.

124. Thus, it is my opinion that Rosenberger discloses a *first computing device* operable to [7.3] *responsive to determining that the first computing device should perform speech recognition on the spoken audio input....*

- 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**

125. Claim 7 requires the *first computing device* to be operable such that, [7.3] *responsive to determining that the first computing device should perform speech recognition on the spoken audio input*, it can [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....*

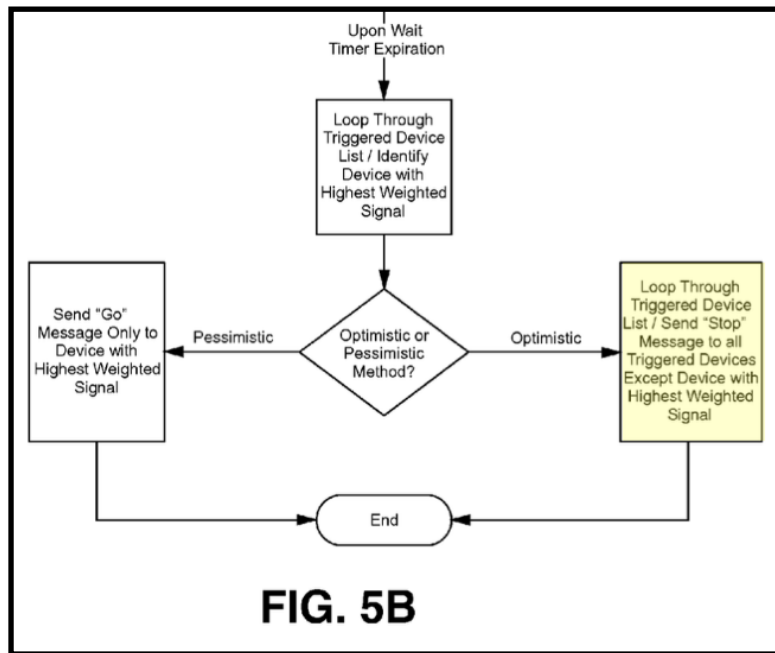
126. Rosenberger discloses this element, especially under Google’s broad interpretation of element [7.3(a)]. *Supra* ¶48-53. For instance, Rosenberger discloses that a “control device” serving as the “coordinating controller” (*first computing device*) is operable such that, responsive to it determining that it should perform speech recognition on *the spoken audio input*, the “coordinating controller” 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*).

127. In this regard, as I explained before, Rosenberger discloses “Coordinating Controller” methods for solving the “duplicate response” problem

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 a user’s utterances. *Supra* ¶¶66-70.

128. As I also explained, Rosenberger discloses a “Coordinating Controller” method for solving the “duplicate response” problem that follows an “optimistic model.” *Id.*; *see also, e.g.*, Rosenberger, 11:27-36, 12:28-52. For Rosenberger’s disclosed “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.

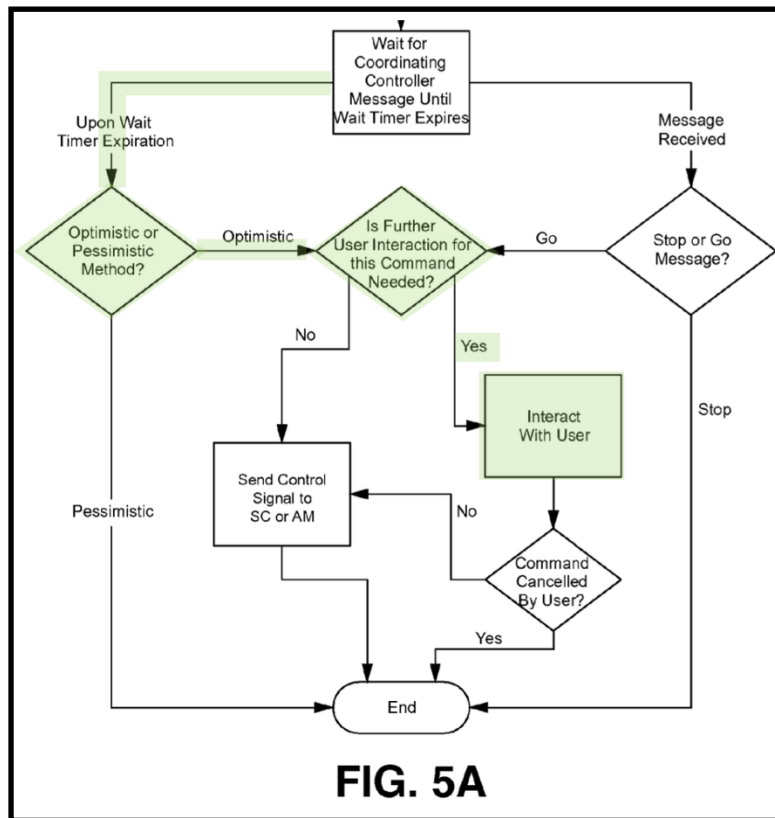
129. FIG. 5B illustrates the logic of the “coordinating controller” for performing “coordination,” including the “coordinating controller” sending “stop” messages to “control devices” that lost “coordination” (i.e., those that do not have the highest WS value) according to the “optimistic” model/method:



Id., FIG. 5B (annotations added & cropped).

130. 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.

131. In turn, FIG. 5A illustrates a control device’s logic for interacting with a user in accordance with Rosenberger’s “Coordinating Controller” methods. In this regard, the below annotated version of 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., FIG. 5A (annotations added & cropped); *see also, e.g., id.*, 5:51-57, 12:28-52, 14:10-45.

132. Moreover, the below annotated version of FIG. 5A illustrates the “optimistic model” and what happens at a “control device” upon receiving a “stop” message from the “coordinating controller.” In particular, 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).

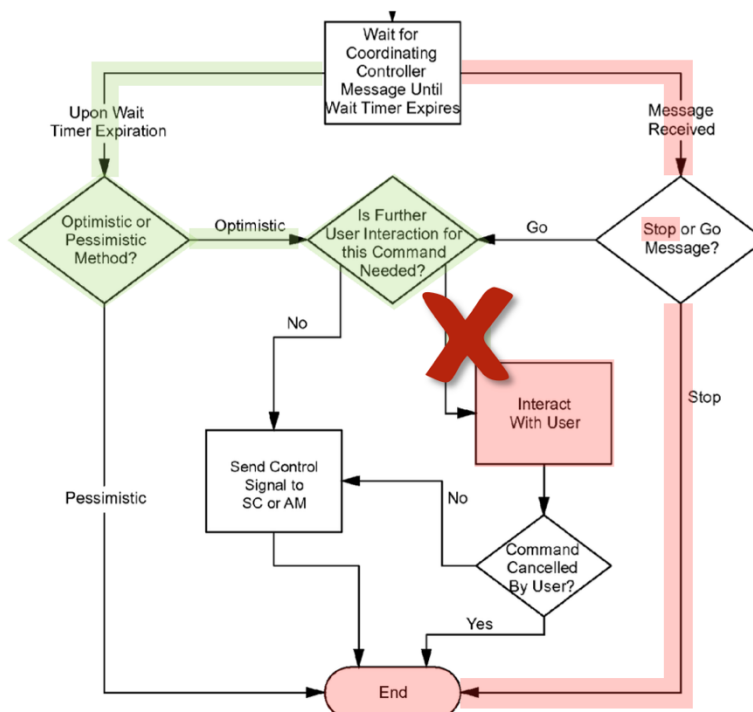


FIG. 5A

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

133. As a first illustrative example, the “coordination” could have been initiated based on a “trigger phrase” (e.g., “Hello Voice Control”) contained in an initial utterance. 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”). *See, e.g., Ex.1008*, p.56 (transcript page 415:17-24) (Dr. Mahon testifying “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 (transcript page 392:13-25):

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*.

134. As a second illustrative example, the “coordination” could have been initiated based on a more general “speech command” (e.g., “Watch”) contained in an initial utterance (e.g., “*Watch* Cable Channel CNN”). 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”). *See, e.g.*, Ex.1008, pp.66-67 (transcript pages 1158:16-1159:20), pp.68-70 (transcript pages 1160:6-1162:3) (Google’s ITC expert for the ’748 Patent 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* ¶52 (collecting testimony from Google’s ITC expert where he contended 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]*).

135. In this way, Rosenberger’s “stop” message constitutes *a command to temporarily refrain from performing speech recognition*, at least under Google’s broad interpretation (*supra* ¶¶48-53) 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.

136. Thus, it is my opinion that Rosenberger discloses a *first computing device* operable to, *responsive to determining that the first computing device should perform speech recognition on the spoken audio input*, [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*.

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

137. Claim 7 requires the *first computing device* to be operable such that, [7.3] *responsive to determining that the first computing device should perform speech recognition on the spoken audio input*, it can [7.3(b)] *perform speech recognition on the spoken audio input....*

138. Rosenberger discloses this element. For instance, Rosenberger discloses that a “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*.

139. In this regard, as I explained before, the selected “control device” (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, e.g., Rosenberger, 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 T2 receives the highest quality speech command, both ***T3 and W4*** will receive a weighted signal that is higher than their own respective weighted signals during the waiting period and ***each will immediately terminate further processing***. 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*.

140. Moreover, Rosenberger is replete with disclosure of each “control device” being configured to perform speech recognition on spoken audio input. 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.’” *Id.*, 6:21-25; *see also, e.g., id.*, 2:1-15, 4:6-24, 4:33-44 (“A preferred characteristic of the control device of the invention is that, within a predetermined time interval after it has produced the weighted signal in response to *recognizing a given speech command*,... the [control device’s] microcontroller proceeds in handling the given speech command.”), 7:5-31, 10:47-52.

141. Thus, it is my opinion that Rosenberger discloses a *first computing device* operable to, *responsive to determining that the first computing device should perform speech recognition on the spoken audio input, [7.3(b)] perform speech recognition on the spoken audio input.*

- 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**

142. Claim 7 requires the *first computing device* to be operable such that, **[7.3]** *responsive to determining that the first computing device should perform*

speech recognition on the spoken audio input, it can [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....

143. Rosenberger discloses this element. For instance, Rosenberger discloses that a “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.

144. For instance, as I explained before, 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. *See, e.g., Rosenberger, 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 T2 receives the highest quality speech command, both T3 and W4 will receive a weighted signal that is higher than their own respective weighted signals during the waiting period and each will immediately terminate further processing. 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***.

145. 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, e.g., id.*, 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 (e.g., the above-noted Model LD3320 from IC Route and Model NLP-5x from Sensory, Inc.) 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.

146. Thus, it is my opinion that Rosenberger discloses a *first computing device operable to, responsive to determining that the first computing device should perform speech recognition on the spoken audio input, [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.*

- 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.**

147. Claim 7 requires the *first computing device* to be operable such that, **[7.3]** *responsive to determining that the first computing device should perform speech recognition on the spoken audio input*, it can **[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.*

148. Rosenberger discloses this element. For instance, Rosenberger discloses that a “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.

149. For instance, as I explained before, the selected “control device” (e.g., the “control device” serving as the “coordinating controller”) may perform speech recognition on subsequent user utterance, determine that such an utterance includes a particular speech command, and then execute an action associated with that particular speech command. *See, e.g., Rosenberger, 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 T2 receives the highest quality speech command, both T3 and W4 will receive a weighted signal that is higher than their own respective weighted signals during the waiting period and each will immediately terminate further processing. 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***.

150. 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., id.*, 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).

151. 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 and executing an action associated with the particular speech command:

Since speech recognition device W8 is closest to the system user and there are no obstructions between the user and W8, it is likely that W8 will have the highest weighted signal and will therefore be chosen by the coordinating controller CC to handle the speech command. Devices T4, T5, and W7 will all disregard the just-recognized speech command and resume listening for new speech commands. W8 will immediately interact with person P4 (e.g. responding with “How may I help you?”). Once user interaction is completed (*e.g. the user says “Lower thermostat two degrees”*), device *W8 will send appropriate* control signals to a system controller SC (e.g. C1). The system controller will then send compatible *instructions to thermostat Tstat1 telling it to lower its set point by two degrees.*

Id., 14:32-45.

152. Rosenberger 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.

153. Thus, it is my opinion that Rosenberger discloses a *first computing device* operable to, *responsive to determining that the first computing device should perform speech recognition on the spoken audio input, [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.*

2. Independent Claims 1 & 11

154. As shown below, (i) independent claim 1 recites *a method* corresponding to the recited functions of independent claim 7 that the *first computing device* is operable to perform and (ii) 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:

Claim 7	Claim 1	Claim 11
[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:	[1.0] A method comprising	[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:
[7.1] receive a spoken audio input that is also received by one	[1.1] receiving, by a first computing device of a plurality	[11.1] receive a spoken audio input that is also received by one

Claim 7	Claim 1	Claim 11
or more additional computing devices in physical proximity to the first computing device;	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;	or more additional computing devices in physical proximity to the first computing device;
[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	[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	[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
[7.3] responsive to determining that the first computing device should perform speech recognition on the spoken audio input:	[1.3] responsive to determining that the first computing device should perform speech recognition on the spoken audio input:	[11.3] responsive to determining that the first computing device should perform speech recognition on the spoken audio input:
[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;	[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;	[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;
[7.3(b)] perform speech recognition on the spoken audio input;	[1.3(b)] performing, by the first computing device, speech recognition on the spoken audio input;	[11.3(b)] perform speech recognition on the spoken audio input;
[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	[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	[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
[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.	[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.	[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.

155. As is evident, (i) [1.0], [1.1], [1.2], [1.3], [1.3(a)], [1.3(b)], [1.3(c)] and [1.3(d)] map to [7.0], [7.1], [7.2], [7.3], [7.3(a)], [7.3(b)], [7.3(c)] and [7.3(d)], respectively, and (ii) [11.0], [11.1], [11.2], [11.3], [11.3(a)], [11.3(b)], [11.3(c)] and [11.3(d)] map to [7.0], [7.1], [7.2], [7.3], [7.3(a)], [7.3(b)], [7.3(c)] and [7.3(d)], respectively.

156. Notably, as discussed above for element [7.0], including elements [7.0(a)] and [7.0(b)], Rosenberger discloses a *non-transitory computer-readable storage medium encoded with instructions that, when executed, cause at least one processor of a first computing device to carry out the functions of claim 11*. *Supra* ¶¶88-98; *see also, e.g.*, Rosenberger, 4:6-14 (“[T]he control device of the invention includes... (b) a programmable microcontroller, operatively coupled to the microphone and comprising embedded software and circuitry”), 6:62-67, 8:8-12 (explaining “device D” has “firmware” loaded into it), 12:28-31 (disclosing that “control devices” include “programming logic”).

157. Thus, Rosenberger discloses each of the elements of claims 1 and 11 for the same reasons that Rosenberger discloses each of the elements of claim 7.

B. Ground 1: Rosenberger Anticipates the Challenged Dependent Claims (2-4, 9-10, 12, 14, 15)

158. As noted above, the Challenged Claims include (i) claims 2-4 that depend from independent claim 1, (ii) claims 9-10 and 15 that depend from

independent claim 7, and (iii) claims 12 and 14 that depend from independent claim 11.

159. As I explained above, it is my opinion that Rosenberger anticipates each of independent claims 1, 7, and 11. Below, I address the additional elements of the dependent Challenged Claims and explain why Rosenberger anticipates the dependent Challenged Claims. Since I used independent claim 7 as the representative independent claim, I will use its dependent claims as the representative dependent claims below.

1. Dependent Claims 2, 9, 12

160. Claim 9 depends from claim 7 and further requires:

[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.*

161. Rosenberger discloses these additional elements.

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

163. As an initial matter, the ’748 Patent defines a device’s *current context* extremely broadly. In this regard, dependent claim 10 that depends from claim 9

provides a long list of possible *current contexts* and recites *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:*

[1] a location of the first computing device, [2] a current time as defined by the first computing device, [3] one or more applications installed at the first computing device, [4] one or more applications currently executing at the first computing device, [5] one or more networks available to the first computing device, [6] one or more other computing devices in proximity to the first computing device, [7] an operating mode of the first computing device, [8] an ambient temperature of the location of the first computing device, [9] an ambient noise level of the location of the first computing device, [10] an ambient light level of the location of the first computing device, [11] a movement of the first computing device, [12] a name of a user of the first computing device, [13] a user identification (UID) of the user of the first computing device, [14] a social media network service account associated with the user of the first computing device, [15] one or more calendars associated with the user of the first computing device, or [16] one or more social relationships of the user of the first computing device.

164. Based on my understanding of dependent claims in patents, dependent 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.

165. The '748 specification describes a device's "context" even more broadly:

- "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 collections of data (e.g., a text data structure, a numeric data structure, or other data structure) that represents a location of the computing device (e.g., a GPS location); information indicating a time as determined by the computing device; information indicating one or more

applications installed at the computing device; information indicating one or more applications currently executing at the computing device; information indicating one or more networks (e.g., wireless networks) available to the computing device; data that represents one or more other computing devices in proximity to the computing device (e.g., within 10 feet, within 100 feet, or other distance); 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, ***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.***” ’748 Patent, 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 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.

³ The ’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.

166. Returning to Rosenberger, Rosenberger discloses determining multiple *current contexts* as that phrase is used 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 I discuss these examples below with reference to element [9.2].

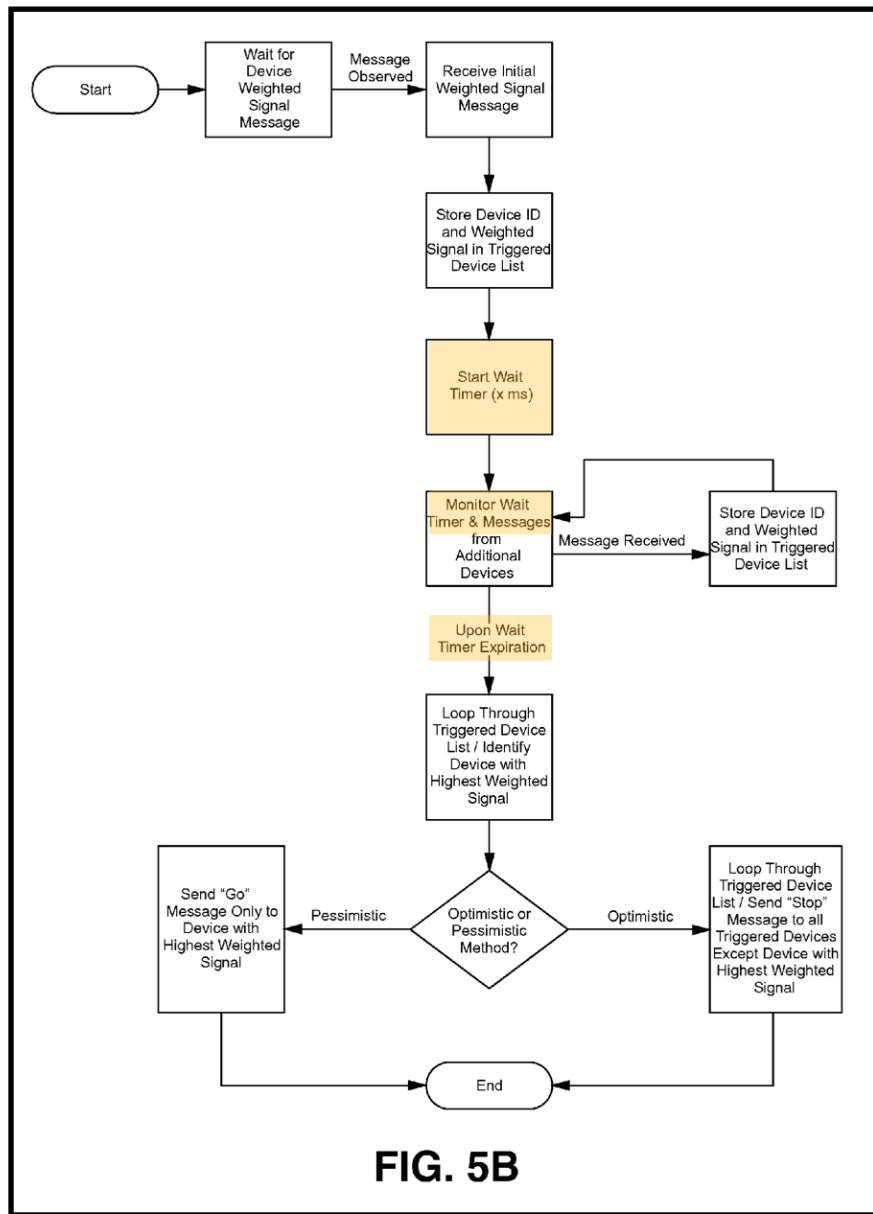
167. Thus, it is my opinion that Rosenberger discloses a *first computing device* operable to [9.1] *determine a current context of the first computing device*.

168. **Second**, Rosenberger discloses that a "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 context* determined for element [9.1], thereby satisfying element [9.2].

169. As a first example, Rosenberger discloses that the "control device" serving as the "coordinating controller" (*first computing device*) determines whether

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.

itself or another “control device” is the “coordination winner” (and thus, should perform speech recognition) based on a current time as defined by the “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:



Rosenberger, FIG. 5B (annotated); *see also, e.g., 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 (“In the 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.”), 13:20-29.

170. As a second example, Rosenberger discloses that the “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, e.g., 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, e.g., 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 (illustrating the same), 12:57-13:10; *see also* element [7.2]. 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* ’748 Patent, 17:10-40.

171. Thus, it is my opinion that Rosenberger discloses a *first computing device* operable to [9.2] *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.

172. Consequently, it is my opinion that Rosenberger discloses a *first computing device* as required by dependent claim 9.

173. Because dependent claims 2 and 12 recite similar subject matter as dependent claim 9, it is also my opinion that Rosenberger anticipates these dependent claims for the same reasons as it anticipates claim 9.

174. In this regard, claim 2 depends from independent claim 1 and recites: *[t]he method of claim 1, further comprising: [2.1] determining, by the first computing device, a current context of the first computing device, [2.2] wherein determining whether the first computing device should perform speech recognition on the spoken audio input is further based on the current context of the first computing device.* Likewise, claim 12 depends from independent claim 11 and recites: *[t]he non-transitory computer-readable storage medium of claim 1, further encoded with instructions that, when executed, cause the at least one processor to: [12.1] determine a current context of the first computing device, [12.2] wherein the instructions cause 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.*

175. As is evident, (i) [2.1] and [2.2] map to [9.1] and [9.2], respectively, and (ii) [12.1] and [12.2] map to [9.1] and [9.2], respectively. Thus, Rosenberger discloses each of the elements of claims 2 and 12 for the same reasons that Rosenberger discloses each of the elements of claim 9.

2. Dependent Claims 3, 10

176. Claim 10 depends from claim 9 and further requires:

[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.

177. Rosenberger discloses this additional element. For instance, as explained above in connection with dependent claim 9, one of the current contexts that the “coordinating controller” determines is a current time as defined by itself.

178. Thus, it is my opinion that Rosenberger discloses a *first computing device* operable to [10.1] *determine one or more of: ... a current time as defined by the first computing device.*

179. Because dependent claim 3 recites similar subject matter as dependent claim 10, it is also my opinion that Rosenberger anticipates claim 3 for the same reasons as it anticipates claim 10.

180. In this regard, claim 3 depends from independent claim 2 and recites:

[3.1] *wherein determining the current context of the first computing device comprises determining 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 sec-end⁵ 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, or one or more social relationships of the user of the first computing device.*

181. As is evident, [3.1] maps to [10.1]. Thus, Rosenberger discloses [3.1] for the same reasons that Rosenberger discloses [10.1].

⁵ For purposes of this IPR only, I have been asked to assume that the phrase *the sec-end first computing device* instead reads as “the first computing device.”

3. Dependent Claims 4, 14, 15

182. Claim 15 depends from claim 7 and further requires:

[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.*

183. Rosenberger discloses these additional elements.

184. **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].

185. 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.” Rosenberger, 10:38-41; *see also, e.g., 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.

186. Rosenberger further discloses, with reference to Figs. 4, 5A, 5B, and 6, 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.

187. Thus, it is my opinion that Rosenberger discloses [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.*

188. **Second**, Rosenberger discloses that a “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].

189. For instance, as discussed with element [15.1], each “weighted signal” includes an indication of an audio input quality associated with a given “control

device.” Moreover, as also explained before in connection with elements [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* ¶¶107-17, 170-71.

190. Thus, it is my opinion that Rosenberger discloses [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.*

191. Consequently, it is my opinion that Rosenberger discloses a *first computing device* as required by dependent claim 15.

192. Because dependent claims 4 and 14 recite similar subject matter as dependent claim 15, it is also my opinion that Rosenberger anticipates these dependent claims for the same reasons as it anticipates claim 15.

193. In this regard, claim 4 depends from claim 1 and recites:

[4.1] *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*

[4.2] wherein determining 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.

194. Likewise, claim 14 depends from claim 11 and recites:

[14.1] 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

[14.2] wherein the instructions cause 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.

195. As is evident, (i) [4.1] and [4.2] map to [15.1] and [15.2], respectively, and (ii) [14.1] and [14.2] map to [15.1] and [15.2], respectively. Thus, Rosenberger discloses each of the elements of claims 4 and 14 for the same reasons that Rosenberger discloses each of the elements of claim 15.

C. Ground 2: Rosenberger+Jang Renders Dependent Claims 2-3, 9-10, 12 Obvious

196. As explained above, it is my opinion that Rosenberger discloses all of the elements of independent claims 1, 7, and 11. While Rosenberger itself discloses the additional elements of claims (i) 2, 9, and 12 and (ii) 3 and 10, as explained in detail below, it is my opinion that Rosenberger in combination with Jang renders the additional elements of these dependent claims obvious as well.

1. Rosenberg+Jang Render Obvious [9.1]-[9.2] & [10.1]

197. Claim 9 depends from claim 7 and further requires: **[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.*

198. As noted before, claim 10 depends from claim 9 and further specifies **[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 long list of possible current contexts that include (i) a current time as defined by the first computing device, (ii) one or more applications currently executing at the first computing device, and (iii) an operating mode of the first computing device.*

199. As I explained above, Rosenberg discloses all of the elements of independent claim 7. While Rosenberg itself discloses the additional elements of claims 9 and 10, Rosenberg and Jang in combination render obvious the additional elements of claims 9 and 10 as well.

200. *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].

201. For instance, Jang discloses determining multiple *current contexts* as that phrase is used in the ’748 Patent (*supra* ¶163-65), 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 10,⁶ and (iii) whether the “electronic device” is “executing an application relating to” a received voice command (*id.*, ¶174-81, FIGs. 15-16), which constitutes *one or more applications currently executing at the first computing device*, as recited in claim 10. I discuss these examples below with reference to element [9.2].

202. Thus, it is my opinion that Jang discloses a *first computing device* operable to [9.1] *determine a current context of the first computing device* in the

⁶ See, e.g., ’748 Patent, 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**.”)

form of [10.1] *a current time as defined by the first computing device,... one or more applications currently executing at the first computing device,... [or] an operating mode of the first computing device....*

203. **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* ¶107-17, 168-71.

204. Similarly, Jang discloses that each “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. *Supra* ¶75-81. Jang discloses embodiments where the arbitration determination is based on multiple, different types of “voice recognition results” (or “voice command results”). *Id.*; *see also, e.g.*, Jang, ¶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.”), ¶205-206:

The priority values may be determined based on the voice recognition results from the electronic devices. For example, the priority values may be assigned in an order of electronic devices ***satisfying better conditions*** to perform the input voice command among a plurality of electronic devices.

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.

205. I will now discuss Jang’s teachings regarding the specific *current contexts* that I identified for elements [9.1], [10.1].

206. 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”). Specifically, 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 is within (less than) that interval:

The voice recognition result received from the other electronic devices includes ***information on time that the user’s voice command was entered***. For instance, when the ***first electronic device 100*** receives a

first voice command *at a first time* and the *second electronic device 10a* receives the first voice command at *a second time*, there might be a tiny difference in the time of recognizing the voice command in consideration of a distance difference between the two devices. However, when the time difference exceeds a predetermined interval, it is difficult to determine the voice command as being generated by the same user at the same time.

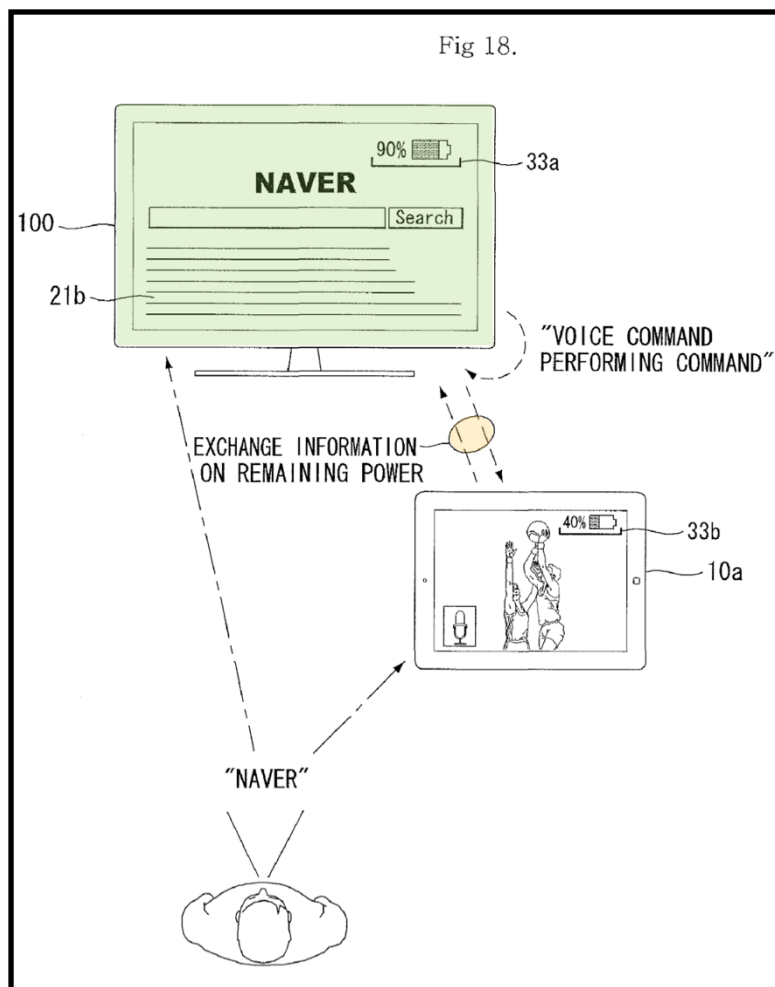
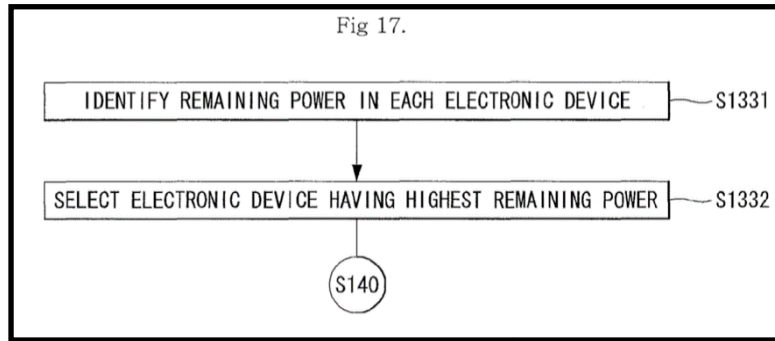
Accordingly, *in sharing the voice recognition results* between a plurality of devices, *time information* received from the devices *may be taken into consideration*. For instance, when a difference in input time between two devices is within a predetermined interval, the controller 180 may determine that the user voice commands have been input at the same time. In contrast, when the difference in input time is more than the predetermined interval, the controller 180 may determine that the voice command input at the first time has been reentered at the second time. *The controlling method for an electronic device according to the embodiments of the present disclosure may apply to the former situation.*

Jang, ¶133-134.

207. With regard 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”). Specifically, Jang explains that the given “electronic device” compares remaining power levels and selects the electronic device “having more remaining power....” *Id.*, ¶182-86.

208. Jang illustrates an electronic device using its determined remaining-power current context to make an arbitration determination with FIG. 18 specifically

illustrating “first electronic device 100” selecting itself based on its determined remaining power level being greater than the remaining power level of the “second electronic device 10”:



Id., FIG. 18 (annotations added).

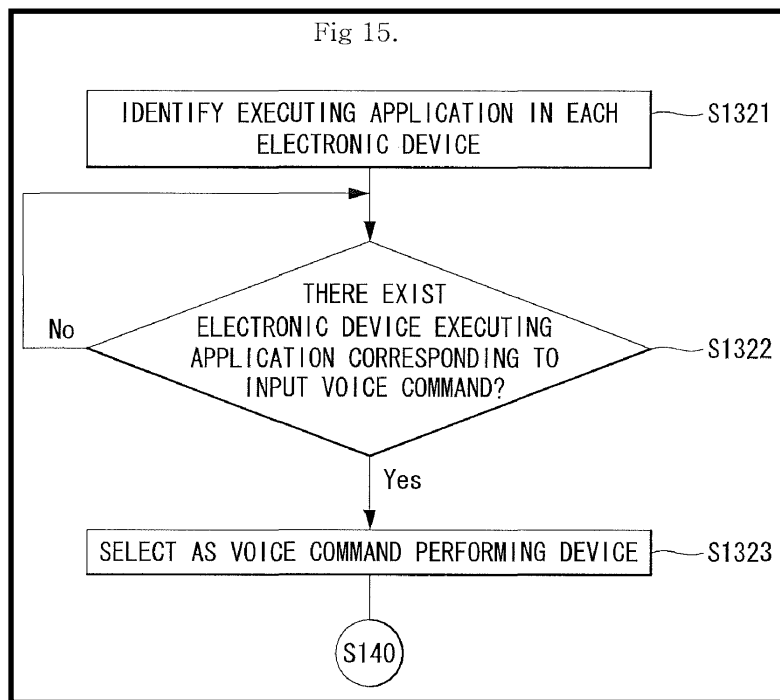
209. With respect to *one or more applications currently executing at the first computing device*, Jang discloses that each “electronic device” is operable to make an arbitration determination based on an application executing at the given “electronic device” (e.g., “first electronic device 100”) and “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 “electronic device” executing an application that corresponds to a user’s voice request:

According to an embodiment, the controller 180 ***identifies*** an ***application in execution*** in ***each electronic device*** (S1321).

Then, the controller 180 ***identifies*** whether there is an ***electronic device executing*** an ***application corresponding*** to an input ***voice command*** among a plurality of electronic devices (S1322), and if any (yes in step S1322), the controller 180 of the first electronic device 100 selects the electronic device as a voice command performing device (S1323).

Id., ¶175-76.

210. Jang illustrates an “electronic device” using its determined current context regarding an executing application to make an arbitration determination:



211. Jang illustrates and describes a specific example where the user’s voice utterance states “transfer a picture to Chulsu” and the “first electronic device 100” selects the “second electronic device 10” as the arbitration winner based on the first electronic device’s current context (e.g., executing a “broadcast application”) being deemed not related to the voice command and the second electronic device’s “voice recognition results” indicating a current context (e.g., executing an “email application”) deemed related to the voice command:

According to an embodiment, the method for controlling a mobile terminal may select 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.

For instance, a voice command saying “*transfer a picture to Chulsu*” enables a predetermined *picture to be transferred* to an electronic

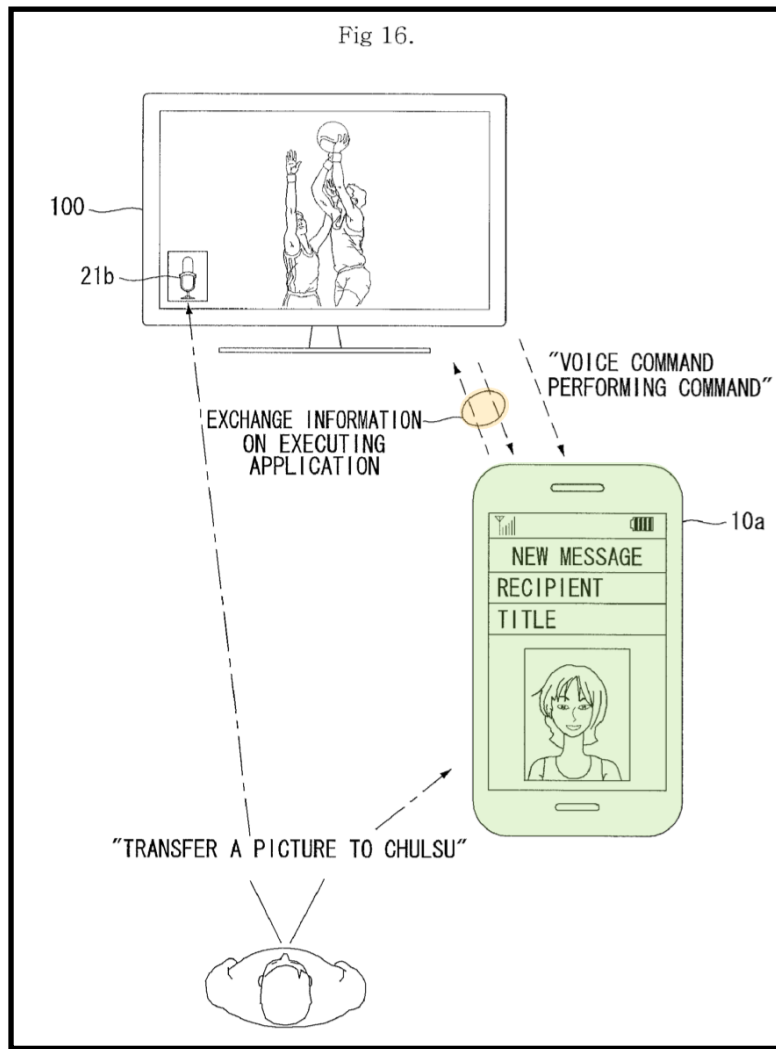
device ***through emailing or MMS mailing***. Accordingly, when there is any electronic device ***executing an application relating to messaging or emailing*** among a plurality of electronic devices, it is most efficient for the corresponding electronic device to perform the voice command.

Referring to FIG. 16, the ***second electronic device 10a*** is ***executing an email application***, and the ***first electronic device 100*** is ***executing a broadcast program***. Under this circumstance, when the voice command saying “transfer a picture to Chulsu” is input to each of the electronic devices, the first electronic device 100 and the second electronic device 10a may exchange the programs (or contents) presently in execution to each other.

The 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.

Accordingly, the controller 180 of the first electronic device 100 may transfer a command to the second electronic device 10 a to enable a function corresponding to the voice command (“transfer a picture to Chulsu”) to be performed. In response to the command, the second electronic device 10 a may perform the voice command.

Id., ¶177-181.



Id., FIG. 16 (annotations added).

212. Thus, it is my opinion that Rosenberger and Jang render obvious a *first computing device* operable to [9.2] *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.*

213. Consequently, it is my opinion that Rosenberger and Jang render obvious a *first computing device* as required by dependent claims 9 and 10.

214. Because dependent claims 2 and 12 recite similar subject matter as dependent claim 9 (*supra* ¶¶160, 173-75), it is also my opinion that Rosenberger and Jang render obvious these dependent claims for the same reasons that they render obvious claim 9. Likewise, because dependent claim 3 recites similar subject matter as dependent claim 10 (*supra* ¶¶176, 179-81), it is also my opinion that Rosenberger and Jang render obvious claim 3 for the same reasons that they render obvious claim 10.

2. Motivation to Combine Rosenberger+Jang

215. In my opinion, 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 Jang's disclosed (i) current time, (ii) remaining power, and/or (iii) application executing. In this way, a POSITA would achieve [9.1], [9.2], and [10.1] (again, assuming Rosenberger does not already disclose these elements). It is my opinion that a POSITA would have been motivated to make this combination for several reasons.

216. **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 when multiple voice-enabled devices detect the same user utterance. *See, e.g.*, Rosenberger, 2:42-59, 9:18-10:4 (describing “duplicate response” problem); Jang, ¶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.”).

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

218. 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* ¶¶66-70, 75-81. 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.,* Rosenberger, 10:34-67, 11:27-31; Jang, ¶122, 127-136, FIG. 11. A POSITA studying Rosenberger would have been motivated to seek out other references that disclose similar arbitration techniques, like Jang, to obtain additional details regarding such techniques.

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

220. For example, Jang expressly explains that Jang’s (i) current time, (ii) remaining power, and/or (iii) application executing would supplement and be incorporated into a “coordination” technique like that described in Rosenberger that involves exchange of an audio quality value (e.g., “weighted signal”). *See, e.g.,* Jang, ¶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.

221. 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.

222. A POSITA would have been motivated to incorporate Jang’s application-executing context when implementing Rosenberger’s “control devices” “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, e.g.*, Rosenberger, 15:62-16:36.

223. Likewise, a POSITA would have been motivated to incorporate Jang’s remaining-power context when implementing Rosenberger’s “control device” as a “[t]abletop device[]... ideally run on batteries” (*id.*, 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, e.g.*, Jang, ¶183-86.

224. **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. 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.

225. For at least these reasons, it is my opinion that a POSITA would have been motivated to combine Rosenberger with Jang before the earliest possible priority date of the '748 Patent (June 24, 2014).

XII. LACK OF SECONDARY CONSIDERATIONS

226. I am unaware of any secondary considerations of non-obviousness relating to the Challenged Claims of the '748 Patent.

227. However, even if any such considerations existed, it is my opinion that they would not overcome the inescapable conclusion that a POSITA would have found the Challenged Claims obvious as of the effective filing date of the '748 Patent. This is because the case for obviousness in view of Rosenberger in view of Jang is so clear and compelling.

XIII. CONCLUSION

228. I declare that all statements made herein of my own knowledge are true and correct and that all statements made on information and belief are believed to be true. I further declare that all statements made herein were made with the knowledge

that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code.

Executed on August 7, 2023 in Lexington, Kentucky

A handwritten signature in blue ink, appearing to read "Michael T. Johnson". The signature is fluid and cursive, with the first name "Michael" and last name "Johnson" being more legible than the middle initial "T.".

Michael T. Johnson

Appendix A

Curriculum Vitae

Michael T. Johnson

Associate Dean for Undergraduate Education and Student Success
Professor and Chair, Electrical and Computer Engineering,
University of Kentucky College of Engineering
453 F. Paul Anderson Tower, Lexington, KY 40506, Phone (859) 257-0717
Email: mike.johnson@uky.edu
Web page: <http://johnson.engineering.uky.edu/>

CURRICULUM VITAE EXECUTIVE SUMMARY

Publications: **52 journal papers, over 150 refereed publications/presentations**
 Citations ≈3400, h-index 31, i10-index 68 (Google scholar August 2023)

Grant funding: **6 major grants, over \$3.5 million in external funding (\$2.7m as PI)**
 Multi-institution, interdisciplinary, international collaborations

Students: **8 PhD graduates, 14 MS graduates**

Leadership: **Associate Dean for Undergrad Education and Student Success, UK COE (2023-)**
 President, Southeast ECE Department Heads Association (2021-22)
 SEC Academic Leadership Development Program, (2019-2020)
 Department Chair, UK ECE (2016-present)
 Director, UK CoE Scholars in Engineering Leadership (2017-2020)
 Chair of Marquette Board of Graduate Studies (2012-2016)
 Director of Graduate Studies, Marquette EECE (2009-2012)
 Senior Visiting Professor, Tsinghua University (2008-2009, 2014-2015)
 Engineering Manager (1993-1996)

Honors: **Featured in “Groundbreaking Thinkers in Wisconsin” series 2007**
 Marquette COE Researcher of the Year 2006-2007
 Engineers and Scientists of Milwaukee Young Engineer of the Year 2006
 HKN EECE Teacher of the year 2005, 2014

RESEARCH SPECIALIZATION

- Speech Processing: speech recognition and enhancement, natural language processing
- Signal Processing: bioacoustics, microphone array processing, signal enhancement
- Pattern Recognition and Machine Learning: statistical estimation and classification

EDUCATION

Doctor of Philosophy, School of Electrical and Computer Engineering, August 2000
Purdue University, West Lafayette, IN GPA: 4.00/4.00
"Incorporating Prosodic Information and Language Structure into Speech Recognition Systems"
Advisor: Dr. Leah H. Jamieson

Master of Science in Electrical Engineering (concentration Digital Signal Processing), Dec. 1994
University of Texas at San Antonio, San Antonio, TX GPA: 4.00/4.00
"An Adaptive Texture Model Based on Markov Random Fields"
Advisor: Dr. Mita Desai

Bachelor of Science, Engineering with Electrical Concentration, April 1990
LeTourneau University, Longview, TX GPA: 3.60/4.00

Bachelor of Science, Computer Science and Engineering with minor in Mathematics, April 1989
LeTourneau University, Longview, TX GPA: 3.57/4.00

ACADEMIC WORK EXPERIENCE

8/2016 – present University of Kentucky, Lexington, KY

- *Department Chair, Electrical and Computer Engineering* (2016-present)
- *Full Professor, Electrical and Computer Engineering* (2016 - present)

8/2000 – 8/2016 Marquette University, Milwaukee, WI

- *Full Professor, Electrical and Computer Engineering*: (2013 - 2016)
- *Director of Graduate Studies, Electrical and Computer Engineering*: (2009 - 2012)
- *Associate Professor with Tenure, Electrical and Computer Engineering* (2007 – p2009)
- *Assistant Professor, Electrical and Computer Engineering*: (2000 – 2007)

2008-2009 and 2014-2015 Tsinghua University, Beijing, China (two 1-year sabbaticals)

- *Senior Visiting Faculty, Electronic Engineering Dept.* While on sabbatical, taught courses and collaborated on research in speech processing with Prof. Liu Jia.

9/1996 – 8/2000 Purdue University, West Lafayette, IN

- *Graduate Research Assistant, Audiology Dept.*: (5/97 to 5/00) Research focused on studying the consistency of speech articulator movement in children and young adults.
- *Graduate Research Assistant, ECE Dept.*: (1/97 to 1/98) Part of an NSF-funded research grant working on interfacing speech recognition and language processing systems.
- *Teaching Assistant, ECE Dept.*: (9/96 to 12/96 and 1/98 to 5/98) Assisted with a course in Speech Processing and taught a laboratory course in Digital Circuits.

1/1993 - 9/1993 University of Texas at San Antonio, San Antonio, TX

- *Graduate Research Assistant, UTSA Image Processing Lab*: Did research in the field of biological image processing as part of masters thesis in mammogram image analysis.

INDUSTRIAL WORK EXPERIENCE

9/1993 - 7/1996 SNC Manufacturing, Oshkosh, WI

- *Engineering Manager, Telecommunications Division* (11/94 - 7/96): Responsible for all new and existing product lines, including design, project management, and personnel.
- *Senior Engineer* (9/93-11/94): Project lead for new products, primarily in the area of SNC's Lyte Lynx line of fiber-optic interfaces for high-voltage environments.

9/1991 - 1/1993 Datapoint Corporation , San Antonio, TX

- *Engineer II*: Hardware engineer on the ArcnetPlus networking and MINX videoconferencing projects. Lead engineer on the PC HUB project for ArcnetPlus.

5/1990 - 9/1991 Micronyx, Inc. / Micro Technology Services, Inc., Richardson, TX

- *Design Engineer*: Telecommunications and embedded microprocessors design, at both hardware and software levels, for Micronyx and its spin-off company MTSI.

5/1989 - 8/1989 GTE , Ft. Wayne, IN *Engineering Intern*: Designed circuits for Business Services, including a project restructuring an analog phone system into a digital carrier system.

6/1988 - 8/1988 NBS , Ft. Wayne, IN *Electrical Technician*: Gained experience with troubleshooting circuit boards, soldering, and designing and building test equipment.

PUBLICATIONS

Peer-reviewed Journal Publications

1. Rahim Soleymanpour, Mohammad Soleymanpour, Anthony J Brammer, Michael T Johnson, Insoo Kim, Speech Enhancement Algorithm Based on a Convolutional Neural Network Reconstruction of the Temporal Envelope of Speech in Noisy Environments, IEEE Access, Volume 11, 5328-5336, January 2023.
2. Julie N. Oswald, Amy M. Van Cise, Angela Dassow, Taffeta Elliott, Michael T. Johnson, Andrea Ravignani, Jeffrey Podos, A Collection of Best Practices for the Collection and Analysis of Bioacoustic Data, Applied Sciences, Volume 12, No. 23, November 2022.
3. Chao Li; Wang, Qiyue Wang, Wenhua Jiao, Michael T. Johnson, Yuming Zhang, Deep Learning-Based Detection of Penetration from Weld Pool Reflection Images, Volume 99 No. 2, Welding Journal, September 2020.
4. Qiyue Wang, Wenhua Jiao, Rui Yu; Michael T. Johnson, Yuming Zhang, Virtual Reality Robot-Assisted Welding Based on Human Intention Recognition, IEEE Transactions on Automation Science and Engineering, Volume 17 No. 2, April 2020.
5. Yi Liu; Liang He; Jia Liu; Michael T. Johnson, Introducing phonetic information to speaker embedding for speaker verification, EURASIP Journal on Audio, Speech, and Music Processing, December 2019.
6. Qiyue Wang, Yongchao Cheng, Wenhua Jiao, Michael T. Johnson, Yuming Zhang, Virtual reality human-robot collaborative welding: A case study of weaving gas tungsten arc welding, Journal of Manufacturing Processes, Volume 49, 2019.
7. Qiyue Wang, Wenhua Jiao, Rui Yu, Michael Johnson, YuMing Zhang, 2019. Modeling of Human Welder Operation in Virtual Reality Human-Robot Interaction Welding System. IEEE Robotics and Automation Letters, 4(3): 2958-2964, July 2019.
8. Ruijie Yan, Liangrui Peng, Shanyu Xiao, Michael T. Johnson, Shengjin Wang, Dynamic temporal residual network for sequence modeling. International Journal on Document Analysis and Recognition, Volume 22, 235–246, July 2019.
9. Liang He, Xianhong Chen, Can Xu, Yi Liu, Jia Liu, Michael T. Johnson, Latent class model with application to speaker diarization, EURASIP Journal on Audio, Speech, and Music Processing, July 2019.
10. Jian Kang, Wei-Qiang Zhang, Weiwei Liu, Jia Liu., Michael T. Johnson, Lattice Based Transcription Loss for End-to-End Speech Recognition, Journal of Signal Processing Systems, Volume 90 Issue 7, pp 1013-1023, July 2018
11. Jian Kang, Wei-Qiang Zhang, WeiWei Liu, Jia Liu, Michael T. Johnson, Advanced recurrent network-based hybrid acoustic models for low resource speech recognition, EURASIP Journal on Audio, Speech, and Music Processing, 2018.
12. Liang He, Xianhong Chen, Can Xu, Jia Liu, Michael T. Johnson, Local Pairwise Linear Discriminant Analysis for Speaker Verification, IEEE Signal Processing Letters Volume 25 No. 10, 1575-1579, 2018.
13. Jeffrey Berry, Andrew Kolb, James Schroeder, Michael T. Johnson, Jaw Rotation in Dysarthria Measured With a Single Electromagnetic Articulography Sensor, American Journal of Speech – Language Pathology, June 2017.
14. Ji, An; Johnson, Michael T; Berry, Jeffrey J., Parallel Reference Speaker Weighting for Kinematic-Independent Acoustic-to-Articulatory Inversion, IEEE/ACM Transactions on Audio, Speech, and Language Processing",24,10, October 2016, 1865-1875.

15. Xu-Kui Yang, Liang He, Dan Qu, Wei-Qiang Zhang, Michael T Johnson, Semi-supervised feature selection for audio classification based on constraint compensated Laplacian score, *EURASIP Journal on Audio, Speech, and Music Processing*, 1, 2016, 1-10.
16. Liu Wei-Wei, Cai Meng, Zhang Wei-Qiang Zhang, Liu Jia, Johnson Michael T., "Discriminative Boosting Algorithm for Diversified Front-End Phonotactic Language Recognition," *Journal of Signal Processing Systems*, February 2016.
17. Wei-Qiang Zhang, Cong Guo, Qiao Zhang, Jian Kang, Liang He, Jia Liu, and Michael T. Johnson, A Speech Enhancement Algorithm Based on Computational Auditory Scene Analysis. *Journal of Tianjin University*, in press, 2015. (Chinese journal. Original language citation: 张卫强, 郭聪, 张乔, 康健, 何亮, 刘加, and Michael T. Johnson, 种基于计算听觉场景分析的语音增强算法, 天津大学学报.)
18. B.T.W. Bochera, K. Cherukurib, J.S. Makib, M. Johnson, D.H. Zitomer, Relating methanogen community structure and anaerobic digester function, *Water Research*, 70 (1), March 2015, 425-435.
19. PM Scheifele, MT Johnson, M Fry, B Hamel, K Laclede, Vocal classification of vocalizations of a pair of Asian Small-Clawed otters to determine stress, *The Journal of the Acoustical Society of America*, 138 (1), EL105-EL109, 2015.
20. Trawicki, Marek B. and Johnson, Michael T., "Beta-order minimum mean-square error multichannel spectral amplitude estimation for speech enhancement", *International Journal of Adaptive Control and Signal Processing*, January 2015.
21. Arik Kershenbaum, Daniel Blumstein, Marie Roch, Michael T. Johnson, et. al., Acoustic sequences in non-human animals: a tutorial review and prospectus, *Biological Reviews*, 2014.
22. C Yu, KK Wójcicki, PC Loizou, JHL Hansen, MT Johnson, "Evaluation of the importance of time-frequency contributions to speech intelligibility in noise", *The Journal of the Acoustical Society of America*, vol. 135, no. 5, May 2014, 3007-3016.
23. Marek B. Trawicki, Michael T. Johnson, "Speech enhancement using Bayesian estimators of the perceptually-motivated short-time spectral amplitude (STSA) with Chi speech priors", *Speech Communication*, vol. 57, no. 2, February 2014, pp101-103.
24. Liu Weiwei, Zhang Weiqiang, Johnson Michael T., Liu Jia, "Homogenous ensemble phonotactic language recognition based on SVM supervector reconstruction", *EURASIP Journal on Audio, Speech, and Music Processing* vol. 2014 no. 1, January 2014, pp 1-13.
25. Junhong Zhao, Wei-Qiang Zhang, Hua Yuan, Michael T Johnson, Jia Liu, Shanhong Xia, "Exploiting contextual information for prosodic event detection using auto-context", *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2013, no. 1, December 2013 pp 1-14.
26. Marek B. Trawicki, Michael T. Johnson, "Distributed multichannel speech enhancement based on perceptually-motivated Bayesian estimators of the spectral amplitude", *IET Signal Processing*, vol. 7, no.4, April 2013, pp. 337-344.
27. An Ji, Michael T. Johnson, Edward J. Walsh, JoAnn McGee, Doug L. Armstrong, Discrimination of individual tigers (*Panthera tigris*) from long distance roars, *The Journal of the Acoustical Society of America*, vol. 133 no. 3, March 2013, pp1762-1769.
28. Yongzhe Shi, Weiqiang Zhang, Jia Liu, Michael T. Johnson, "RNN language model with word clustering and class-based output layer", *EURASIP Journal on Audio, Speech, and Music Processing* vol. 2013 no. 1, January 2013, pp1-7.

29. Peter M. Scheifele, Michael T. Johnson, David C. Byrne, John G. Clark, Ashley Vandlik, Laura W. Kretschmer, Kristine E. Sonstrom, "Noise impacts from professional dog grooming forced-air dryers", *Noise and Health*, vol. 14 no. 60, October 2012, p224-226.
30. Wen-Lin Zhang, Wei-Qiang Zhang, Bi-Cheng Li, Dan Qu and Michael T. Johnson, "Bayesian Speaker Adaptation Based on a New Hierarchical Probabilistic Model", *IEEE Transactions on Speech and Language Processing*, vol. 20 no. 7, July 2012, pp2002-2015.
31. Yuxiang Shan, Yan Deng, Jia Liu, Michael T. Johnson, "Phone lattice reconstruction for embedded language recognition in LVCSR", *EURASIP Journal on Audio Speech and Music Processing*, vol. 2012, no. 15, April 2012, pp1-13.
32. Marek B. Trawicki, Michael T. Johnson, "Distributed multichannel speech enhancement with minimum mean-square error short-time spectral amplitude, log-spectral amplitude, and spectral phase estimation". *Signal Processing*, vol. 92 no. 2, February 2012, pp 345-356.
33. Peter M. Scheifele, Michael T. Johnson, Laura W. Kretschmer, John G. Clark, Deborah Kemper, Gopu Potty, "Ambient habitat noise and vibration at the Georgia Aquarium", *The Journal of the Acoustical Society of America*, vol. 132 no. 2, February 2012, EL88-EL94.
34. Wei Qiang Zhang, Liang He, Yan Deng, Jia Liu, Michael T. Johnson, "Time-Frequency Cepstral Features and Heteroscedastic Linear Discriminant Analysis for Language Recognition", *IEEE transactions on audio, speech, and language processing* vol. 19 no. 2, February 2011, pp.266-276.
35. Kuntoro Adi, Michael T. Johnson, and Tomasz S. Osiejuk, "Acoustic Censusing using Automatic Vocalization Classification and Identity Recognition," *Journal of the Acoustical Society of America*, vol. 127, no. 2, February 2010, pp 874-883.
36. Yao Ren, Michael T. Johnson, Patrick J. Clemins, Michael Darre, Sharon Stuart Glaeser, Tomasz S. Osiejuk, and Ebenezer Out-Nyarko, "A Framework for Bioacoustic Vocalization Analysis Using Hidden Markov Models", *Algorithms*, vol. 2 no. 3, November 2009, pp 1410-1428.s
37. Yanmin Qian, Jia Liu, Michael T. Johnson "Efficient Embedded Speech Recognition for Very Large Vocabulary Mandarin Car-Navigation Systems," *IEEE Transactions on Consumer Electronics*, vol. 55, no. 3, August 2009, 1496-1500.
38. Yao Ren, Michael T. Johnson, Jidong Tao, "Perceptually motivated wavelet packet transform for bioacoustic signal enhancement", *Journal of the Acoustical Society of America*, vol. 124, no. 1, July 2008, 316-327.
39. Jidong Tao, Michael T. Johnson, Tomasz S. Osiejuk, "Acoustic model adaptation for ortolan bunting (*Emberiza hortulana* L.) song type classification", *Journal of the Acoustical Society of America*, vol. 123, no. 3, March 2008, 1582-1590.
40. Michael T. Johnson, Xiaolong Yuan, Yao Ren, "Speech signal enhancement through adaptive wavelet thresholding", *Speech Communication*, vol. 49, No. 2, February 2007, pp 123-133.
41. Patrick Clemins and Michael T. Johnson, "Generalized perceptual linear prediction (gPLP) features for animal vocalization analysis", *Journal of the Acoustical Society of America*, Vol. 120, No. 1, July 2006, pp 527-534.

42. Richard J. Povinelli, Michael T. Johnson, Andrew C. Lindgren, Felice Roberts, and Jinjin Ye, "Statistical Models of Reconstructed Phase Spaces for Signal Classification", *IEEE Transactions on Signal Processing*, vol. 54, no. 6, June 2006, 2178-2186.
43. Michael T. Johnson and Richard J. Povinelli, "Generalized phase space projection for nonlinear noise reduction", *Physica D*, No. 201, February 2005, pp 306-317.
44. Kevin M. Indrebo, Richard J. Povinelli, and Michael T. Johnson, "Sub-banded Reconstructed Phase Spaces for Speech Recognition", *Speech Communication*, Vol. 48, No. 7, July 2006, pp 760-774.
45. Michael T. Johnson, "Capacity and Complexity of HMM Duration Modeling Techniques", *IEEE Signal Processing letters*, Volume 12, No. 5, May 2005, pp 407-410.
46. Patrick J. Clemins, Michael T. Johnson, Kirsten M. Leong, and Anne Savage, "Automatic Classification and Speaker Identification of African Elephant (*Loxodonta Africana*) vocalizations", *Journal of the Acoustical Society of America*, Volume 117, No. 2, February 2005, pp 956-963.
47. Michael T. Johnson, Richard J. Povinelli, Jinjin Ye, Xiaolin Liu, Andrew Lindgren, and Kevin Indrebo, "Time-Domain Isolated Phoneme Classification using Reconstructed Phase Spaces", *IEEE Transactions on Speech and Audio Processing*, Vol. 13, No. 4, July 2005, pp 458-466.
48. Richard J. Povinelli, Michael T. Johnson, Andrew C. Lindgren, and Jinjin Ye, "Time Series Classification using Gaussian Mixture Models of Reconstructed Phase Spaces", *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 6, June 2004, 779-783.
49. Kelly Stonger and Michael T. Johnson, "Optimal Calibration of PET Crystal Position Maps using Gaussian Mixture Models", *IEEE Transactions on Nuclear Science*, Vol. 51, No. 1, 2004, pp 85-90.
50. Yang Liu, Mary P. Harper, Michael T. Johnson, and Leah H. Jamieson, "The Effect of Pruning and Compression on Graphical Representations of the Output of a Speech Recognizer", *Computer Speech and Language*, Volume 17, 2003, pp. 329-256.
51. A.M. Surprenant, S.L. Hura, M.P. Harper, L.H. Jamieson, G.Long, S.M. Thede, A. Rout, T.H. Hsueh, S.A. Hockema, M.T. Johnson, P. Srinivasan, and C.M. White, "Familiarity and Pronouncibility of Nouns and Names", *Behavior Research, Methods, Instruments, & Computers*, Vol. 31, Nov. 1999, pp. 638-649.
52. Anne Smith, Michael Johnson, Clare McGillem, and Lisa Goffman "On the Assessment of Stability and Patterning of Speech Movements", *Journal of Speech, Language, and Hearing Research*, Vol. 43, 2000, pp 277-286.

Peer-reviewed Conference Publications and Presentations

53. Mohammad Soleymanpour, Michael T. Johnson, Rahim Soleymanpour, Jeffrey Berry, Synthesizing Dysarthric Speech Using Multi-speaker TTS for Dysarthric Speech Recognition, International Conference on Acoustics, Speech, and Signal Processing (ICASSP), May 2022, Singapore.
54. Narjes Bozorg, Michael T Johnson, Mohammad Soleymanpour, Autoregressive Articulatory WaveNet Flow for Speaker-Independent Acoustic-to-Articulatory Inversion,

- 2021 international Conference on Speech Technology and Human-Computer Dialog (SpeD), pp. 156-161, October 2021.
55. Subash Khanal, Michael T Johnson, Mohammad Soleymanpour, Narjes Bozorg, Mispronunciation Detection and Diagnosis for Mandarin Accented English Speech, 2021 international Conference on Speech Technology and Human-Computer Dialog (SpeD), pp. 62-67, October 2021.
 56. Mohammad Soleymanpour, Michael T Johnson, Jeffrey Berry, Dysarthric Speech Augmentation Using Prosodic Transformation and Masking for Subword End-to-end ASR, 2021 international Conference on Speech Technology and Human-Computer Dialog (SpeD), pp. 42-46, October 2021.
 57. Mohammad Soleymanpour, Michael T Johnson, Jeffrey Berry, Comparison in Suprasegmental Characteristics between Typical and Dysarthric Talkers at Varying Severity Levels, 2021 international Conference on Speech Technology and Human-Computer Dialog (SpeD), pp. 52-56, October 2021.
 58. Mohammad Soleymanpour, Michael T Johnson, Jeffrey Berry, Increasing the Precision of Dysarthric Speech Intelligibility and Severity Level Estimate, International Conference on Speech and Computer, pp. 670-679, September 2021.
 59. Subash Khanal, Michael T. Johnson, Narjes Bozorg, Articulatory Comparison of L1 and L2 Speech for Mispronunciation Diagnosis, IEEE Spoken Language Technology Workshop (SLT) 2021, January 2021, Shenzhen China.
 60. Narjes Bozorg and Michael Johnson, Acoustic-to-Articulatory Inversion with Deep Autoregressive Articulatory-WaveNet, Interspeech 2020, October 2020, Shanghai China.
 61. Jiaqi Xie, Ruijie Yan, Shanyu Xiao, Liangrui Peng, Michael T. Johnson, Weiqiang Zhang, Dynamic Temporal Residual Learning for Speech Recognition, International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2020, May 2020, Barcelona, Spain.
 62. Jeffrey Berry and Michael T. Johnson, Perceived Acoustic Working Space Modulates Sensorimotor Learning, Motor Speech Conference, Santa Barbara, California, February 2020.
 63. Narjes Bozorg and Michael T. Johnson, Reference speaker selection for kinematic-independent acoustic-to-articulatory-inversion, J. Acoust. Soc. Am., vol. 145, no. 3, pp. 1932–1932, Mar. 2019, presentation at the Spring 2019 Meetings of the Acoustical Society of America.
 64. McGowan, Kevin B; Johnson, Michael T; Combs, Aleah; Soleymanpour, Mohammad; Acoustic, non-invasive measurement of velopharyngeal aperture using a high frequency tone, presentation at the Spring 2019 Meetings of the Acoustical Society of America.
 65. Narjes Bozorg, Michael T. Johnson, Jeffrey Berry, Comparing Articulography Consistency Between L1 and L2 Speakers, IEEE International Symposium on Signal Processing and Information Technology, Ajman, United Arab Emirates, December 2019.
 66. Narjes Bozorg, Michael T. Johnson, MLLR-PRSW for Kinematic-Independent Acoustic-to-Articulatory Inversion, IEEE International Symposium on Signal Processing and Information Technology, Ajman, United Arab Emirates, December 2019.
 67. Kevin McGowan, Michael Johnson, Aleah Combs, Mohammad Soleymanpour, Acoustic, non-invasive measurement of velopharyngeal aperture using a high frequency tone, Annual Spring Meeting of the Acoustical Society of America, Louisville, Kentucky, May 2019. (JASA Vol.145(3), pp.1931-1931.)

68. Kevin McGowan, Michael Johnson, Aleah Combs, Mohammad Soleymanpour, Acoustic, non-invasive measurement of velopharyngeal aperture using a high frequency tone, International Congress of Phonetic Sciences (ICPhS) 2019, Melbourne, August 2019.
69. Narjes Bozorg and Michael T. Johnson, Comparing Performance of Acoustic-to-Articulatory Inversion for Mandarin Accented English and American English Speakers, International Symposium on Signal Processing and Information Technology, Louisville, KY, December 2018.
70. Sidrah Liaqat, Narjes Bozorg, Neenu Jose, Patrick Conrey, Antony Tamasi and Michael T. Johnson, Domain Tuning Methods for Bird Audio Detection, Detection and Classification of Acoustic Scenes and Events 2018, Surrey, UK, November 2018.
71. Chao Li, Qiyue Wang, Jinsong Chen, Michael Johnson, Qiang Chang, YuMing Zhang, Machine Learning Based Detection of Weld Joint Penetration from Weld Pool Reflection Images, 14th IEEE International Conference on Automation Science and Engineering, Munich, Germany, 2018.
72. Yi Liu, Liang He, Jia Liu, Michael T. Johnson, Speaker Embedding Extraction with Phonetic Information, Interspeech Hyderabad, India, 2018.
73. Yi Liu, Liang He, Yao Tian, Zhuzio Chen, Jia Liu, Michael T. Johnson, Comparison of multiple features and modeling methods for text-dependent speaker verification, Automatic Speech Recognition and Understanding Workshop (ASRU), Okinawa, Japan, 2017.
74. Yi Liu, Liang He, Weiqiang Zhang, Jia Liu, Michael T. Johnson, Investigation of Frame Alignments for GMM-based Digit-prompted Speaker Verification, Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, Honolulu, Hawaii, November 2018.
75. Weilian Song, Tawfiq Salem, Nathan Jacobs, Michael T. Johnson, Detecting the Presence of Bird Vocalizations in Audio Segments Using a Convolutional Neural Network Architecture, 4th International Symposium on Acoustic Communication by Animals, Omaha, NE, July 2017.
76. Peter M. Scheifele, Giulia Raponi, Haylea Roark, and Lesa Scheifele, Michael T. Johnson, Neenu Jose, Vocalization and Hearing Acuity of the Asian Small-Clawed Otter (*Amblonyx cinereus*) in Response to Anthropogenic Noise, 4th International Symposium on Acoustic Communication by Animals, Omaha, NE, July 2017.
77. Jeff Berry, Andrew Kolb, James Schroeder, Michael T Johnson, Jaw rotation in dysarthria measured with a single electromagnetic articulography sensor, Proceedings of Spring 2017 Meetings on Acoustics, Boston, June 2017
78. Jeffrey J Berry, Abigail Stoll, Deriq Jones, Seyedramin Alikiaamiri, Michael T Johnson, Second language pronunciation training using acoustic-to-articulatory inversion, Proceedings of Spring 2017 Meetings on Acoustics, Boston, June 2017.
79. Berry, Jeffrey J; Bagin, Ramie; Schroeder, James; Johnson, Michael T; Sensitivity and specificity of auditory feedback driven articulatory learning in virtual speech, Proceedings of Spring 2017 Meetings on Acoustics, Boston, June 2017.
80. Andrew J Kolb, Michael T Johnson, Jeffrey Berry, Interpolation of Tongue Fleshpoint Kinematics from Combined EMA Position and Orientation Data, Sixteenth Annual Conference of the International Speech Communication Association (Interspeech), September 2015.

81. Yi Liu, Yao Tian, Liang He, Jia Liu, Michael T Johnson, Simultaneous Utilization of Spectral Magnitude and Phase Information to Extract Supervectors for Speaker Verification, ASV Spoofing Challenge 2015, presented at Interspeech 2015, September 2015.
82. Gui Jin, Michael T Johnson, Jia Liu, Xiaokang Lin, Voice conversion based on Gaussian mixture modules with Minimum Distance Spectral Mapping, 2015 5th International Conference on Information Science and Technology (ICIST), April 2015, pp. 356-359.
83. Berry, Jeffrey, Cassandra North, and Michael T. Johnson, Dynamic aspects of articulating with a virtual vocal tract in dysarthria, International Conference on Motor Speech, 2014.
84. Berry, Jeffrey, A Kolb, C North, and Michael T. Johnson, Acoustic and Kinematic Characteristics of Vowel Production through a Virtual Vocal Tract in Dysarthria, Interspeech, Singapore, September 2014.
85. Berry, Jeffrey, John Jaeger, M Wiedenhoeft, B Bernal, and Michael T. Johnson, Consonant Context Effects on Vowel Sensorimotor Adaptation, Interspeech, Singapore, September 2014.
86. Ji, An, Michael T. Johnson, and Jeffrey Berry, Palate-referenced Articulatory Features for Acoustic-to-Articulator Inversion, Interspeech, Singapore, September 2014.
87. Jianglin Wang and Michael T. Johnson, "Physiologically-motivated Feature Extraction for Speaker Identification", International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2014, Florence, Italy.
88. Jeffrey Berry, Cassandra North, and Michael T. Johnson, "Sensorimotor adaptation of speech using real-time articulatory resynthesis", International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2014, Florence, Italy.
89. An Ji, Michael T. Johnson, and Jeffrey Berry, "The Electromagnetic Articulography Mandarin Accented English (EMA-MAE) Corpus of Acoustic and 3D Articulatory Kinematic Data", International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2014, Florence, Italy.
90. Jeffrey Berry, An Ji, and Michael T. Johnson, "Dynamic aspects of vowel production in Mandarin-accented English", American Speech-Language-Hearing Association 2013, Chicago, IL, November 2013.
91. Jeffrey Berry, Christine Belchel, Cassandra North, and Michael T. Johnson, "Learning novel articulatory-acoustic mappings in dysarthria", American Speech Language Hearing Association 2013, Chicago, IL, November 2013.
92. Jianglin Wang and Michael T. Johnson, "Vocal source features for bilingual speaker identification", China SIP, July 2013, Beijing China.
93. An Ji, Michael T. Johnson, Jeffrey Berry, "Articulatory space calibration in 3D Electromagnetic Articulography", China SIP, July 2013, Beijing China.
94. Jeffrey Berry, Cassandra North, Benjamin Meyers, Michael T. Johnson, Speech sensorimotor learning through a virtual vocal tract, Proceedings of Spring 2013 Meetings on Acoustics, Montreal, June 2013.
95. An Ji, Jeffrey Berry, Michael T. Johnson, Vowel production in Mandarin accented English and American English: Kinematic and acoustic data from the Marquette

University Mandarin accented English corpus, Proceedings of Spring 2013 Meetings on Acoustics, Montreal, June 2013.

96. Jianglin Wang, MT Johnson, "Residual Phase Cepstrum Coefficients with Application to Cross-lingual Speaker Verification", Interspeech 2012, Portland, September 2012.
97. Trawicki, Marek B. and Michael T. Johnson. "Improvements of the Beta-Order Minimum Mean-Square Error (MMSE) Spectral Amplitude Estimator using Chi Priors," Interspeech 2012, Portland, September 2012.
98. An Ji, Michael T. Johnson, Jeffrey Berry, "Tracking articulator movements using orientation measurements", 2012 International Conference on Audio, Language and Image Processing (ICALIP), Shanghai, China, July 2012.
99. Marek B. Trawicki, Michael T. Johnson, An Ji, Tomasz S. Osiejuk, "Multichannel speech recognition using distributed microphone signal fusion strategies", 2012 International Conference on Audio, Language and Image Processing (ICALIP), Shanghai, China, July 2012.
100. Jianglin Wang, An Ji, Michael T. Johnson, "Features for phoneme independent speaker identification", 2012 International Conference on Audio, Language and Image Processing (ICALIP), Shanghai, China, July 2012.
101. Michael T. Johnson and Patrick Clemins, "Individual Identification using Hidden Markov Models for Population Monitoring and Assessment", Invited Presentation to Special Session "Topical Meeting on Signal Processing of Subtle and Complex Acoustic Signals in Animal Communication", Acoustical Society of America Spring 2010 Meeting, Baltimore, Maryland, April 2010.
102. Marek Trawicki and Michael T. Johnson, "Optimal Distributed Microphone Phase Estimation," International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2009, Taiwan, April 2009, pp 2177-2180.
103. Yao Ren and Michael T. Johnson, "Auditory Coding based Speech Enhancement", International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2009, Taiwan, April 2009, pp 4685-4688.
104. Marek B. Trawicki, Yao Ren, Michael T. Johnson, Tomasz S. Osiejuk, Distributed Multi-Channel Speech Enhancement of Ortolan Bunting (*Emberiza Hortulana*) Vocalizations, Acoustic Communication by Animals, Corvallis, Oregon, 2008, pp. 276-277.
105. Kuntoro Adi, Michael T. Johnson, Tomasz S. Osiejuk Hidden Markov Model (HMM) based animal acoustic censusing, Acoustic Communication by Animals, Corvallis, Oregon, 2008, pp. 3-4.
106. Ebenezer Otu-Nyarko, Peter Scheifele, Michael Darre, Michael Johnson, Classification of stressful vocalizations of captive laying chickens using the Hidden Markov Model (HMM), Acoustic Communication by Animals, Corvallis, Oregon, 2008, pp. 176-177.
107. Kristine Sonstrom, Peter Scheifele, Michael Darre, Michael Johnson, The classification of vocalizations to identify social groups of beluga whales in the St. Lawrence River Estuary using the Hidden Markov Model, Acoustic Communication by Animals, Corvallis, Oregon, 2008, pp 251-252.

108. Kuntoro Adi, Kristine E. Sonstrom, Peter M. Scheifele, Michael T. Johnson, "Unsupervised validity measures for vocalization clustering," International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2008, April 2008, pp 4377-4380.
109. Yao Ren, Michael T. Johnson, "An Improved SNR estimator for speech enhancement," International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2008, April 2008, pp 4901-4904.
110. Mohamed A. Mneimneh, Michael T. Johnson, Richard J. Povinelli, "A Heart Cell Model for the Identification of Myocardial Ischemia," International Conference on Health Informatics, Funchal Madeira, Portugal, January 2008.
111. Z. W. Slavens, R. S. Hinks, J. A. Polzin, and M. T. Johnson, "Improved MR Image Magnification by Generalized Interpolation of Complex Data," Proceedings of the 15th Annual Meeting of ISMRM, Berlin, Germany, (Abstract #1887), May 2007.
112. Craig A. Struble, Richard J. Povinelli, Michael T. Johnson, Dina Berchanskiy, Jidong Tao, Marek Trawicki, "Combined Conditional Random Fields and n-Gram Language Models for Gene Mention Recognition, Proceedings of the Second BioCreative Challenge Evaluation Workshop, Madrid, Spain, April 2007.
113. Xi Li, Jidong Tao, Michael T. Johnson, Joseph Soltis, Anne Savage, Kirsten M. Leong,, John D. Newman, Stress and Emotion Classification using Jitter and Shimmer Features, International Conference on Acoustics Speech and Signal Processing 2007 (ICASSP07), Honolulu, Hawaii, April 2007, pp 1081-1084.
114. Mohamed A. Mneimneh, Richard J. Povinelli, Michael T. Johnson " An Integrative Approach for the Measurement of QT interval" Computers in Cardiology, Valencia, Spain, September 2006, pp 329-332.
115. Mohamed A. Mneimneh, Edwin E. Yaz, Michael T. Johnson, Richard J. Povinelli "Adaptive Kalman Filter Approach for the Baseline Removing Baseline Wandering " Computers in Cardiology, Valencia, Spain, September 2006, pp. 253-256.
116. Richard J. Povinelli, Mohamed A. Mneimneh, Michael T. Johnson "Cardiac Model Based Approach to QT Estimation" Computers in Cardiology, Valencia, Spain, September 2006, pp. 333-336.
117. Patrick Clemens, Marek B. Trawicki, Kuntoro Adi, Jidong Tau, and Michael T. Johnson, "Generalized Perceptual Features for Vocalization Analysis across Multiple Species", International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2006, Toulouse, France, May 2006.
118. Kuntoro Adi, Michael T. Johnson, Cepstral moment normalization for robust individual identification of ortolan bunting (*emberiza hortulana L*), presented at the Spring 2006 meetings of the Acoustical Society of America, May 2006, Providence, RI
119. Anthony D. Ricke, Richard J. Povinelli, Michael T. Johnson. "Segmenting Heart Sound Signals," Computers in Cardiology, Leon, France, September 2005.
120. Marek Trawicki and Michael T. Johnson, "Automatic Song-Type Classification and Speaker Identification of Norwegian Ortolan Bunting (*Emberiza Hortulana*) Vocalizations", IEEE International Conference on Machine Learning in Signal Processing (MLSP), Mystic, Connecticut, September 2005.

121. Patrick J. Clemens and Michael T. Johnson, "Unsupervised Classification of Beluga Whale Vocalizations", Spring 2005 Meetings of the Acoustical Society of America, Vancouver, May 2005.
122. Kevin M. Indrebo, Richard J. Povinelli, Michael T. Johnson. "Third-Order Moments of Filtered Speech Signals For Robust Speech Recognition," International Conference on Non-Linear Speech Processing (NOLISP) 2005, Barcelona, Spain, 151-157.
123. Jidong Tao and Michael T. Johnson, "Comparison and Evaluation of Animal Vocalization Enhancement Techniques", Fall 2004 Meetings of the Acoustical Society of America, San Diego, November 2004.
124. Kuntoro Adi and Michael T. Johnson, "Automatic Song-type classification and individual identification of Ortolan Bunting (*Emberiza Hortulana* L) vocalizations", Fall 2004 Meetings of the Acoustical Society of America, San Diego, November 2004.
125. Heather E. Ewalt and Michael T. Johnson, "Combining Multi-source Wiener Filtering with Parallel Beamformers to Reduce Noise from Interfering Talkers", International Conference on Signal Processing (ICSP) 2004, Beijing, August 2004.
126. Kevin M. Indrebo, Richard J. Povinelli, and Michael T. Johnson, "A comparison of reconstructed phase spaces and cepstral coefficients for multi-band phoneme classification", ICSP 2004, Beijing, August 2004.
127. Patrick Clemens and Michael T. Johnson, "Generalized perceptual features for animal vocalization classification", Spring 2004 Meetings of the Acoustical Society of America, New York, May 2004.
128. Andrew Lindgren, Michael T. Johnson, and Richard J. Povinelli, Joint Frequency Domain and Reconstructed Phase Space Features for Speech Recognition", International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2004, Montreal, May 2004.
129. Mike Zimmerman, Richard J. Povinelli, Michael T. Johnson, and Kris Ropella, "A Reconstructed Phase Space Approach for Distinguishing Ischemic from Non-Ischemic ST Changes using Holter ECG Data", Computers in Cardiology, Thessalonica, Greece, September, 2003.
130. Patrick Clemens and Michael T. Johnson, "Automatic Classification of African Elephant (*Loxodonta africana*) Follicular and Luteal Rumbles", 1st International Conference on Acoustic Communication by Animals, Baltimore, Maryland, July 2003, pp 81-82.
131. Patrick Clemens and Michael T. Johnson, "Automatic Type Classification and Speaker Identification of African Elephant (*Loxodonta Africana*) Vocalizations", Spring 2003 Meetings of the Acoustical Society of America, Nashville, May 2003.
132. Heather E. Ewalt and Michael T. Johnson, "Multiple Speech Signal Enhancement using a Microphone Array", Spring 2003 Meetings of the Acoustical Society of America, Nashville, May 2003.
133. Kevin M. Indrebo, Richard J. Povinelli, and Michael T. Johnson, "A Combined Sub-band and Reconstructed Phase Space Approach to Phoneme Classification", Non-Linear Speech Processing (NOLISP) 2003, Le Croisic, France, May 2003.

134. Jinjin Ye, Michael T. Johnson, and Richard J. Povinelli, "Study of attractor variation in the reconstructed phase space of speech signals, NOLISP 2003, Le Croisic, France, May 2003.
135. Jinjin Ye, Michael T. Johnson, and Richard J. Povinelli, "Phoneme classification over reconstructed phase space using principal component analysis", NOLISP 2003, Le Croisic, France, May 2003.
136. Xiaolin Liu, Richard J. Povinelli, and Michael T. Johnson, "Vowel classification by global dynamic modeling", NOLISP 2003, Le Croisic, France, May 2003.
137. Andrew C. Lindgren, Michael T. Johnson, and Richard J. Povinelli, "Speech Recognition using Reconstructed Phase Space Features", Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2003, Hong Kong, April 2003, pp I-60 – I-63.
138. Patrick Clemins and Michael T. Johnson, "Application of Speech Recognition to African Elephant (*Loxodonta Africana*) Vocalizations", Proceedings of ICASSP 2003, Hong Kong, April 2003, pp I-484 – I-487.
139. Michael T. Johnson, Andrew C. Lindgren, Richard J. Povinelli, and Xiaolong Yuan, "Performance of Nonlinear Speech Enhancement using Phase Space Reconstruction", Proceedings of ICASSP 2003, Hong Kong, April 2003, pp I-872 – I-875.
140. Xiaolin Liu, Richard Povinelli, and Michael T. Johnson, "Detecting Determinism in Speech Phonemes", IEEE Digital Signal Processing Workshop 2002, October 2002.
141. Jinjin Ye, Richard J. Povinelli, and Michael T. Johnson, "Phoneme Classification Using naïve Bayes Classifier in Reconstructed Phase Space", IEEE Digital Signal Processing Workshop 2002, October 2002.
142. Richard J. Povinelli, Michael T. Johnson, Nabeel A.O. Demerdash, and John F. Bangura, "A Comparison of Phase Space Reconstruction and Spectral Coherence Approaches for Diagnostics of Bar and End-Ring Connector Breakage and Eccentricity Faults in Polyphase Induction Motors using Motor Design Particulars", IEEE Industry Applications Society meetings 2002, October 2002.
143. Richard J. Povinelli, Felice M. Roberts, Kristina M. Ropella, and Michael T. Johnson, "Are Nonlinear Ventricular Arrhythmia Characteristics Lost, As Signal Duration Decreases?", Computers in Cardiology 2002, September 2002.
144. Patrick Clemins and Michael T. Johnson, "Automatic Speech Recognition and Speaker Identification of Animal Vocalizations", Measuring Behavior 2002, August 2002, pp 41-46.
145. Patrick Clemins, Heather Ewalt and Michael Johnson, "Time-aligned SVD Analysis for Speaker Identification", International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2002, May 2002, p 4160.
146. Michael T. Johnson and Leah H. Jamieson, "Temporal Features for Broadcast News Segmentation", ISCA Workshop on Prosody in Speech Understanding and Recognition, October 2001.
147. Michael T. Johnson and Mary P. Harper, "Near Minimal Weighted Word Graphs for Post-processing Speech", International Workshop on Automatic speech Recognition and Understanding (ASRU), December 1999.

148. M.P. Harper, M.T. Johnson, L.H. Jamieson, S.A. Hockema, and C.M. White, "Interfacing a CDG Parser with an HMM Word Recognizer Using Word Graphs", Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 1999, March 1999, pp 733-736.
149. Michael T. Johnson, Leah H. Jamieson, and Mary P. Harper "Interfacing Acoustic Models with Natural Language Processing Systems", Proceedings of the International Conference on Spoken Language Processing (ICSLP) 1998, December 1998, pp 2419-2422.
150. A.M. Supranant, S.L. Hura, M.P. Harper, L.H. Jamieson, G. Long, S.M. Thede, A. Rout, T.H. Hsueh, S.A. Hockema, M.T. Johnson, J.B. Laflen, P. Srinivasa, and C.M. White, "Familiarity and Pronouncibility of Nouns and Names: The Purdue Proper Name Database", Spring 1998 Meetings of the Acoustical Society of America, June 1998.
151. Michael Johnson and Mita Desai, "An Adaptive Approach for Texture Modelling", Proceedings of the International Conference on Image Processing (ICIP) 1994, November 1994.
152. Harold Longbotham, Michael Johnson, Jack Harris, and Redouan Rouzky, "The FatBear: a nonarithmetic pico filter", Proceedings of the SPIE, Image Algebra and Morphological Image Processing IV, July 1993, pp 128-139.

Selected Invited Presentations

- Keynote Speaker, Weld Penetration Identification Based on Convolutional Neural Network, The 2019 International Workshop on Intelligentized Welding Manufacturing, University of Kentucky, Lexington, Kentucky, November 2019.
- Invited Colloquium Speaker, Acoustic to Articulatory Inversion for Computer-Aided Pronunciation Training, University of Texas at San Antonio Department of Electrical and Computer Engineering Colloquium Series, April 2019.
- Keynote Speaker, IEEE Joint Louisville-Lexington Chapters Holiday Dinner, Louisville, Kentucky, December 2018.
- Invited speaker, Shanghai Jiaotong University Colloquium Series, "Electromagnetic Articulography for Speaker Independent Acoustic-to-Articulatory Inversion, April 2015.
- Invited speaker, University of Memphis Institute of Intelligent Systems Colloquium Series, "Applying Speech Technology to Animal Vocalizations", April 2014.
- Invited speaker, NIMBioS Investigative Workshop on Analyzing Animal Vocal Sequences, "Using HMMs to model vocal structure and sequence for classification and identification", October 2013.
- Featured Keynote Speaker, GE Healthcare Technology symposium, "Innovations and Applications of Artificial Intelligence and Bayesian Modeling", August 2011.
- "Individual Identification using Hidden Markov Models for Population Monitoring and Assessment", Invited Presentation to Special Session "Topical Meeting on Signal Processing of Subtle and Complex Acoustic Signals in Animal Communication", at Acoustical Society of America Spring 2010 Meeting, Baltimore, Maryland, April 2010.
- Featured speaker at ShARE World Seminar 2008, "Innovation in Computation", Tsinghua University, Beijing China, December 2008.

- Keynote speaker for annual Tau Beta Pi banquet, “The Dr. Dolittle Project: Applying Speech Processing to Animal Vocalizations,” November 2007.
- Presentation to Beijing Institute of Technology annual colloquia, “The Dr. Dolittle Project: Applying Speech Processing to Animal Vocalizations,” Beijing, China, May 2007.
- Presentation to Tsinghua University Center for Speech Technology, “The Dr. Dolittle Project: Applying Speech Processing to Animal Vocalizations,” Beijing, China, May 2007.
- Presentation to the Chinese Academy of Sciences, “The Dr. Dolittle Project: Applying Speech Processing to Animal Vocalizations,” Institute of Acoustics, Beijing, China, May 2007.
- Michael T. Johnson, “The Dr. Dolittle Project: Applying speech technology to animal vocalizations”, Sigma Xi annual banquet, April 27, 2007.
- Michael T. Johnson, “Speech Research at Marquette – The Dr. Dolittle Project”, Marquette Trustees Committee on Academic Affairs and Planning, March 3, 2004.
- Michael T. Johnson, “Speech Recognition and Chaos Theory: A new approach to signal analysis”, Lynde Bradley Science Club, Rockwell Automation, November 11, 2003.
- Michael T. Johnson, “The Dr. Dolittle Project: Applying human speech processing algorithms to other species”, UW Milwaukee Computer Science Department Colloquium, November 7, 2003.
- Michael T. Johnson, “Talking with the Animals: Automatic Classification and Identification of Animal Vocalizations using Speech Technology”, Keynote Speaker for Physics Club of Milwaukee Annual Banquet, June 5, 2002.
- Michael T. Johnson, “Speech Processing Research at Marquette (Elephants and Cocktail Parties)”, COE National Research Council meeting, April 26, 2002.
- Michael T. Johnson, “Speech and Signal Processing Research at Marquette”, Invited speaker at Sigma Xi annual meeting, February 20, 2002.

PhD Dissertation and MS Thesis

- Michael T. Johnson, "Incorporating Prosodic Information and Language Structure into Speech Recognition Systems", PhD Dissertation, Purdue University School of Electrical and Computer Engineering, August 2000.
- Michael T. Johnson, "An Adaptive Texture Model Based on Markov Random Fields", Masters Thesis, University of Texas at San Antonio, August 1994.

Student PhD Dissertations (primary advisor)

- Soleymanpour, Mohammad, “Synthesizing Dysarthric Speech Using Multi-speaker TTS for Dysarthric Speech Recognition”, PhD Dissertation, University of Kentucky, August 2022.
- Narjes Bozorg, “Articulatory Wavenet: Deep Autoregressive Model for Acoustic-to-Articulatory Inversion”, PhD Dissertation, University of Kentucky, December 2020.

- An Ji, “Speaker Independent Acoustic-to-Articulatory Inversion”, PhD Dissertation, Marquette University, December 2014.
- Jianglin Wang, "Physiologically-motivated feature extraction methods for speaker recognition", PhD Dissertation, Marquette University, December 2013.
- Jidong Tao, “Acoustic Model Adaptation for Automatic Speech Recognition and Animal Vocalization Classification”, PhD Dissertation, Marquette University, May 2009.
- Marek Trawicki, “Distributed Multichannel Processing for Signal Enhancement”, PhD Dissertation, Marquette University, May 2009.
- Kuntoro Adi, “Hidden Markov model based animal acoustic censusing: learning from speech processing technology”, Ph.D. Dissertation, Marquette University, May 2008.
- Patrick J. Clemens, “Automatic Classification of Animal Vocalizations”, PhD Thesis, Marquette University, May 2005.

Student MS Theses (primary advisor)

- Subash Khanal, “Mispronunciation Detection and Diagnosis in Mandarin Accented English speech, Masters Thesis, University of Kentucky, May 2020.
- Jose, Neenu, “Speaker and Gender Identification using Bioacoustic Data Sets”, Masters Thesis, University of Kentucky, May 2018.
- Vonderhaar, Joe, “Speaker-specific Adaptation of Maeda Synthesis Parameters for Auditory Feedback”, Masters Thesis, Marquette University, May 2017.
- Jones, Deriq, “Development of Kinematic Templates for Automatic Pronunciation Assessment using Acoustic-to-Articulatory Inversion”, Masters Thesis, Marquette University, May 2017.
- Kelly Vonderhaar, Analysis of Interference between Electromagnetic Articulography and Electroglottograph Systems, Masters Thesis, Marquette University, August 2016.
- Xiangyu Zhou, Least-squares Mapping from Kinematic data to Acoustic Synthesis Parameters for Rehabilitative Acoustic Learning, Masters Thesis, Marquette University, May 2016.
- Andrew Kolb, Development of Software Tools and Analysis methods for the Use of Electromagnetic Articulography Data in Speech Research”, Masters Thesis, Marquette University, May 2015.
- Zac Slavens, “Generalized interpolation applied to MR image magnification and gradient nonlinearity correction”, Masters Thesis, Marquette University, May 2008.
- Xi Li, “SPEech Feature Toolbox (SPEFT) Design, and Emotional Speech Feature Extraction”, M.S. Thesis, Marquette University, August 2007.
- Anthony D. Rieke, “Automatic Frame Length, Frame Overlap, and Hidden Markov Model Topology for Automatic Speech Recognition of Animal Vocalizations”, M.S. Thesis, Marquette University, December 2006.
- Jinjin Ye, “Speech Recognition Using Time Domain Features from Phase Space Reconstructions”, M.S. Thesis, Marquette University, May 2004.
- Franck Hounkpevi, “Triphone Creation Through Rule-based Trajectory Interpolation for Continuous Speech Recognition”, M.S. Thesis, Marquette University, May 2003.

- Xiaolong Yuan, “Auditory Model-Based Bionic Wavelet Transform for Speech Enhancement”, M.S. Thesis, Marquette University, May 2003.
- Andrew Lindgren, “Speech Recognition using Features extracted from Phase Space Reconstructions”, M.S. Thesis, Marquette University, May 2003.
- Heather Ewalt, “Speech Signal Enhancement Using a Microphone Array”, M.S. Thesis, Marquette University, December 2002.

Selected News and Magazine Articles

- “Listening to the Animal Kingdom”, Discover Magazine, Marquette University, 2008, pp 14-15.
- “Decoding Calls of the Wild”, by Mark Johnson, feature article in Milwaukee Journal Sentinel, part of “Brainpower: Groundbreaking Thinkers in Wisconsin”, October 2008.
- “Say What?”, cover article on Dr. Dolittle Project, Weekly Reader Science, Spring 2007.
- “Animal Talk”, by Elena Cabral, Scholastic News Edition 5/6, December 18, 2006.
- “Animal “Speech” Project Aims to Decode Critter Communication”, by Maryann Mott, National Geographic Online, <http://news.nationalgeographic.com/news/2006/09/060926-dolittle-project.html>, September 2006.
- “Parsing the Puffin’s Patois”, article on Dr. Dolittle Project by Rachel Metz, Wired News online, July 2006.
- “Quacking the Code”, by Katie Pelech, article in Milwaukee Magazine, March 2006, p. 30.
- “What Jumbo tells Dumbo” by Alex Antunes, feature article in Computers in Science and Engineering (CISE) magazine, published by IEEE Computer Society, September-October 2005, pp 3-6.
- “Discoveries with Disney”, article about collaborative project with Disney’s Animal Kingdom published in MU Magazine, Fall 2003, p. 17.

GRANTS

Major Grants Funded

1. Michael T. Johnson (PI), Aaron Cramer (co-PI), Dan Ionel (co-PI), and Janet Lumpp (co-PI), GAANN Fellowship Program in Electrical and Computer Engineering at the University of Kentucky, Department of Education, 10/1/2019 – 9/30/2022. Total Budget \$600,702 over 3 years. Supplement of \$51,000 in 2020 for 1 additional fellow.
2. Jeffrey Berry (PI) and Michael T. Johnson (co-PI), “Speech Sensorimotor Control in Non-progressive Dysarthria”, National Institutes of Health R-15, 7/1/2018 – 6/30/21. Total Budget \$446,956 over 3 years (\$166,428 to University of Kentucky as subcontract to Marquette University).
3. Michael T. Johnson (PI) and Jeffrey Berry (co-PI), “RI: Small: Speaker Independent Acoustic-Articulator Inversion for Pronunciation Assessment”, National Science Foundation CISE Directorate, 2013. Total Budget \$449,643 over 3 years.
4. Michael T. Johnson (PI) and Jeffrey Berry (co-PI), “EAGER: Acoustic-Articulator Modeling for Pronunciation Analysis”, National Science Foundation CISE Directorate, 2011. Total Budget: \$149,896.00 over 1 year.
5. Michael T. Johnson (PI), Elizabeth Mugganthalder (Fauna Communications Research Institute), Peter Scheifele (National Undersea Research Center, University of Connecticut), Mike Darre (Animal Sciences, University of Connecticut), Anne Savage (Disney Animal Kingdom), “The Dr. Dolittle Project: A Framework for Classification and Understanding of Animal Vocalizations”, NSF IT-R program under CISE Directorate, 2003. Total Budget: \$1,200,000 over 4 years.
6. Michael T. Johnson (PI) and Richard J. Povinelli, “Integration of Stochastic and Dynamical Methods for Speech Technology”, NSF IT-R program under CISE directorate, 2001. Total Budget: \$360,000 over 3 years.
7. Richard J. Povinelli (PI), Nabeel A. O. Demerdash, Edwin Yaz, and Michael T. Johnson, “A Novel Approach to Fault Modeling, Diagnostics, and Prediction in Motor Drive Systems”, NSF Division of Controls, Networks, and Computational Intelligence, 2003. Total Budget: \$340,000 over 3 years.

Additional Grants Funded

1. Peter Scheifele (PI), Leesa Scheifele (co-PI), and Michael T. Johnson (co-PI), Measuring Hearing Acuity and Vocalization Signals of the Asian Small-Clawed Otter (*Amblonyx cinereus*) in Response to Anthropogenic Noise, Indianapolis Zoo, March 2017, Funded \$14,300.
2. Kevin McGowan (PI) and Michael T. Johnson (co-PI), Apparatus and Procedure for Ultrasound Investigation of Velopharyngeal port aperture, University of Kentucky Igniting Research Collaborations project, Funded \$27,726 in 2018-2019.
3. Jeffrey Berry (PI) and Michael T. Johnson (co-PI), “Speech Rehabilitation Using a Virtual Vocal Tract”, Marquette Regular Research Grant (RRG). Funded \$6,000 in 2014-15.
4. Michael T. Johnson (PI) and Jeffrey Berry (co-PI), Research Experiences for Undergraduates (REU) supplement to “RI: Small: Speaker Independent Acoustic-Articulator Inversion for Pronunciation Assessment”, NSF REU program, CISE directorate. Funded \$6,000 in 2013-2014, \$6000 in 2014-2015, \$6000 in 2015-2016.
5. Michael T. Johnson (PI) and Jeffrey Berry (co-PI), OISE International supplement to “EAGER: Acoustic-Articulator Modeling for Pronunciation Analysis”, co-funded by CISE and OISE directorates. Funded \$19,840 in 2012-2013.

6. Michael T. Johnson (PI) and Jeffrey Berry (co-PI), Research Experiences for Undergraduates (REU) supplement to “EAGER: Acoustic-Articulator Modeling for Pronunciation Analysis”, NSF REU program, CISE directorate. Funded \$8,000 in 2011-2012.
7. Richard J. Povinelli (PI) and Michael T. Johnson (PI), “An examination of phoneme confusability in spoken English”, Marquette Regular Research Grant (RRG). Funded \$15,000 in 2011-2012.
8. Michael T. Johnson (PI) et. al., Research Experiences for Undergraduates (REU) supplement to “The Dr. Dolittle Project”, NSF REU program, CISE directorate. Funded at \$6,000 in 2003-04, \$15,000 in 2004-2005, \$12,000 in 2005-2006, \$15,000 in 2006-2007 Total Budget: \$48,000.
9. Michael T. Johnson (PI) and Richard J. Povinelli, REU supplement to “Integration of Stochastic and Dynamical Methods for Speech Technology”, NSF REU program, CISE directorate. Funded at \$15,000 in 2001-02, 2002-03, and 2003-04, Total Budget: \$45,000.
10. Richard J. Povinelli (PI), Nabeel A. O. Demerdash, Edwin Yaz, and Michael T. Johnson, REU supplement to “A Novel Approach to Fault Modeling, Diagnostics, and Prediction in Motor Drive Systems”, NSF REU Program, CNCI Division. Funded at \$6,000 in 2003-04 and 2004-05. Total Budget to-date: \$6,000.

PROFESSIONAL ACTIVITIES

Technical Committees and Publication editing

- Area Editor, Speech Enhancement, Speech Communication Journal, 2014-present
- Member, IEEE Signal Processing Society Speech and Language Technical Committee (SLTC), 2014-2017

Selected Publication Review Participation

- Speech Communication Journal
- International Conference on Acoustics, Speech and Signal Processing
- Interspeech IEEE Automatic Speech and Recognition Workshop
- Workshop on Spoken Language Technology
- IEEE Transactions on Speech and Audio Processing
- Journal of the Acoustical Society of America
- IEEE Sensors Journal
- IEEE Transactions on VLSI
- IEEE Transactions on Signal Processing
- IEEE Signal Processing Letters
- IEEE Transactions on Neural Systems
- Journal of Applications of Signal Processing
- Animal Behavior Journal
- EURASIP Journal on Audio, Speech and Music Processing
- International Journal on Adaptive Control and Signal Processing
- Biosystems Engineering Journal
- Acoustics Research Letters Online
- International Conference on Decision and Control
- Midwest Symposium on Circuits and Systems
- Iasted Circuits and Systems Conference
- Workshop on Signal Processing and Applications
- American Controls Conference
- Journal of Zhejiang University Science C (Computers & Electronics)
- Journal of Medical Engineering & Physics
- Open Signal Processing Journal
- IEEE Spoken Language Technology workshop
- International Conference on Signal Processing (ICSP)
- Journal of Computers
- International Journal of Electronics and Communications
- Franklin Institute journal

Selected Funding Review Participation

- National Science Foundation
Have reviewed for CCLI, CAREER, CONICYT, and ITR programs, as well as various regular programs in the CISE and Biological Sciences directorate
- International: Austrian Science Fund, Icelandic Research Fund
- Medical College of Wisconsin Clinical and Translational Science Institute (CTSI)

Professional and honorary society memberships

- ECE Department Heads Association (ECEDHA), 2016-present
ECEDHA Awards Committee, 2018-present
- Southeast ECEDHA (SECEDHA), 2016-present
Secretary 2019-2020, Vice President 2020-2022, President 2022-2023
- Southeastern Center for EE Education (SCEEE) Board of Directors, 2020-2023
- Senior Member IEEE, membership 1994-present
- IEEE Signal Processing Society, 1998-present
- Acoustical Society of America (ASA), 2001-present
- Eta Kappa Nu, 1996-present
- Upsilon Pi Epsilon, 2001-present
- Senior Member Sigma Xi, 2001-2012
- Association of Computer Machinery (ACM), 1998-2010
- IEEE Computer Society, 1998-2010
- IEEE Circuits and Systems Society, 2001-2009
- Association of Computational Linguistics (ACL), 1998-2002
- International Speech Communication Association (ISCA), 2001-2006

Other professional activities and awards

- Selected for SEC Academic Leadership Development Program, 2019-2020
- Featured in “Groundbreaking Thinkers in Wisconsin” article series 2007
- Researcher of the Year, Marquette College of Engineering, 2006-2007
- Young Engineer of the Year, Engineers and Scientists of Milwaukee (ESM), 2006
- Registered Professional Engineer, State of Wisconsin

TEACHING AND EDUCATIONAL ACTIVITIES

Courses Taught

Undergraduate

EECE Freshman Seminar
Computer Hardware
Digital Logic Design
Linear Systems Analysis
Embedded Systems Design
Circuits 1
Signals and Systems Lab
Introduction to Engineering
Understanding Leadership

Graduate and Joint Undergraduate/Graduate

Digital Signal Processing
Speech Processing
Optimal/Adaptive Signal Processing
Pattern Recognition
Analysis of Algorithms
Information Theory
Professional Research Writing

Teaching awards

Marquette Eta Kappa Nu (HKN) honor society EECE Teacher of the Year, 2014, 2005

Selected Teaching, Education, and Pedagogy activities

UK Global Engagement Academy, Spring 2022
Center for Enhancement of Learning and Teaching (CELT) workshop, Spring 2021
UK Counseling Center, ECE faculty meeting guest on Student Mental Health, Fall 2021
UK Unconscious Bias Seminar, Fall 2019
UK Experienced Leadership Academy, Fall 2018
UK CoE Teaching Community of Practice, Fall 2017
UK Chairs Academy, Fall 2016
Chair, CoE Teaching Best Practices Committee, Spring 2017
Coordinator for Marquette EECE Freshman Seminar, 2011-2015
Participant in KEEN/Lafferty Teaching Development Workshop, Fall 2015, Spring 2016
Mentored numerous Research Experience for Undergraduates (REU) through NSF grants
International Teaching Experience, Tsinghua University, 2008-2009, 2011, 2013, 2014-2015
Signals and Systems Concept Inventory site coordinator for NSF study, 2003-2006
Developed and gave numerous DSP/Audio Workshops for middle and high school outreach
Participated with IEEE Signal Processing Education Technical Committee activities
Developed introductory DSP module for General Engineering freshman course
Participated in MU Assessment seminar by Barbara Wolvoord, May 2005
Participated in NSF Engineering Education Scholars (EES) program for new faculty
Participated in Prentice Hall Symposium on Education, Chicago, IL, 2001
Participated in NSF/ASEE Visiting Scholars Teaching Workshop, Fall 2000 and Spring 2001
Member, Advisory committee MU Center on Teaching and Learning (2005-2010)

Student Mentoring Activities (2000-present)

MS and PhD Committees
Undergraduate Research Projects
Senior Design Projects
Individual Student Mentoring

ACADEMIC SERVICE

Academic Service at the University of Kentucky

University:

- Steering Committee for Wildcat Foundations first-year experience program (2017-2019)
- Represented UK in SEC Academic Leadership Development Program (ALDP), 2019-2020

College:

- Chair, Engineering Technology Faculty Committee (2020 – present)
ET Chair search, LST Faculty search (2020-2021), CPT Faculty search (2021-2022)
- CoE Funkhouser (2019, 2021, 2023) and Grehan (2018-19) Building Committees
- CoE Recruitment Advisory Committee (2019-present)
- CoE Search Committee for Senior Director of Philanthropy (2018)
- Director, CoE Scholars in Engineering Leadership Program (2017-present)
- CoE Search Committee for Director of Marketing and Communications (2017)
- Chair, CoE Teaching Best Practices committee (Spring 2017)
- CoE Chair's Committee (2016-present, Chair 2018-2019)

Departmental:

- ECE Department Chair (2016-present)

Academic Service at Marquette University

University:

- University Board of Graduate Studies (2009-2016; chair, 2012-2016)
- University Search Committee, Vice Provost for Enrollment Management (2015-2016)
- University Steering committee on academic integrity (2012-2014)
- University Subcommittee on academic integrity (2011)
- COF, Subcommittee on Nominations and Elections (chair, 2004-2007).
- Faculty mentor for Dissertation Boot Camps (2008, 2010)
- Advisory Committee for Center on Teaching and Learning (2005-2010)

College and Departmental:

- COE Research Activity Committee (2011)
- COE Dean of Engineering Search Committees (2003, 2009)
- COE Freshman Programs Committee (2006-2007)
- COE Discovery Towers Building Committee (2004-2005)
- BME Biocomputer Program Advisory Board (2001-2016)
- EECE Graduate Committee (2000-2003, 2005-2016)
- EECE Undergraduate Committee (2003-2008, 2012-2015)
- EECE Director of Graduate Studies (2009-2012)
- EECE Department Grad Program Assessment Coordinator (2009-2012)
- Numerous EECE faculty search and other subcommittees

Selected public service and outreach activities

- UK CoE E-day, every spring annually since 2016
- UK CoE and ECE prospective student tours, several times per month since 2016

- Guest speaker, Triangle UK Leadership Academy students, July 2023
- UK ECE Middle School STEM program, November 2018
- UK ECE STEM Outreach camp for girls, June 2016, June 2017
- Guest speaker, Marquette Girls Who Code outreach program, March 2016.
- Coordinator for Marquette EECE open house activities, Fall 2015.
- Faculty advisor for Marquette's HKN chapter and Sigma Phi Delta engineering fraternity
- Guest speaker, MOTIF English corner, Beijing, China. "Diversity", September 2014, "Humor", March 2015, "Languages of the World", June 2015.
- Senior Visiting Scholar, Tsinghua University, China 2014-15, 2008-09, Summer 2011, Summer 2013
- COE iPalm2 outreach program to high school students, August 2011, 2012, 2013
- Represented Marquette COE on visits to Harbin Institute of Technology, China, 2011
- Panelist for Marquette University Advancement workshop, 2011
- Marquette ORSP Brown Bag panelist "One thing led to another" research series, 2010
- Spoke to students at Rosenwald Dunbar elementary school in Nicholasville, KY, about the Dolittle project and careers in engineering, 2008
- Spoke to students at Cardinal Valley elementary school in Lexington, KY, about the Dolittle project and careers in engineering, 2006
- Participated in Marquette University High School Junior Career Day, 2005
- Spoke to "Future Problem Solvers" gifted program at Templeton Middle School about Artificial Intelligence, 2005
- Participated in Innovation Mondays meetings with local industry, 2004-2005
- COE representative for "We are Marquette" World Café Conversation, 2004
- EECE representative and presenter for GEMS Advanced Education Fair, 2004
- Spoke at MSOE about graduate school opportunities in EECE, 2002-2003
- Presented at Lynde Bradley Science Club, Rockwell Automation, 2003
- Presented at UW Milwaukee Computer Science Department Colloquium, 2003
- Marquette ORSP Brown Bag panelist on Pre-proposal Contact with Sponsors, 2003
- Keynote on Global Research, Marquette COE National Advisory Council, 2002
- Keynote Speaker at Physics Club of Milwaukee annual banquet, 2002
- Spoke at UWM Elementary Education class on careers in science and technology, 2002
- Participated in Nathan Hale H.S. Career Day regarding careers in engineering, 2002
- Active participant in Marquette EECE and COE open houses, 2000-2015
- Blackjack dealer for Marquette engineering student council casino nights, 2000-15
- Regular participant in Marquette parent lunch program, 2000-2015
- Member, Eastbrook Church & worship choir, 2001-2016
- Numerous student and student group activities, both UK and Marquette, 2000-present
- Numerous research talks, colloquia, and seminars, 2000-present
- Numerous other volunteer and community activities, 2000-present

LITIGATION EXPERIENCE

US District Court

Snap-On Technologies, Inc. vs. Hunter Engr. Company, 98-C-369, 98-C-837, 00-C-929, 00-C-1511 (2002)

Jurisdiction: USDC Eastern District of Texas
Client: Hunger Engineering Company (Defendant)
Law Firm: Kirkland and Ellis LLP
Activities: Declaration, Deposition

EVS Codec Tech LLC and Saint Lawrence Comm. LLC vs. ZTE and ZTE USA Inc., 3:19-cv-00385 (2019)

Jurisdiction: USDC Northern District of Texas
Client: ZTE Inc. (Defendant)
Law Firm: K&L Gates LLP
Activities: Source Code Review, Technical consulting services

Nuance Communications, Inc. vs. MModal LLC., 1:17-cv-01484-MN (2019)

Jurisdiction: USDC District of Delaware
Client: MModal LLC (Defendant)
Law Firm: Duane Morris, LLP and Latham & Watkins, LLP
Activities: Reports (3), Source Code Review, Deposition

VoiceAge EVS LLC vs. Apple Inc, 1:20-cv-01061-CFC (2020)

Jurisdiction: USDC District of Delaware
Client: Apple Inc. (Defendant)
Law Firm: WilmerHale LLP
Activities: Source Code Review, Technical consulting services

Freshub vs. Amazon Inc., 1:19-cv-00885-ADA (2020)

Jurisdiction: USDC Western District of Texas
Client: Amazon (Defendant)
Law Firm: Fenwick and West LLP
Activities: Reports (3), Source Code Review, Depositions (2), Trial Testimony

VB Assets, LLC vs. Amazon.com, Inc., 1:19-cv-01410-MN (2021)

Jurisdiction: USDC District of Delaware
Type of Litigation: Patent Infringement
Client: Amazon (Defendant)
Law Firm: Fenwick and West LLP
Activities: Reports (2), Depositions (2), Technical Consulting (currently active)

Vocalife LLC vs. Sonos, Inc., 2:21-cv-00129 (2021)

Jurisdiction: USDC Eastern District of Texas
Client: Sonos (Defendant)
Law Firm: Lee, Sullivan, Shea, and Smith LLP
Activities: Declaration

Jawbone Innovations, LLC vs. Meta Platforms, Inc., 6-23-cv-00158 (2023)

Jurisdiction: USDC Western District of Texas
Client: Meta Platforms (Petitioner)
Law Firm: Allen & Overy LLP
Activities: Technical Consulting (currently active)

Patent Trial and Appeal Board (PTAB) Inter Partes Review (IPR)

LG Electronics vs. Saint Lawrence Communication, IPR2015-01874, IPR2015-01875 (2015)

Client: LG Electronics (Petitioner)

Law Firm: Mayer Brown LLP

Activities: Declarations (2)

ZTE Corp. vs. Saint Lawrence Communications, IPR2016-00704, IPR2016-00705 (2016)

Client: ZTE Inc. (Petitioner)

Law Firm: Finnegan, Henderson, Farabow, Garrett & Dunner LLP

Activities: Declarations (2), Deposition

Google LLC vs. Koninklijke Philips, IPR2017-00437 (2017)

Client: Koninklijke Philips (Patent Owner)

Law Firm: Fitzpatrick, Cella, Harper, & Scinto

Activities: Declaration, Deposition

Google LLP vs. Sonos, Inc., IPR2021-00475 (2022)

Client: Sonos (Patent Owner)

Law Firm: Ackerman LLP and Lee, Sullivan, Shea, and Smith LLP

Activities: Declaration

Sonos, Inc. vs. Google LLP, IPR2022-01592, 2022-01594, 2023-0118, 2023-0119 (2022)

Client: Sonos (Petitioner)

Law Firm: Lee, Sullivan, Shea, and Smith LLP

Activities: Declarations (4) (currently active, all 4 IPRs instituted)

Sonos, Inc. vs. Google LLP, IPR2023-(TBD), 2023-(TBD) (2023)

Client: Sonos (Petitioner)

Law Firm: Lee, Sullivan, Shea, and Smith LLP

Activities: Technical consulting (currently active)

International Trade Commission (ITC)

Google LLP vs. Sonos, Inc. 337-TA-1329, 337-TA-1330 (2022)

Client: Sonos (Respondent)

Law Firm: Lee, Sullivan, Shea, and Smith LLP

Activities: Technical Consulting, 4 reports, 2 depositions, hearing testimony (currently active)

Other Jurisdictions

Google LLP vs. Sonos, Inc. Docket T-952-20 (2021)

Jurisdiction: Canadian Federal Court

Type of Litigation: Patent Infringement

Client: Sonos (Defendant)

Law Firm: ROBIC, LLP and Lee, Sullivan, Shea, and Smith LLP

Activities: Reports (2), Source Code Review, Trial Testimony