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. 11,024,311

DECLARATION OF DR. MICHAEL T. JOHNSON IN SUPPORT OF PETITION FOR *INTER PARTES* REVIEW OF U.S. PATENT NO. 11,024,311

TABLE OF CONTENTS

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 1-3, 8-12, and 14-20 (the "Challenged Claims") of U.S. Patent No. 11,024,311 (the "'311 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. BACKGOUND 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 to this report 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. 11,024,311 (the "'311 Patent")

EXHIBIT	DESCRIPTION
1003	US Patent Publication 2017/0357478 (“Piersol”)
1004	US Patent Publication 2017/0083285 (“Meyers”)
1005	Japanese Publication No. 2003/223188 (“Masahide”)
1006	English translation of Masahide
1007	US Patent Publication 2013/0073293 (“Jang”)
1008	*Certain Audio Players and Components Thereof*, Inv. No. 337-TA-1330, Order No. 14
1009	US Patent No. 9,318,107 (“’107 Patent)
1010	Relevant Excerpts of ’107 Patent file history
1011	K. Paliwal, Filter-Bank Energies For Robust Speech Recognition (1999), *available at* https://maxwell.ict.griffith.edu.au/spl/publications/papers/euro99_kkp_fbe.pdf
1012	K. Paliwal , Filter-Bank Energies For Robust Speech Recognition (1999), *available at* https://www.semanticscholar.org/paper/FILTER-BANK-ENERGIES-FOR-ROBUST-SPEECH-RECOGNITION-Paliwal/a7777b5132f4b82430925190f8850d3d2090cbc6
1013	US Patent Publication 2014/0163978 (“Basye”)
1014	Reetika & Saini, Impact of Router’s Buffer Size on Packet Loss Due To Congestion in TCP, IJERT, Vol. 3 Issue 5 (May 2014), *available at* https://www.ijert.org/research/impact-of-routers-buffer-size-on-packet-loss-due-to-congestion-in-tcp-IJERTV3IS051297.pdf
1015	Krzyzanowski, Quality of Service (Sept. 29, 2012), *available at* https://people.cs.rutgers.edu/~pxk/417/notes/03-qos.html.
1016	Goetze and Appell, Voice activity detection driven acoustic event classification for monitoring in smart homes, IEEE Xplore (Dec. 2010), *available at* https://www.researchgate.net/publication/224215429_Voice_activity_detection_driven_acoustic_event_classification_for_monitoring_in_smart_homes
1017	Venter and Hanekom, Automatic detection of African elephant (Loxodonta africana) infrasonic vocalisations from recordings, ScienceDirect (2010).
1018	Miralles et al., SAMARUC a programmable system for passive acoustic monitoring of cetaceans (2013).

EXHIBIT	DESCRIPTION
1019	Solera-Urena *et al.*, Real-Time Robust Automatic Speech Recognition Using Compact Support Vector Machines, IEEE Transactions on Audio Speech and Language Processing (May 2012), *available at* https://www.researchgate.net/publication/228109902_Real-Time_ Robust_Automatic_Speech_Recognition_Using_Compact_Support_Vector_Machines.
1020	*Certain Audio Players and Components Thereof*, Inv. No. 337-TA-1330, Redacted excerpts of Google's Pre-Hearing Brief (May 31, 2023)
1021	*Certain Audio Players and Components Thereof*, Inv. No. 337-TA-1330, Redacted excerpts of Google's Post-Hearing Brief (July 10, 2023)
N/A	Google's Patent Owner's Response in IPR2022-01592 (June 30, 2023) ("IPR2022-01592 POR")

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 (here, Sonos) 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.

15. I also understand that an independent claim can be found invalid, even though a dependent claim to which it refers is valid.

16. I have considered these principles in forming my opinions contained in this Declaration.

B. Anticipation

17. 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.

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

19. 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

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

21. 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 priority 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.

22. 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 priority date of the claimed invention. I understand that the hypothetical POSITA is presumed to have knowledge of all pertinent prior art references.

23. 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.

24. 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.

25. 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.

26. 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.

27. 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 priority date of the claimed invention.

D. Claim Construction

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

29. 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.

30. 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.

31. 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.

E. Priority Date

32. I understand that a subject patent can assert priority to a patent application filed earlier than the subject patent.

33. However, I understand that such an assertion does not necessarily mean the subject patent is entitled to the date of the earlier-filed patent application.

34. In this regard, I understand that the subject patent is only entitled to the date of the earlier-filed patent application if the earlier-filed patent application sufficiently discloses the invention claimed in the subject patent in a manner compliant with 35 U.S.C. § 112(a).

35. I understand that § 112(a) requires that the earlier-filed patent application (i) reasonably conveys to one of ordinary skill in the art that the inventors

were in possession of the claimed subject matter (i.e., the "written description support" requirement) and (ii) enables a POSITA to practice the claimed invention without undue experimentation (i.e., the "enablement" requirement).

36. Lastly, I understand that broad disclosure of a given concept in an earlier-filed patent application often does not provide written description support for a narrower invention relating to that given concept.

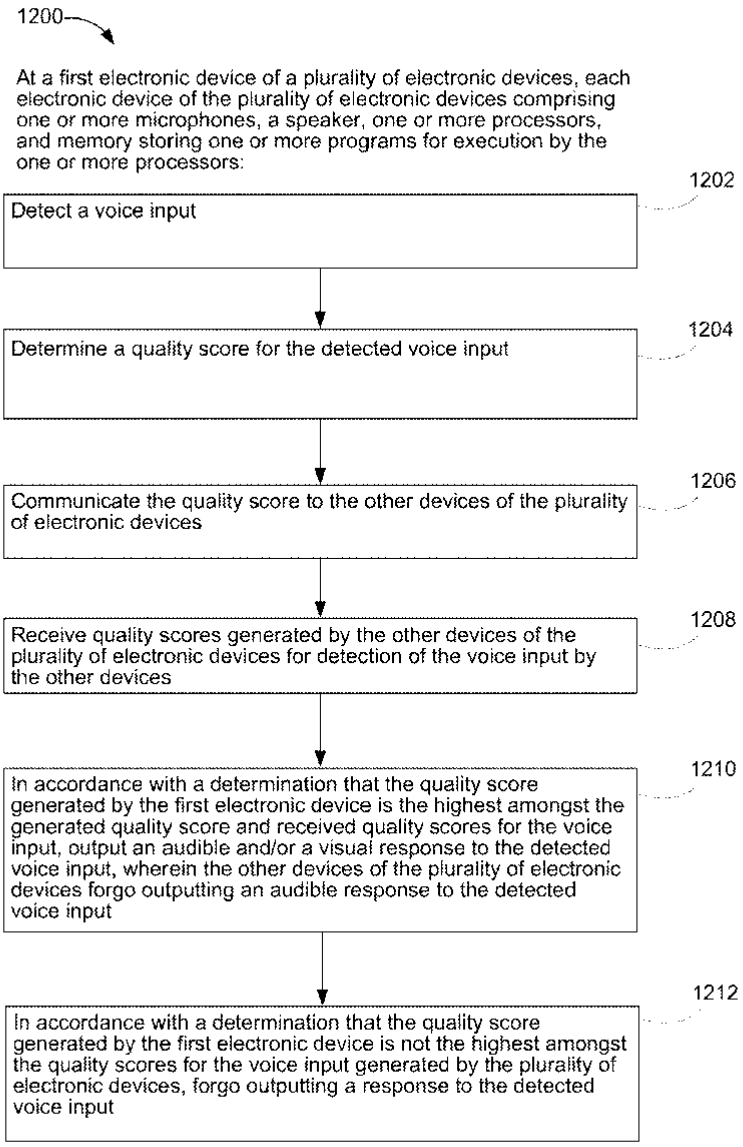
V. OVERVIEW OF THE '311 PATENT[1]

37. The '311 Patent "relates generally to voice interfaces and related devices, including but not limited to methods and systems for coordination amongst multiple voice interface devices." '311 Patent, 1:44-47.

38. More specifically, much like the prior art discussed below, the '311 Patent explains that "[a] location (e.g., a room or space within a home) may include multiple devices that include voice assistant systems" and that it is "desirable for there to be a leader amongst the voice assistant devices that is responsible for responding to the user's voice inputs, in order to reduce user confusion." *Id.*, 1:60-2:3; *see also, e.g.*, *id.*, 7:7-14.

39. To that end, the '311 Patent discloses "coordination" or "leadership negotiation" techniques amongst "voice-activated electronic devices" (or simply, "electronic devices"), which are best summarized by FIG. 6:

[1] All emphasis set forth herein has been added unless noted otherwise.



Id., FIG. 6; *see also, e.g., id.*, 20:55-22:20, 29:5-30:3.

40. As shown above, the disclosed "coordination" or "leadership negotiation" techniques seek to select a single "leader amongst themselves to respond to the user…." *Id.*, 7:4-14.

41. The ’311 Patent identifies a wide-range of functions that could be performed as part of an “electronic device” responding to a voice input, including:

- “generating and providing a spoken response to a voice command (e.g., speaking the current time in response to the question, ‘what time is it?’)”;
- providing “voice message responses to the user, where the response may include, for example, requests for additional information to fulfill the request, confirmation that the request has been fulfilled, notice that the request cannot be fulfilled, and so forth”;
- “output[ting] responses to the voice input and subsequent voice inputs”;
- “detect[ing] and process[ing] or pre-process[ing] subsequent voice inputs from the user (whether it be the leader processing the voice inputs and generating the responses, the leader pre-processing the voice inputs for transmission to the voice assistance server 112 which generates the responses, or the leader simply transmitting the voice inputs to the voice assistance server 112 which generates the response)”;
- “reading a news story or a daily news briefing prepared for the user”;
- “streaming media content requested by a user (e.g., ‘play a Beach Boys song’)”;
- “playing a media item stored on the personal assistant device or on the local network”;
- “changing a state or operating one or more other connected devices within the operating environment 100 (e.g., turning lights, appliances or media devices on/off, locking/unlocking a lock, opening windows, etc.)”; and
- “issuing a corresponding request to a server via a network 110”

Id., 7:50-63, 10:28-34, 22:4-20.

VI. PERSON OF ORDINARY SKILL IN THE ART

42. As discussed below, it is my opinion that the Challenged Claims of the ’311 Patent are not entitled to a priority date any earlier than October 3, 2016. In

other words, the effective priority date for the Challenged Claims is no earlier than October 3, 2016.

43. 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 '311 Patent, I considered these factors during my analysis of the '311 Patent and its file history, and I also considered my own experiences working and teaching in the field of speech processing.

44. Based on the foregoing, it is my opinion that a POSITA for the '311 Patent as of the effective priority date (i.e., no earlier than October 3, 2016) 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.

45. 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 field of invention and/or level of ordinary skill were slightly different.

VII. CLAIM CONSTRUCTION

46. 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 '311 Patent claim terms do not require construction and can be afforded their plain and ordinary meaning for purposes of this IPR.

A. "Detecting a Voice Input"

47. Each '311 claim recites *detecting a voice input from a user* (cls. 1, 10) or *detect a voice input* (cl. 16).[2]

48. Google contends that *detecting a voice input* (i) refers "to a process for identifying that an audio input includes voice" (IPR2022-01592 POR, p.12), (ii) "refers to a process separate from performing speech recognition (e.g., speech to text) on the voice input" (*id.*, p.13), and (iii) can be satisfied by "voice activity detection" or "hotword/wakeword detection" (*id.*, pp.3-9, 16). In other words, Google contends *detecting a voice input* requires identifying voice without

[2] For clarity, I have italicized claim language throughout this Declaration.

performing "speech recognition," which Google narrowly interprets as being a mutually exclusive process from hotword/wakeword detection.

49. For purposes of this IPR, I have been asked to assume that Google's interpretation of *detecting a voice input* is correct.

B. "Second Values… From Other Electronic Devices"

50. Each '311 claim recites *receiving second values associated with the voice input from other electronic devices of the plurality of electronic devices* (cls. 1, 10) or *receive respective second values from other electronic devices of the plurality of electronic devices* (cl. 16).

51. In the parties 1330 ITC dispute, Google successfully argued that each of these terms – *second values* and *other electronic devices* – "can be satisfied by a single device, score, or value." Ex.1008, p.35. In other words, Google successfully argued that the '311 claims only require a claimed *electronic device* to receive a single *second value* from a single *other electronic device*. *Id.*, p.39-40.

52. For purposes of this IPR, I have been asked to assume that Google's interpretation of *second values… from other electronic devices of the plurality of electronic devices* is correct.

C. Dependent Claim 11

53. Dependent claim 11 recites:

> **[11.0]** *The first electronic device of claim 10, the one or more programs further comprising instructions for:*

[11.1] *recognizing a command in the voice input; and*

[11.2] *in accordance with a determination that a type of the command is related to the first electronic device, responding to the detected voice input.*

54. Google contends that claim 11 "requires a single device programmed to both arbitrate based on a 'first value' (claim 10) and to 'respond[]' to a voice command in accordance with a determination that the command is related to a device." IPR2022-01592 POR, p.26. In other words, Google contends claim 11 requires a device be "configured to perform arbitration based on both a 'first value' and the 'relevance' of a command to" a device. *Id.*, p.26-27.

55. Yet, according to Google, claim 11 covers: (i) a device "***overriding*** the outcome of the arbitration based on scores based on a determination that a type of the command is related to the device" (*id.*, p.27) and (ii) the disclosure at 22:65-23:8 of the '311 Patent describing "relevance of the command" being used to "narrow leadership candidates" before arbitrating on "scores":

> In some implementations, multiple leadership determination criteria (e.g., quality scores, relevance of command, and state of the device, as described above) may be implemented as a multi-step leadership determination (e.g., ***determine relevance of the command*** and the device state ***to narrow leadership candidates, then determine leader based on quality scores***; determine device with highest score, then check if command relevance or device state criteria apply to the other devices) or as ***a weighted determination (e.g., each criterion is accounted for and weighted in a weighted score)***.

Id., p.26.

56. For purposes of this IPR, I have been asked to assume that Google's interpretation of claim 11 is correct.

D. "Confidence Level of Detection of the Voice Input"

57. Dependent claim 18 depends on claim 17. Dependent claim 17 recites *[t]he system of claim 16, wherein the first value comprises a first quality score of the voice input.* In turn, dependent claim 18 recites *[t]he system of claim 17, wherein the first quality score comprises a confidence level of detection of the voice input.*

58. In the parties 1330 ITC dispute, Google interpreted *a confidence level of detection of the voice input* broadly.

59. In this respect, the '311 specification only refers to *confidence level* in two passages. First, it baldly refers to "confidence level of detection" in one section of the specification titled "Example Process for Device Leadership Negotiation" as follows:

> In some implementations, the quality score includes a confidence level of detection of the voice input; the quality score is a confidence level value for the voice input sample.

'311 Patent, 30:44-46.

60. Second, the other reference to a "confidence level" states:

> Based on the comparing, the electronic device 190 determines (1106) that the first voice input corresponds to a first user of the plurality of users. For example, the user identification module 560 identifies a voice model in voice models data 550 that best matches the first voice input, and in accordance with the identification of the match determines that the user speaking the first voice input is the user to which the matching

> voice model corresponds. In some implementations, the user identification module 560 also determines a ***confidence level*** or some other similar measure of ***the quality or closeness of the match between a voice model and the voice input***, and identifies a match only if the match is the best and the ***confidence level*** is above a predefined threshold.

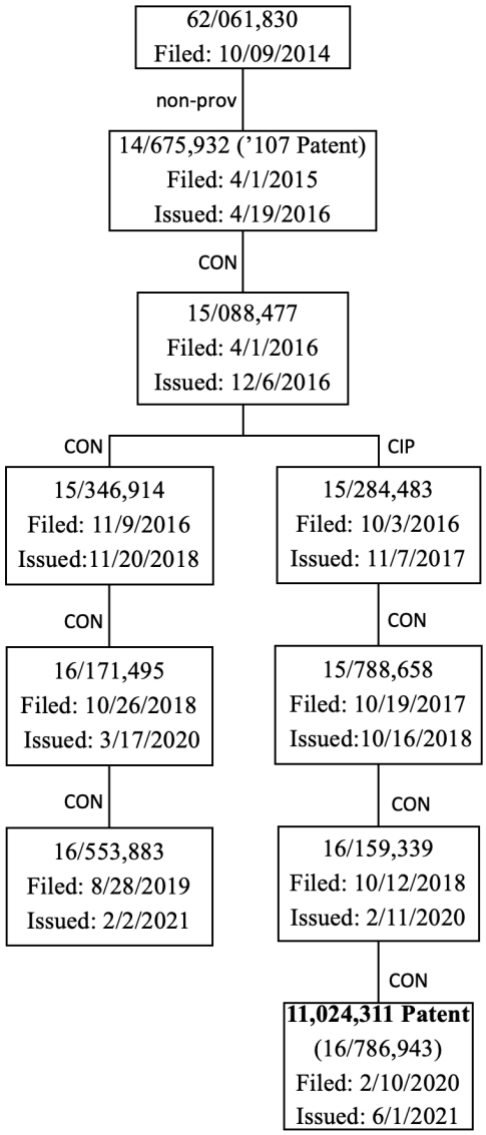
'128 Patent, 24:35-48.

61. Despite the '311 specification informing a POSITA that *a confidence level of detection of the voice input* refers to a likelihood or probability of the quality or closeness of the match between a voice model (e.g., a model the represents voice existing in sensed sound) and the voice input perceived by the device, Google says this term should not be so limited. For instance, Google argued in the 1330 ITC dispute that this term broadly refers to "a measure reflecting the confidence that the audio input is voice." Ex.1020, p.6 (document page 89). In other words, Google broadly contended *a confidence level of detection* "is a measure of the likelihood that an audio signal includes speech, and indicates a level of confidence that a voice input was detected." Ex.1021, p.6 (document page 70). For purposes of this IPR, I have been asked to assume that Google's interpretation *confidence level of detection* is correct.

VIII. EFFECTIVE PRIORITY DATE

62. The '311 Patent issued on June 1, 2021 from a patent application filed on February 10, 2020. As illustrated below, the '311 Patent is part of a complicated chain of patent filings ultimately claiming priority back to provisional patent

application number 62/061,830 (the "'830 Provisional") filed on October 9, 2014 and non-provisional patent application number 14/675,932 (the "'932 Application") filed on April 1, 2015, which issued as US Patent 9,318,107 (the "'107 Patent"). Of note, the '311 Patent is part of a continuation-in-part (CIP) branch of the family tree that developed on October 3, 2016 when non-provisional patent application number 15/284,483 (the "'483 CIP Application") was filed.



63. The ’311 Patent’s specification (which is purportedly the same as the specification of the ’483 CIP Application) differs significantly from the ’107 Patent’s specification.

64. Based on my analysis of the ’107 Patent’s specification, it is my opinion that the Challenged Claims of the ’311 Patent are not entitled to the priority date of the ’107 Patent because its specification does not support Google’s narrow interpretation of *detecting a voice input*. To the contrary, it is my opinion that Google’s narrow interpretation of *detecting a voice input* is inconsistent with the ’107 Patent’s teachings for at least two reasons.

65. ***First***, the ’107 Patent clearly explains that *detecting a voice input* simply refers to a device perceiving or sensing a voice input via the device’s microphone and does not require any sort of “process[ing],” much less “process[ing] for identifying that an audio input includes voice.” IPR2022-01592 POR, p.12.

66. In this respect, the ’107 Patent states that a user’s “utterance” (which is *a voice input*) is “detected” simply by a device’s microphone perceiving or sensing the utterance; any “process[ing]” occurs after the voice ***has been detected***:

- “In general, system 100 illustrates a user 102 speaking an utterance 104 that is ***detected by microphones*** of computing devices 106, 108, and 110.” ’107 Patent, 3:56-58.
- “When the user 102 speaks the utterance 104, ‘OK computer,’ each computing device that has a microphone in the vicinity of the user 102 ***detects and processes*** the utterance 104. Each computing device ***detects***

the utterance 104 ***through*** an audio input device such as ***a microphone***." '107 Patent, 5:3-7.

67. ***Second***, the '107 Patent clearly explains that hotword/wakeword detection is a form of "speech recognition." In other words, hotword/wakeword detection and speech recognition are not mutually exclusive processes at least in the context of the '311 Patent.

- "In an example environment, the ***hotword*** used to invoke the system's attention are the words 'OK computer.' Consequently, ***each time*** the words 'OK computer' are spoken, it is picked up by a microphone, conveyed to the system, which ***performs speech recognition*** techniques ***to determine whether the hotword was spoken*** and, if so, awaits an ensuing command or query." '107 Patent, 1:58-64.

68. It is notable that the Examiner of the '107 Patent confirmed that the '107 Patent's specification discloses that hotword/wakeword detection is a form of "speech recognition."

69. In this regard, during prosecution, Google amended its pending claims in its August 20, 2015 Office Action Response to distinguish hotword detection from "automated speech recognition," as shown in relevant part below:

1. (Currently amended) A computer-implemented method comprising:

receiving, by a first computing device, audio data that corresponds to an utterance;

determining, based on an analysis of the audio data and without performing automated speech recognition processing on the audio data, a first value ~~corresponding to~~ that reflects a first likelihood that the utterance includes a particular hotword;

Ex.1010, p.2.

70. In the subsequent October 1, 2015 Office Action, the Examiner correctly pointed out that there was no written description support in the '107 Patent's specification for distinguishing hotword detection from "speech recognition." Specifically, the Examiner stated the following:

> **NOTE**: During current analysis of the present invention it is realized that ***contradictory to the claim language***, ***the specification in fact states that speech recognition is performed***, which from a technical standpoint is ***almost inherent*** given the nature of command and control. ***In order to spot a keyword some sort of speech recognition must be performed***. There is no indication of waveform analysis that would suggest otherwise in the specification. Therefore being consistent with both the spec and technological field and in view of prior art Rosenberger, it is believed the claims mean that audio input is processed with ASR but segmented into two portions, where a first portion has speech recognition performed on it, however the remaining portion of the audio is not processed until probabilistic weighted values are compared between devices. The remaining portion of audio therefore does not have ASR performed on it as consistent with the claims. For instance a user states "Hello thermostat" where SR is performed, the following command afterwards "turn up 2 degrees" is not processed until a decision is made between the devices to process the remaining command. This is precisely taught in newly incorporated Rosenberger.

Id., p.15.

71. The Examiner also cited paragraphs in the '107 Patent's specification that directly contradicted Google's attempt to distinguish hotword detection from "automated speech recognition":

> Consequently, each time the words "OK computer" are spoken, it is picked up by a microphone, conveyed to the system, which performs speech recognition techniques to determine whether the hotword was

> spoken and, if so, awaits an ensuing command or query. Accordingly, utterances directed at the system take the general form [HOTWORD] [QUERY], where "HOTWORD" in this example is "OK computer" and "QUERY" can be any question, command, declaration, or other request that can be speech recognized, parsed and acted on by the system, either alone or in conjunction with the server via the network.

Id., pp.16-17 (emphasis original).

72. In the After-Final Response dated December 7, 2015, Google did not attempt to traverse the Examiner's interpretation of the '107 Patent's specification but instead, amended the claims, as shown in relevant part below:

1. (Currently amended) A computer-implemented method comprising:

receiving, by a first computing device, audio data that corresponds to an utterance;

before beginning automated speech recognition processing on the audio data, processing the audio data using a classifier that classifies audio data as including a particular hotword or as not including the particular hotword;

determining, based on ~~an analysis of the audio data and without performing automated speech recognition processing on the audio data~~ the processing of the audio data using the classifier that classifies audio data as including a particular hotword or as not including the particular hotword, a first value that reflects a first likelihood that the utterance includes ~~a particular~~ the particular hotword;

receiving a second value that reflects a second likelihood that the utterance includes the particular hotword, as determined by a second computing device;

~~transmitting, to the second device, the first value that reflects the first likelihood that the utterance includes the particular hotword;~~

comparing the first value that reflects the first likelihood that the utterance includes the particular hotword and the second value that reflects the second likelihood that the utterance includes the particular hotword; and

based on comparing the first value that reflects the first likelihood that the utterance includes the particular hotword to the second value that reflects the second likelihood that the utterance includes the particular hotword, determining whether to ~~perform~~ begin performing automated speech recognition processing on the audio data.

Id., p.33; *see also id.*, pp.36-37, 39-40 (applying same amendments to other independent claims).

73. From the Examiner's October 1, 2015 Office Action and the language of this amendment, a POSITA would appreciate that this amended claim ***cannot*** be interpreted to cover detecting a hotword ***without*** performing speech recognition.

74. ***First***, the Examiner's October 1, 2015 Office Action expressly explained with specific citations to the '107 Patent's specification that the prior claim language attempting to recite hotword detection "without performing automated speech recognition processing on the audio data" contradicted the '107 Patent's disclosures. Google changed the claim language to get around this glaring problem identified by the Examiner.

75. ***Second***, a POSITA would readily appreciate that the claimed "classifier" utilizes speech recognition and that this amendment in the context of the '107 Patent requiring "[a] classifier that classifies audio data as including a particular hotword or as not including the particular hotword" merely refers to limited-vocabulary speech recognition – that is, speech recognition limited to identifying a particular hotword, as opposed to large-vocabulary speech recognition.

76. In this regard, Google added the following dependent claims in the After-Final Response dated December 7, 2015, each identifying specific speech recognition techniques for the claimed "classifier" of independent claim 1:

23. (New) The method of claim 1, wherein processing the audio data using a classifier that classifies audio data as including a particular hotword or as not including the particular hotword comprises:

extracting filterbank energies or mel-frequency cepstral coefficients from the audio data.

24. (New) The method of claim 1, wherein processing the audio data using a classifier that classifies audio data as including a particular hotword or as not including the particular hotword comprises:

processing the audio data using a support vector machine or a neural network.

Ex. 1010, p.41

77. These amendments appear to correspond to the following disclosure from the '107 specification:

> The audio subsystem provides the processed audio data to a hotworder. The hotworder compares the processed audio data to known hotword data and computes a confidence score that indicates the likelihood that the utterance 104 corresponds to a hotword. The hotworder may extract audio features from the processed audio data such as ***filterbank energies or mel-frequency cepstral coefficients***. The hotworder may use ***classifying windows*** to process these audio features such as by using ***a support vector machine or a neural network***. Based on the processing of the audio features, the hotworder 124 computes a confidence score of 0.85, hotworder 126 computes a confidence score of 0.6, and hotworder 128 computes a confidence score of 0.45.

'107 Patent, 5:12-28.

78. As of the '107 Patent's earliest possible priority date (Oct. 9, 2014), a POSITA would have readily understood that extracting "filterbank energies or mel-frequency cepstral coefficient" and "processing [] audio data using a support vector machine or a neural network" to classify speech data as containing a hotword or not

refer to well-known ***speech recognition*** techniques. *See, e.g.*, Ex.1011, p.1, 4 (1999[3] publication stating "Mel frequency cepstral coefficients (MFCCs) are perhaps the most widely used features for speech recognition," stating "filter-bank energies (FBEs)" "have been used in 50's, 60's and 70's for speech recognition," and showing "FBEs perform at least as good as (and sometimes even better than) the MFCCs for robust speech recognition."); Ex.1019, p.2:

> SVMs [support vector machines] offer several advantages over artificial neural networks (ANNs) that have attracted the attention of the speech processing community. Nevertheless, their high computational requirements prevent them from being used in practice in ***automatic speech recognition (ASR)***, where ***ANNs have proven to be successful***…. Support vector machines (SVMs) have shown superior performance than ANNs in a variety of tasks…. [S]everal authors have suggested the ***use of SVMs in ASR*** ….

Id., p.4 ("[T]hese characteristics make SVMs a promising future alternative to Gaussian mixture models and artificial neural networks for the problem of acoustic modeling in robust speech recognition."); *see also, e.g.*, Meyers, ¶26-27:

> In some cases, a keyword spotter may use simplified ***ASR (automatic speech recognition) techniques***. For example, an expression detector may use a Hidden Markov Model (HMM) recognizer that performs acoustic modeling of the audio signal and compares the HMM model of the audio signal to one or more reference HMM models that have been created by training for a specific trigger expression…. In practice, an HMM recognizer may produce multiple feature scores, corresponding to different features of the HMM models. An expression detector may use ***a support vector machine (SVM) classifier*** that receives the one or more feature scores produced by the HMM recognizer. The SVM classifier produces a confidence score indicating

[3] Ex. 1012 (identifying paper was published in 1999).

> the likelihood that an audio signal contains the trigger expression. The confidence score is compared to a confidence threshold to make a final decision regarding whether a particular portion of the audio signal represents an utterance of the trigger expression.

79. ***Third***, in the context of the '107 Patent, the clause "before beginning automated speech recognition processing on the audio data" in the second limitation and "determining whether to begin performing automated speech recognition processing on the audio data" in the final limitation ***does not*** mean that "processing the audio data using a classifier that classifies audio data as including a particular hotword or as not including the particular hotword" is done without speech recognition.

80. Instead, these claimed aspects are merely consistent with the specification's ***repeated*** disclosure of the voice-enabled device first performing speech recognition to detect a hotword in the user's utterance and then initiating speech recognition on the speech in the utterance ***following*** the hotword. *See, e.g.*, '107 Patent, 1:58-2:3, 3:67-4:3, 6:24-30, 6:50-55, 6:56-63, 7:2-4, 7:27-29, 7:46-49, 7:56-62, 8:3-15, 9:28-33, 10:66-11:2, 11:51-54.

81. In other words, these claimed aspects are merely consistent with the specification's disclosure of ***first*** performing limited-vocabulary speech recognition to identify a hotword and ***then*** performing large-vocabulary speech recognition on the remainder of the utterance following the hotword.

82. Considering the Examiner's rejection, a POSITA's knowledge, and the '107 Patent expressly stating "***each time*** the [hotword is] spoken, it is picked up by a microphone, conveyed to the system, which ***performs speech recognition*** techniques ***to determine whether the hotword was spoken*** and, if so, awaits an ensuing command or query" ('107 Patent, 1:58-64), a POSITA would conclude that the claimed "classifier" utilizes speech recognition.

83. Accordingly, it is my opinion that the earliest effective priority date for the Challenged Claims of the '311 Patent is October 3, 2016, which is the date when the '483 CIP Application was filed.

IX. OVERVIEW OF THE PRIOR ART

A. Overview of Piersol

84. US Patent Publication 2017/0357478 ("Piersol") is titled "Intelligent Device Arbitration and Control," was filed on September 16, 2016, and claims priority to a provisional filing on June 11, 2016. Ex.1003. Thus, Piersol qualifies as prior art to the '311 Patent, which has an effective priority date no earlier than October 3, 2016, for the reasons I explained above.

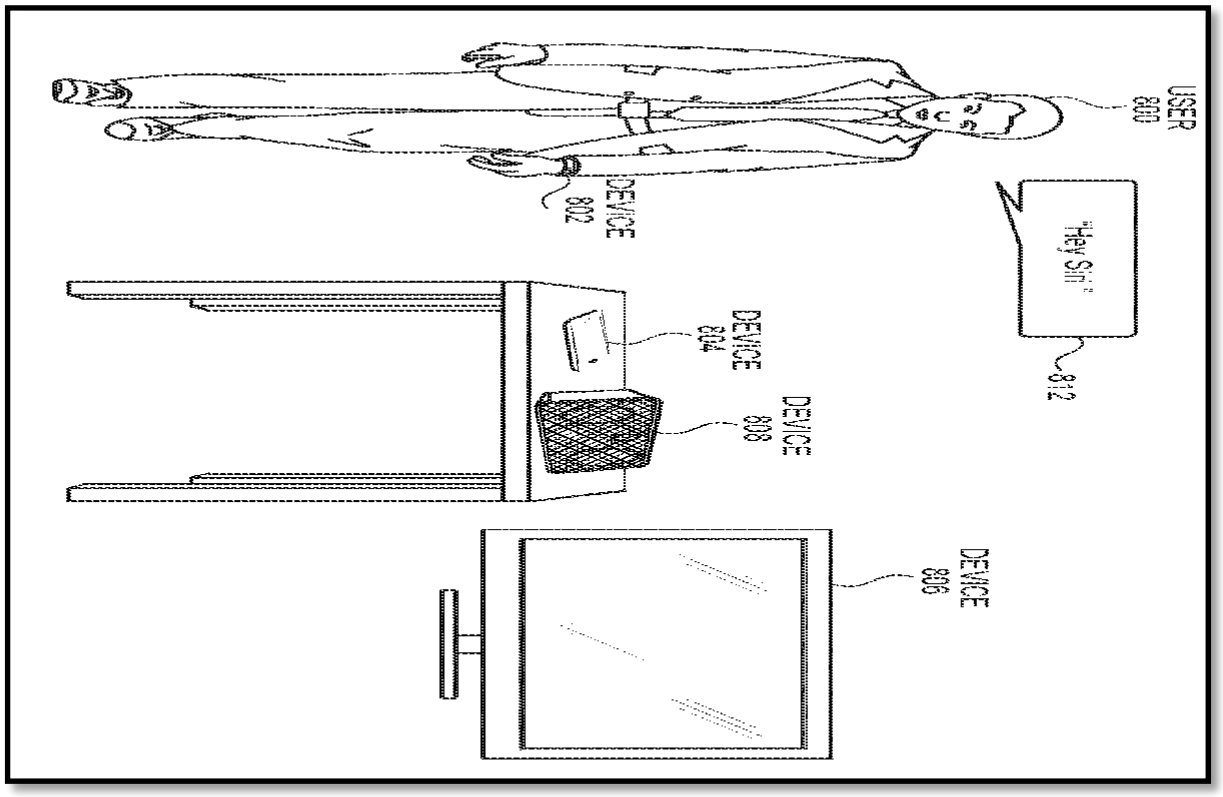
85. Piersol is generally directed at addressing the problem where multiple "electronic devices have virtual assistant services" in a system detect the same voice input from a user containing a request and multiple voice-enabled devices handle the

user's request "result[ing] in a confusing experience for the user" and/or "duplicative or conflicting operations" being performed:

> Many contemporary electronic devices offer virtual assistant services to perform various tasks in response to spoken user inputs. In some circumstances, multiple electronic devices having virtual assistant services may concurrently operate in a shared environment. As a result, a user input may cause each of the multiple electronic devices to respond when the user input contains a trigger phrase or command recognized by each of the virtual assistant services on the electronic devices. This in turn may result in ***a confusing experience for the user***, as multiple electronic devices may ***simultaneously begin to*** listen and/or ***prompt for additional input***. Further, the multiple electronic devices may ***perform duplicative or conflicting operations*** based on the same user input.

Id., ¶3; *see also id.*, ¶33, 295.

86. Piersol provides the following illustration of a system comprising multiple "electronic devices" that detect the same voice input from a user:



Id., FIG. 8B.

87. Piersol summarizes its disclosed solution as follows:

> In one example process, a first electronic device samples an audio input using a microphone. The first electronic device broadcasts a first set of one or more values based on the sampled audio input. Furthermore, the first electronic device receives a second set of one or more values, which are based on the audio input, from a second electronic device. Based on the first set of one or more values and the second set of one or more values, the first electronic device determines whether to respond to the audio input or forego responding to the audio input.

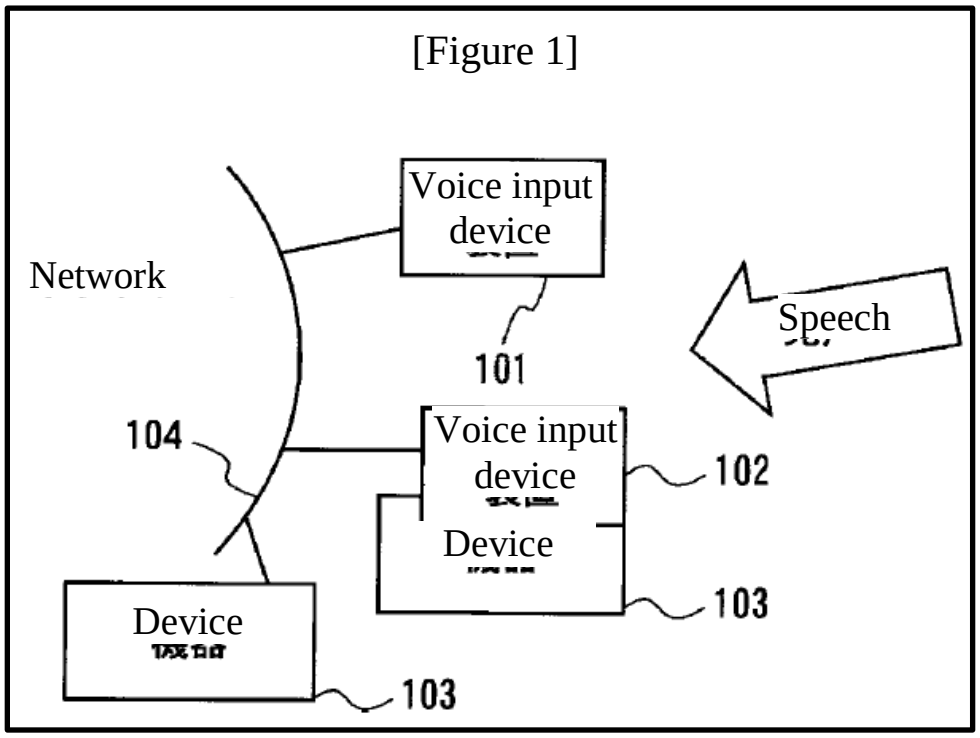
Id., Abstract. In this way, Piersol discloses techniques whereby, "[i]n response to detecting [a] spoken instruction [], electronic devices 802-808 initiate an arbitration process to identify (e.g., determine) an electronic device for responding to the spoken instruction [] from user 800." *Id.*, ¶247.

B. Overview of Masahide

88. Japanese Patent App. Pub. No. JP2003223188 ("Masahide") was filed on January 29, 2002 and published on August 8, 2003. Ex.1006. Thus, Masahide qualifies as prior art to the '311 Patent, which has an ***earliest possible*** priority date of October 9, 2014.

89. Masahide generally relates to "devices that handle voice, and in particular, to voice input systems, voice input methods, and voice input programs where the speech of a user may be input into ***a plurality*** of voice inputs." *Id.*, ¶1. Masahide recognized that, when multiple voice input devices are located in a room, it may be necessary for the user to specify a particular target device for the voice input, and additionally, voice input may need to be suppressed for devices other than the intended voice input device. *Id.*, ¶2. Accordingly, Masahide seeks to "solve[] the question of what to do with each voice input device when the voice of the user is input to ***a plurality*** of networked voice input devices." *Id.*, ¶11.

90. Masahide provides the following illustration of a system comprising multiple "voice input devices" that detect the same voice input from a user:



Id., FIG. 2; *see also id.*, ¶4, 11, 13-14.

91. Masahide summarizes its disclosed solution as follows:

> [T]he voice input method of the present invention is characterized by detecting voice information input into a plurality of voice input devices connected to a network, respectively, and sharing information regarding said voice detected by each voice input device connected to the network with other voice input devices via the network as determinative information, with each voice input device connected to the network using the determinative information in the applicable voice input device itself, and other voice input devices to determine processing and execution of the voice information.

Id., ¶5.

92. Masahide explains that the "determinative information" used to perform arbitration and select one voice input device to interact with the user can take various forms, including signal-to-noise ratio (SNR) information. *See, e.g.*, *id.*, ¶49-51, FIG. 6, ¶52, 54-56, FIG. 7.

C. Overview of Meyers

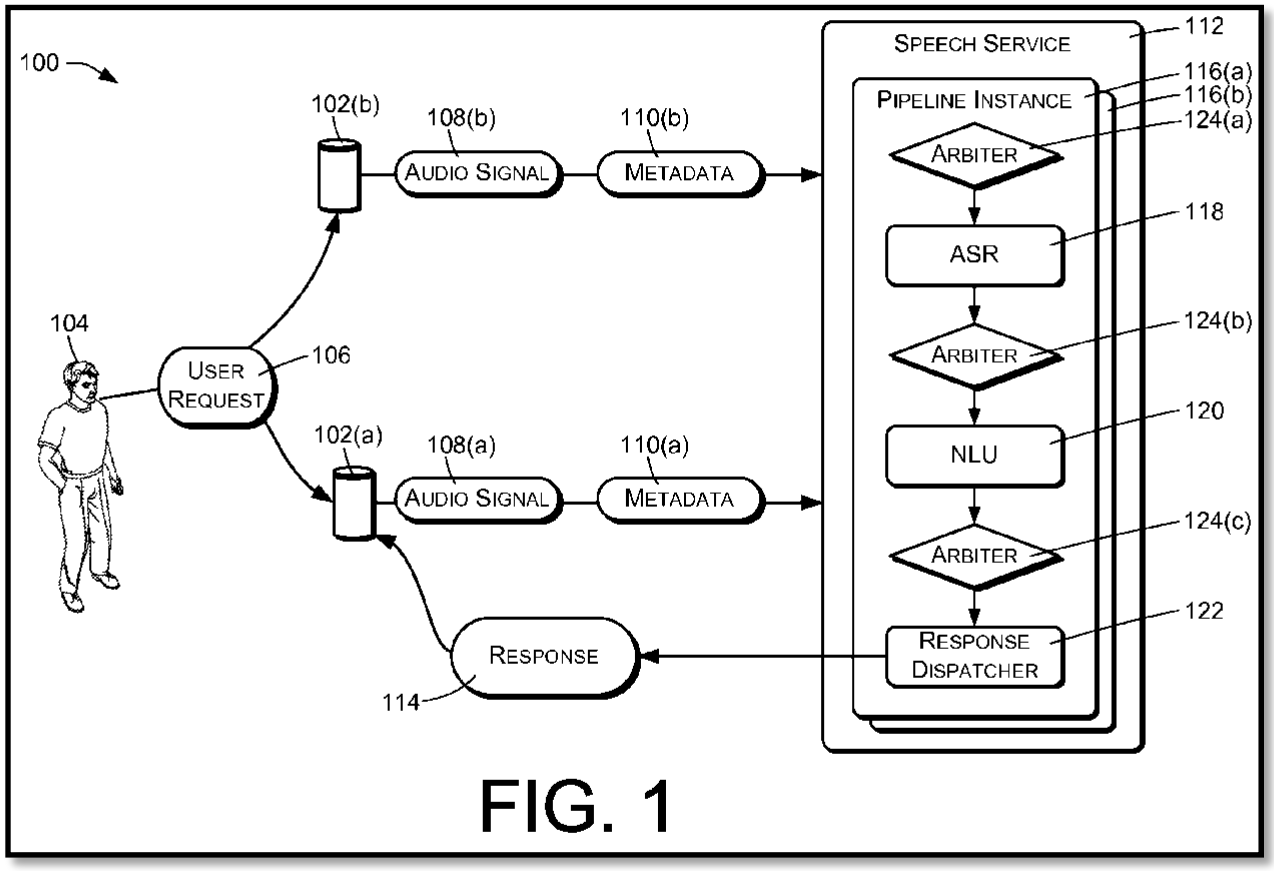
93. US Patent Publication 2017/0083285 ("Meyers") is titled "Device Selection for Providing a Response" and was filed on September 21, 2015. Ex.1004. Thus, Meyers qualifies as prior art to the '311 Patent, which has an effective priority date no earlier than October 3, 2016, for the reasons I explained above.

94. Meyers is generally directed at addressing the problem where multiple "speech interface devices" in a system detect the same user utterance and multiple speech interface devices "independently attempt to process and respond to the user utterance as if it were two separate utterances":

> In situations in which multiple speech interface devices are near each other, such as within a single room or in adjoining rooms, each of the speech interface devices may receive a user utterance and each device may independently attempt to process and respond to the user utterance as if it were two separate utterances. The following disclosure relates to techniques for ***avoiding such duplicate efforts and responses***, among other things.

Id., ¶14; *see also id.*, ¶31, 36, 66.

95. Meyers provides the following illustration of a system comprising multiple "speech interface devices" (102(a), 102(b)) that detect the same voice input from a user:



Id., FIG. 1.

96. Meyers summarizes its disclosed solution for "avoiding [] duplicate efforts and responses" as follows:

> A system may use multiple speech interface devices to interact with a user by speech. All or a portion of the speech interface devices may detect a user utterance and may initiate speech processing to determine a meaning or intent of the utterance. Within the speech processing, arbitration is employed to select one of the multiple speech interface devices to respond to the user utterance. Arbitration may be based in part on metadata that directly or indirectly indicates the proximity of the user to the devices, and the device that is deemed to be nearest the user may be selected to respond to the user utterance.

Id., ¶Abstract. In this way, Meyers discloses arbitration techniques "for determining which of the devices 102 the user 104 most likely wants to respond to the user request 106." *Id.*, ¶31.

97. Meyers discloses that a "speech service 112" performs the disclosed arbitration techniques and the "speech service 112" could take the form of "a network-accessible service implemented by multiple server computers that support devices 102 in the homes or other premises of many different users" or "one or more of the devices 102 may include or provide the speech service 112." *Id.*, ¶30.

98. Meyers explains that each speech interface device is configured to "detect" voice input containing "a user request 106." *See, e.g.*, Meyers, ¶15 ("In the described embodiments, each speech interface device ***detects*** that a user is speaking a command…."), ¶21 ("The devices 102 may be located near enough to each other so that both of the devices 102 may ***detect*** an utterance of the user 104.").

99. Meyers further explains that "[i]n certain implementations, the user request 106 may be prefaced by a wakeword or other trigger expression that is spoken by the user 104 to indicate that subsequent user speech is intended to be received and acted upon by one of the devices 102." *Id.*, ¶23. Meyers continues by explaining "[t]he device 102 may ***detect*** the wakeword and interpret subsequent user speech as being directed to the device 102" where "[a] wakeword in certain

embodiments may be a reserved keyword that is ***detected*** locally by the speech interface device 102." *Id.*

100. In particular, Meyers' speech interface device includes a "voice activity detector 922" that performs "voice activity detection" and "wake word detector 920" that performs "wakeword detection":

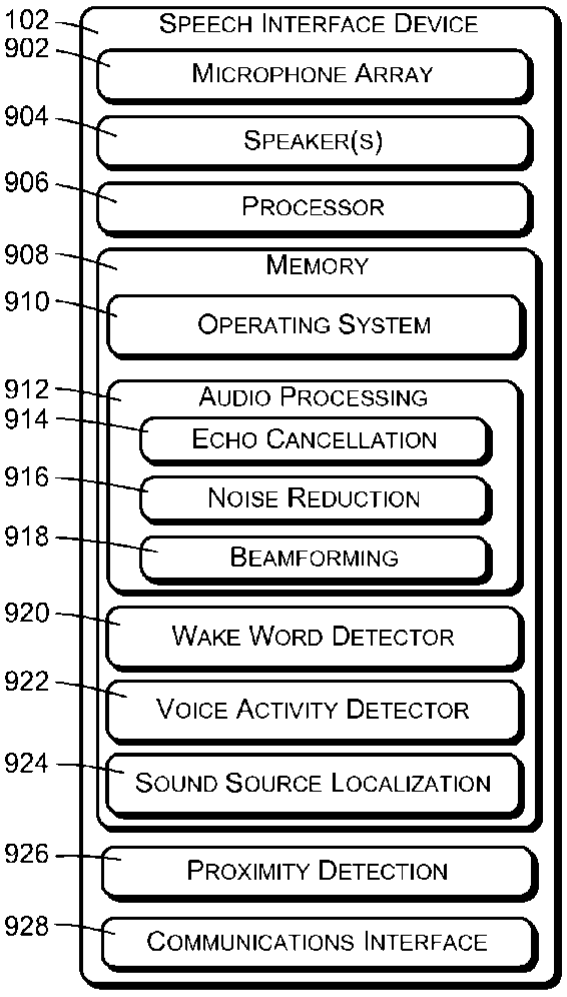


FIG. 9

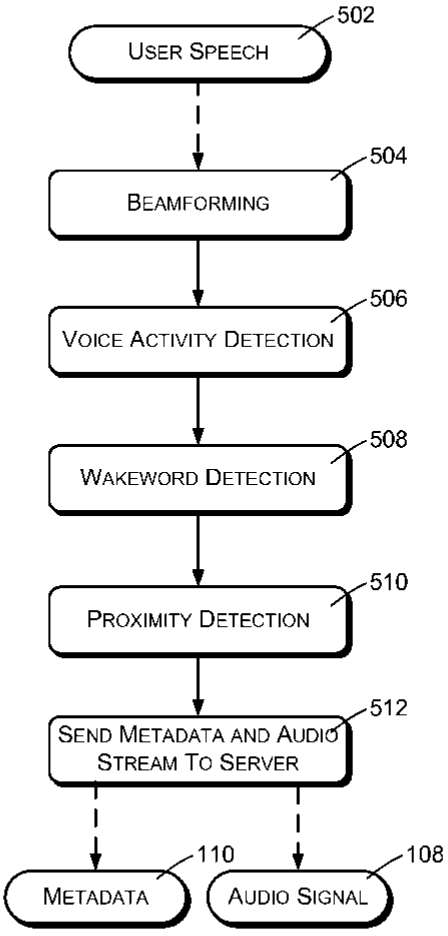


FIG. 5

101. With respect to the voice activity detector and voice activity detection (VAD), Meyers explains:

> VAD ***determines*** the level of ***voice presence*** in an audio signal by analyzing a portion of the audio signal to evaluate features of the audio signal such as signal energy and frequency distribution. The features are quantified and compared to reference features corresponding to reference signals that are known to contain human speech. The comparison produces a score corresponding to the degree of similarity between the features of the audio signal and the reference features. The score is used as an indication of the detected or likely level of speech presence in the audio signal.

Id., ¶102; *see also id.*, ¶120.

102. With respect to the wake word detector (or "expression detector") and wakeword detection, Meyers explains:

> In certain implementations, each device 102 may have an expression detector that analyzes an audio signal produced by a microphone of the device 102 to ***detect the wakeword***, which generally may be a predefined word, phrase, or other sound. Such an expression detector may be implemented using keyword spotting technology, as an example. A keyword spotter is a functional component or algorithm that evaluates an audio signal to ***detect the presence*** a predefined word or expression in the audio signal. ***Rather than producing a transcription*** of the words of the speech, ***a keyword spotter generates a true/false*** output to indicate whether or not the predefined word or expression was represented in the audio signal.

Id., ¶24; *see also id.*, ¶104, 119.

103. Meyers also explains that wakeword detection may be performed on an audio signal "within which voice activity has been detected…." *Id.*, ¶103.

104. Upon detecting a voice signal, Meyers discloses that each speech interface device is configured to determine "metadata" associated with the detected voice signal that includes "one or more signal attributes." *See, e.g., id.*, ¶16, 40. Meyers discloses numerous signal attributes, including (i) "the amplitude of the

audio signal," (ii) "the signal-to-noise ratio of the audio signal," (iii) "the level of voice presence detected in the audio signal," and (iv) "the confidence level with which a wakeword was detected in the audio signal," among other examples. *Id.*, ¶40; *see also id.*, ¶70, 108.

105. With regard to "the level of voice presence detected in the audio signal" attribute, Meyers explains:

> VAD determines the level of voice presence in an audio signal by analyzing a portion of the audio signal to evaluate features of the audio signal such as signal energy and frequency distribution. The features are quantified and compared to reference features corresponding to reference signals that are known to contain human speech. The comparison produces ***a score corresponding to the degree of similarity*** between the features of the audio signal and the reference features. The score is used as ***an indication of the detected or likely level of speech presence in the audio signal***.

Id., ¶102; *id.*, ¶108 ("[T]he VAD 506 may produce a voice presence level, ***indicating the likelihood a person is speaking*** in the vicinity of the device 102.").

106. With regard to "the confidence level with which a wakeword was detected in the audio signal" – or "the level of confidence with which a wakeword or other trigger expression was detected" (*id.*, ¶70) – attribute, Meyers explains "[t]he wakeword [detector] may produce ***a wakeword confidence level***, corresponding to ***the likelihood that the user 104 has uttered the wakeword***." *Id.*, ¶108; *see also id.*, ¶27, 106, 119.

107. Meyers explains that "[t]he device whose audio signal 108 has the strongest set of attributes is selected as the winner of the arbitration." *Id.*, ¶66. In this regard, Meyers provides the following examples:

> For example, the metadata 110 associated with an audio signal 108 may comprise the amplitude of the audio signal 108, and the action 212 may comprise selecting the device 102 producing the audio signal 108 having the ***highest amplitude***. The metadata 110 may comprise or may indicate the level of human voice presence detected in the audio signal 108, and the action 212 may comprise selecting the device 102 producing the audio signal 108 having the ***highest level of detected voice presence***. Similarly, the metadata may comprise or may indicate a signal-to-noise ratio of the audio signal 108 and the action 212 may comprise selecting the device 102 providing the audio signal 108 having the ***highest signal-to-noise ratio***. As another example, the metadata 110 may comprise or indicate the level of confidence with which a wakeword or other trigger expression was detected by the device 102, and the action 212 may comprise selecting the device 102 that detected the trigger expression with the ***highest level of confidence***.

Id., ¶70.

D. Overview of Jang

108. U.S. Patent Publication 2013/0073293 ("Jang") was published on March 21, 2013 from a patent application filed on September 20, 2011. Ex.1007. Thus, Jang qualifies as prior art to the '311 Patent, which has an ***earliest possible*** priority date of October 9, 2014.

109. Jang discloses a system of "electronic devices" (*see, e.g., id.*, ¶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.

110. Jang explains that the "voice recognition unit" enables each electronic device to recognize a user's voice command and there can be scenarios where 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 10 a 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 10 a ***generates the same output in response to the user's voice command***.").

111. 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 performing arbitration independently (*see, e.g., id.*, ¶148-52, FIG. 9) or a "managing" device performing arbitration on behalf of the electronic devices in the system.

112. In this regard, Jang discloses that multiple electronic devices may detect the same voice input.

> The voice recognition unit 182 can carry out voice recognition upon voice signals input through the microphone 122 of the electronic device 100 or the remote control 10 and/or the mobile terminal shown in FIG. 1; the voice recognition unit 182 can then obtain at least one recognition candidate corresponding to the recognized voice. For example, the voice recognition unit 182 can recognize the input voice signals by ***detecting voice activity from the input voice signals***, carrying out sound analysis thereof, and recognizing the analysis result as a recognition unit.

Id., ¶110; *see also id.*, ¶127 ("For example, the TV 100 receives a voice command saying 'next channel' from the user. Other electronic devices (for example, the tablet PC 10 a) than the first electronic device 100, which are connected to the first electronic device over a network, may receive the user's voice command.").

113. Each electronic device may then compute a respective "voice recognition 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 respond to the voice input. *Id.*, ¶130-132, 135-140, 149-151.

114. According to Jang, the "voice recognition results" "indicate whether or not recognition of the voice command input was successful." *Id.*, ¶12. In this regard, Jang discloses that a "voice recognition result" (sometimes instead referred to as a "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.

X. OVERVIEW OF THE GROUNDS

115. As explained below, it is my opinion that each of the Challenged Claims –claims 1-3, 8-12, 14-20– is invalidated by Piersol and that each of Challenged Claims 1-3, 8-11, 14, 16-17, and 19-20 is invalidated by Masahide, according to the following grounds:

Ground	Reference(s)	Claims
Ground 1A	Piersol (§102)	Independent cls. 1, 10, 16 Dependent cls. 2-3, 8-9, 11, 14-15, 17-20
Ground 1B	Piersol + Meyers (§103)	Independent cls. 1, 10, 16 Dependent cls. 2-3, 8-9, 11-12, 14-15, 17-20
Ground 2A	Masahide (§102)	Independent cls. 1, 10, 16 Dependent cls. 2-3[4], 8-9, 11, 14, 17, 19-20
Ground 2B	Masahide + Jang (§103)	Independent cls. 1, 10, 16 Dependent cls. 2-3, 8-9, 11, 14, 17, 19-20

XI. THE CHALLENGED CLAIMS ARE UNPATENTABLE

116. For the reasons I explain in greater detail below, it is my opinion that a POSITA would have considered the Challenged Claims to be unpatentable as anticipated and obvious over the prior art discussed in this Declaration.

[4] Masahide alone also renders claim 3 obvious.

A. Ground 1A: Piersol Anticipates Independent Claims 1, 10, & 16

117. As explained in detail below, it is my opinion that Piersol anticipates independent claims 1, 10, and 16. I provide my analysis of claim 16 first and then discuss claims 1 and 10.

1. Piersol Anticipates Claim 16

118. As explained below, it is my opinion that Piersol discloses each element of claim 16 and thus, anticipates claim 16.

119. Claim 16 recites in full (with bracketed numerals inserted for ease of reference):

> **[16.0]** *A system comprising a plurality of* **[16.0(a)]** *electronic devices, each of the electronic devices including* **[16.0(b)]** *one or more microphones,* **[16.0(c)]** *one or more processors, and* **[16.0(d)]** *memory storing one or more programs for execution by the one or more processors, each of the electronic devices programmed to:*
>
> **[16.1]** *detect a voice input;*
>
> **[16.2]** *determine a first value associated with the detected voice input;*
>
> **[16.3]** *receive respective second values from other electronic devices of the plurality of electronic devices;*
>
> **[16.4]** *compare the first value and the respective second values; and*
>
> **[16.5]** *in accordance with a determination that the first value is higher than the respective second values, respond to the voice input.*

a. [16.0] Preamble

120. The preamble of claim 16 recites **[16.0]** a *system comprising a plurality of* **[16.0(a)]** *electronic devices, each of the electronic devices including* **[16.0(b)]** *one or more microphones*, **[16.0(c)]** *one or more processors*, *and* **[16.0(d)]** *memory*

storing one or more programs for execution by the one or more processors, each of the electronic devices programmed to perform the recited functions.

121. To the extent that the preamble is limiting, Piersol discloses the elements of the preamble. In this regard, Piersol discloses that "FIG. 1 is a block diagram illustrating a system and environment for implementing a digital assistant according to various examples." Piersol, ¶14, 37. In this "system 100," a "digital assistant can include client-side portion 102… executed on user device 104," which "can be any suitable electronic device," and "[s]econd user device 122 can be similar or identical to user device 104" and communicatively coupled with one another "via a direct communication connection, such as Bluetooth, NFC, BTLE, or the like, or via a wired or wireless network, such as a local Wi-Fi network." *Id.*, ¶39, 41, 44.

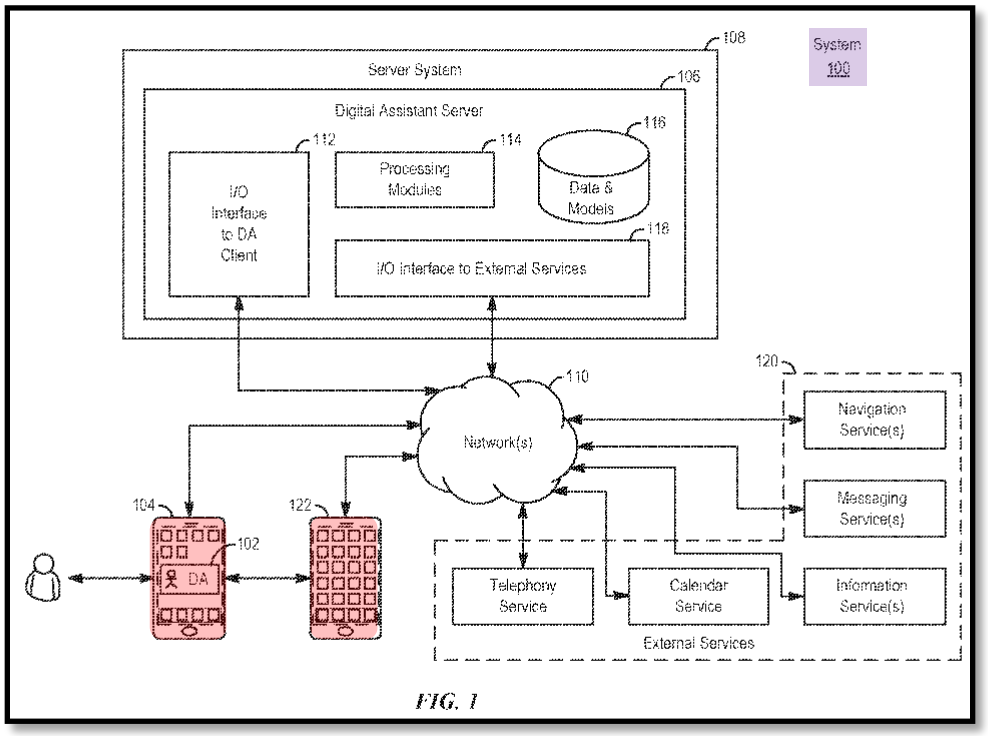
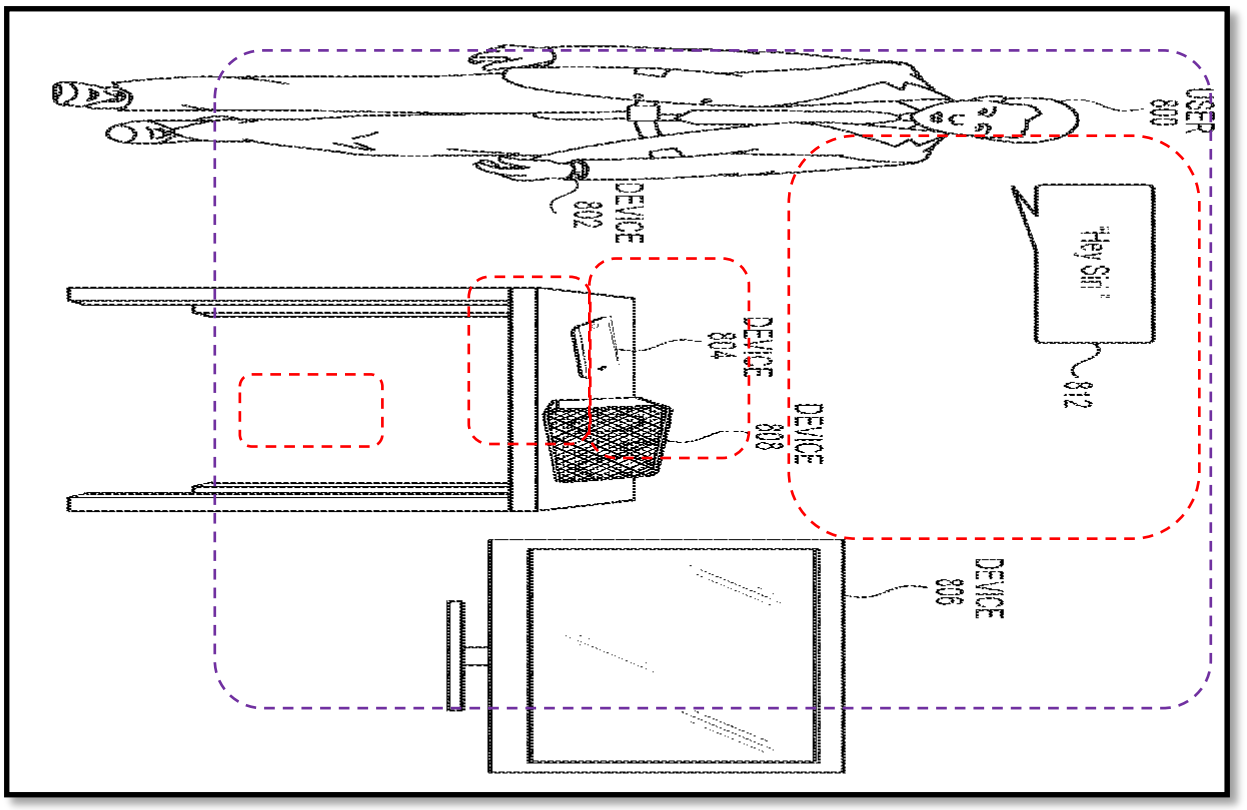


FIG. 1

Id., FIG. 1 (annotations added); *see also*, *e.g.*, *id.* ¶41 ("User device 104 can be any suitable electronic device. For example, user devices can be a portable multifunctional device (e.g., device 200, described below with reference to FIG. 2A), a multifunctional device (e.g., device 400, described below with reference to FIG. 4), or a personal electronic device (e.g., device 600, described below with reference to FIG. 6A-B.)."), ¶41 (also listing examples of "user device 104").

122. Piersol provides another illustration of a system and environment for implementing a digital assistant according to various examples in FIGs. 8A-8C that show "electronic devices 802, 804, 806, and 808 of user 800" where "[o]ne or more of the devices 802-808 may be any of devices 104, 122, 200, 400, 600, and 1200 (FIGS. 1, 2A, 3, 4, 5A, 6A-6B, and 12) in some embodiments." Piersol, ¶245.

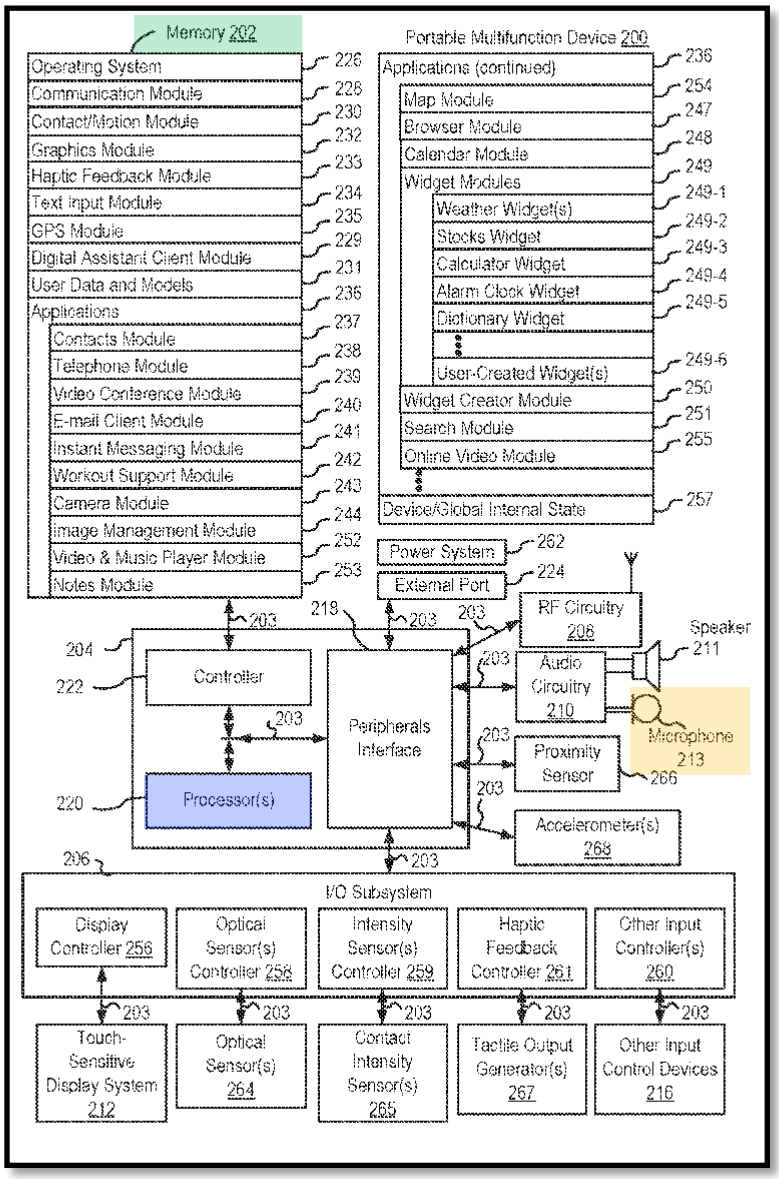


Id., FIG. 8B (annotations added); *see also* FIGs. 8A, 8C.

123. Piersol provides several block diagrams of "user devices" and "electronic devices" including FIG. 2A, which "is a block diagram illustrating a portable multifunction device [200] implementing the client-side portion of a digital assistant in accordance with some embodiments." Piersol, ¶15; *see also, e.g., id.*, ¶47.

124. Piersol discloses that "[d]evice 200 includes memory 202 (which optionally includes one or more computer-readable storage mediums), memory controller 222, one or more processing units (CPUs) 220, peripherals interface 218,

RF circuitry 208, audio circuitry 210, speaker 211, microphone 213, input/output (I/O) subsystem 206, other input control devices 216, and external port 224." *Id.*



Id., FIG. 2A (annotations added); *see also, e.g.*, *id.*, ¶17, FIG. 3, ¶18, FIG. 4, ¶19, FIG. 5A, ¶20-21, FIGs. 6A-6B, ¶30, 321 ("FIG. 12 shows a functional block diagram of an electronic device 1200 configured in accordance with the principles of the

various described embodiments, including those described with reference to FIGS. 8A-C and 10A-C."), 322-23, FIG. 12.

125. Piersol explains "[m]emory 202 may include one or more computer-readable storage mediums. The computer-readable storage mediums may be tangible and non-transitory." Piersol, ¶51. Piersol further explains "a non-transitory computer-readable storage medium of memory 202 can be used to store instructions (e.g., for performing aspects of processes 1000 …) for use by or in connection with an instruction execution system, apparatus, or device, such as a computer-based system, processor-containing system, or other system that can fetch the instructions from the instruction execution system, apparatus, or device and execute the instructions." Piersol, ¶52. In turn, "FIGS. 10A-C illustrate process 1000 for operating a digital assistant according to various examples. Process 1000 is performed, for example, using one or more electronic devices (e.g., devices 104, 106, 200, 400, or 600) implementing a digital assistant…. [M]ethod 1000 provides an efficient way for arbitrating among multiple devices for responding a user input." Piersol, ¶294-95.

126. Piersol also explains "[t]he one or more processors 220 run or execute various software programs and/or sets of instructions stored in memory 202 to perform various functions for device 200 and to process data." Piersol, ¶53.

127. As evident from the above citations, Piersol uses different phrases to refer to the same "device" including "user device," "electronic device," and "multifunction device," among others. For this Declaration, I will refer to Piersol's "device" as an "electronic device" for consistency with the claim language.

128. Thus, it is my opinion that Piersol discloses **[16.0]** a *system comprising a plurality of* **[16.0(a)]** *electronic devices, each of the electronic devices including* **[16.0(b)]** *one or more microphones*, **[16.0(c)]** *one or more processors, and* **[16.0(d)]** *memory storing one or more programs for execution by the one or more processors, each of the electronic devices programmed to* perform the recited functions, as discussed in further detail below.

b. [16.1] Detect a Voice Input

129. Claim 16 recites that *each of the electronics devices [is] programmed to:* **[16.1]** *detect a voice input* …. As noted above, Google contends *detect[ing] a voice input* (i) refers "to a process for identifying that an audio input includes voice," (ii) "refers to a process separate from performing speech recognition (e.g., speech to text) on the voice input," and (iii) can be satisfied by "voice activity detection" or "hotword/wakeword detection." *Supra* §VII.A.

130. Piersol discloses this element. In this regard, Piersol explains that each electronic device is configured to detect a "spoken trigger," which may take the form of "Hey Siri":

> In some examples, the electronic device determines respective values and/or broadcasts the values in response to ***detecting*** a spoken trigger. With reference to FIGS. 8A-C, the spoken trigger is the phrase "Hey Siri". In response to the spoken trigger "***Hey Siri***", each of the electronic devices 802-808 broadcasts the respective set of one or more values, as described. If an audio input does not contain a spoken trigger, the electronic device foregoes determining and/or broadcasting a set of one or more values.

Piersol, ¶254; *see also, e.g.*, ¶246.

131. A POSITA would appreciate that Piersol's "spoken trigger" (e.g., "Hey Siri") is synonymous with what was also referred to in the art as a "hotword" or "wakeword." *See, e.g.*, '311 Patent, 6:2-5 ("[A] 'hotword' is a user defined or predefined word or phrase used to 'wake-up' or ***trigger*** a voice-activated device to attend/respond to a spoken command that is issued subsequent to the hotword").

132. Piersol explains that the electronic device detects a spoken trigger by its microphone first sampling an audio input and then the electronic device "determin[ing] whether the audio input comprises a spoken trigger":

> At block 1002, a first electronic device having ***a microphone samples*** an audio input. In some examples, the audio input can be received via a microphone (e.g., microphone 213) of the electronic device…. At block 1004, the first electronic device optionally ***determines whether*** the audio input comprises a spoken trigger (e.g., "Hey Siri")."

Piersol, ¶297-98; *see also id.*, ¶330 ("In some examples, sampling the audio input comprises ***determining*** (e.g., with determining unit 1214) ***whether*** the audio input comprises a spoken trigger (e.g., block 1004 of FIG. 10A).").

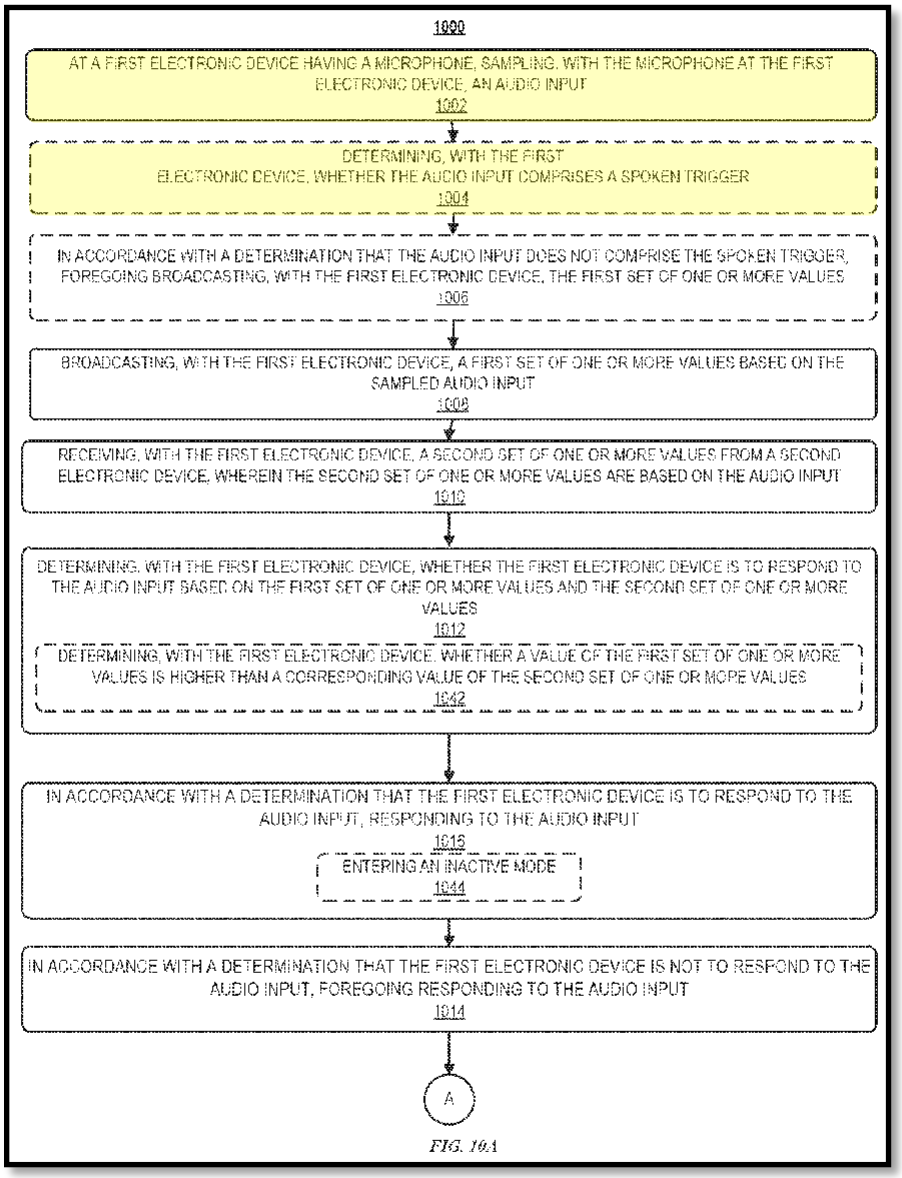


FIG. 10A

Id., FIG. 10A (annotations added).

133. In this way, Piersol discloses *detect[ing] a voice input* by virtue of hotword detection.

134. Notably, Piersol discloses detecting a spoken trigger independent of Piersol's disclosures of "speech-to-text" or "ASR systems." In this respect, Piersol describes "digital assistant system 700," which may be implemented on a cloud

server (Piersol, ¶202) or on an electronic device (*id.*, ¶205), as including "speech-to-text (STT) processing module 730" (*id.*, ¶212) that "can include one or more ASR systems" (*id.*, ¶215). Piersol explains "each ASR system can include one or more speech recognition models (e.g., acoustic models and/or language models) and can implement one or more speech recognition engines" that "can be used to process the extracted representative features of the front-end speech pre-processor to produce intermediate recognitions results (e.g., phonemes, phonemic strings, and sub-words), and ultimately, text recognition results (e.g., words, word strings, or sequence of tokens)." *Id.* Piersol further explains, "[o]nce STT processing module 730 produces recognition results containing a text string (e.g., words, or sequence of words, or sequence of tokens), the recognition result can be passed to natural language processing module 732 for intent deduction."

135. Thus, it is my opinion that Piersol discloses that *each of the electronics devices [is] programmed to:* **[16.1]** *detect a voice input*.

c. [16.2] Determine a First Value Associated with the Detected Voice Input

136. Claim 16 recites that *each of the electronics devices [is] programmed to:…* **[16.2]** *determine a first value associated with the detected voice input ….*

137. Piersol discloses this element. In this regard, Piersol explains that each electronic device is configured to determine (or generate) a "score" (also referred to as a set of one or more "values") "in response to detecting a spoken trigger":

> In some examples, the electronic device ***determines respective values*** and/or broadcasts the values ***in response to detecting a spoken trigger***. With reference to FIGS. 8A-C, the spoken trigger is the phrase "Hey Siri". In response to the spoken trigger "Hey Siri", each of the electronic devices 802-808 broadcasts the respective ***set of one or more values***, as described. If an audio input does not contain a spoken trigger, the electronic device foregoes determining and/or broadcasting a set of one or more values.

Piersol, ¶254.

> In some examples, each of electronic devices 802-808 broadcasts multiple values (e.g., an aggregated ***score*** and a "device type" value)…. In some examples, each of electronic devices 802-808 broadcasts a single ***score***, which is calculated according to one or more functions using some or all of the above-discussed values. For example, a single score for broadcasting can be calculated based on a predetermined weighting of ***received audio quality*** (e.g., signal-to-noise ratio), type of device, ability to perform task, and device state.

Id., ¶261.

> Turning to FIG. 8B, user 800 provides audio input 812, which includes only a spoken trigger ("Hey Siri") and does not specify a task. In response, each of electronic devices 802-808 ***generates*** and broadcasts ***a set of one or more values***, as described.

Id., ¶267.

138. Piersol explains that, in response to detecting a spoken instruction comprising the "spoken trigger," "electronic devices 802-808 ***initiate an arbitration process*** to identify (e.g., determine) an electronic device for responding to the spoken instruction 834 from user 800." *Id.*, ¶247. Each electronic device determines and broadcasts a set of values or score that "may include one or more values" "based on the audio input as sampled on the respective device." *Id.*

139. Piersol discloses that these "one or more values" or score can take various forms (e.g., a single value/score derived from multiple values with different weights) and comprise various information "relevant to implementing intelligent device arbitration":

> The above-discussed exemplary values correspond to ***various metrics*** relevant to implementing intelligent device arbitration as described herein. It will be appreciated that ***any number of values*** corresponding to these metrics can be broadcasted and/or that ***some or all of the values can be used to provide a single value*** using one or more functions. The single value may thereafter be broadcasted during device arbitration as described herein. It will be further appreciated that the one or more functions can ***assign different weights to different values***, respectively.

Piersol, ¶253; *id.*, ¶261 ("[A] ***single score*** for broadcasting can be calculated based on a ***predetermined weighting*** of received audio quality (e.g., signal-to-noise ratio), type of device, ability to perform task, and device state."), ¶267 ("In some examples, each of electronic devices 802-808 calculates ***a single value aggregating some or all of the exemplary values*** disclosed herein.").

140. More specifically, as one example, Piersol explains that a value/score could indicate an "audio quality," such as "an energy level of the audio input," and comprise "signal-to-noise ratio, sound pressure, or a combination thereof":

> In some examples, a set of one or more values broadcasted by an electronic device includes any number of values. One exemplary value indicates an ***energy level of the audio input*** as sampled on the electronic device. The energy level of the sampled audio input may, for instance, be indicative of the proximity of the electronic device to the user. In some examples, the energy level is measured using known

> metrics for audio quality, such as ***signal-to-noise ratio***, sound pressure, or a combination thereof.

Piersol ¶248; *id.*, ¶261 ("[A] single score for broadcasting can be calculated based on a predetermined weighting of ***received audio quality*** (e.g., ***signal-to-noise ratio***), type of device, ability to perform task, and device state.")

141. As another example, Piersol explains that a value/score could indicate "an acoustic fingerprint of the audio input" or "a confidence value that expresses the likelihood that the audio input originates from [a] particular user":

> Another exemplary ***value indicates an acoustic fingerprint of the audio input***, that is, ***whether the audio input is likely provided by a particular user.*** In some examples, the electronic device analyzes the audio input and calculates ***a confidence value*** that expresses the likelihood that the audio input originates from the particular user (e.g., a user that has enrolled in the virtual assistant service on the electronic device). In some examples, the electronic device determines whether the audio input is from an authorized user by comparing the confidence value with a predetermined threshold value, and broadcasts a value based on the comparison. In some examples, the set of one or more values may include ***a value corresponding to the confidence value*** as well.

Piersol, ¶249.

142. As yet another example, Piersol explains that a value could indicate "a type of the electronic device" in instances "when the spoken instruction specifies a task to be performed":

> Another exemplary value indicates a type of the electronic device. For example, as illustrated in FIG. 8A, the electronic device 804 is a mobile phone. It will be appreciated that predetermined values can be used by electronic devices 802-808 to ***indicate different device types***, including

> but not limited to, speaker, television, smart watch, laptop, tablet, mobile device, set-top box, headphones, or any combination thereof. In some examples, each of the electronic devices 802-808 ***broadcasts a "device type" value only when the spoken instruction specifies a task to be performed***, as discussed in more details below.

Piersol, ¶250.

143. In connection with FIG. 10, Piersol describes an electronic device determining and then broadcasting a first set of values:

> At block 1008, the first electronic device broadcasts ***a first set of one or more values based on the sampled audio input***. In some examples, a value of the first set of values is based on a signal to noise ratio of speech of the audio input sampled with the first electronic device. In some examples, a value of the first set of values is based on a sound pressure of the audio input sampled with the first electronic device. In some examples, the first electronic device identifies a confidence value indicative of a likelihood that the audio input was provided by a particular user, and a value of the first set of values is based on the confidence value.

Piersol, ¶297.

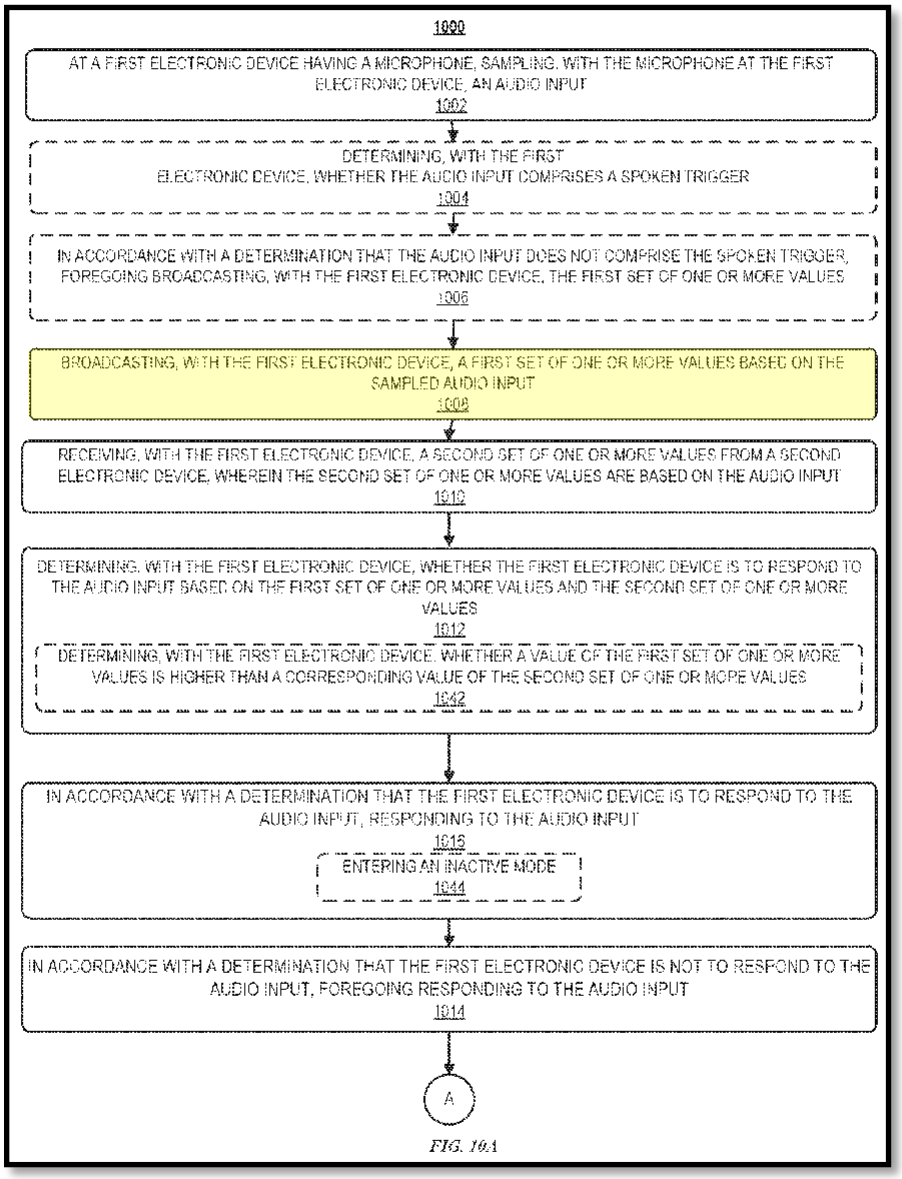


FIG. 10A

Id., FIG. 10A (annotations added).

144. Thus, it is my opinion that Piersol discloses that *each of the electronics devices [is] programmed to:…* **[16.2]** *determine a first value associated with the detected voice input*.

d. [16.3] Receive Respective Second Values from Other Electronic Devices of the Plurality of Electronic Devices

145. Claim 16 recites that *each of the electronics devices [is] programmed to:…* **[16.3]** *receive respective second values from other electronic devices of the plurality of electronic devices ….*

146. Piersol discloses this element. For instance, Piersol discloses that each electronic device is configured to receive a respective one or more sets of values or score from one or more other electronic devices:

> Each of the electronic devices 802-808 may ***receive sets of values from other devices***. By way of example, the electronic device 802 may receive sets of values from electronic devices 804-808, the electronic device 804 may receive sets of values from electronic devices 802, 806, 808, and so on. After sets of values have been ***exchanged*** among electronic devices 802-808, each device determines whether it is to respond to the audio input by analyzing the sets of values, as discussed below.

Piersol, ¶255.

147. In connection with FIG. 10, Piersol describes an electronic device receiving one or more values from another electronic device: "At block 1010, the first electronic device ***receives*** a second set of one or more values from a second electronic device. The second set of one or more values is based on the audio input sampled at the second electronic device." Piersol, ¶299.

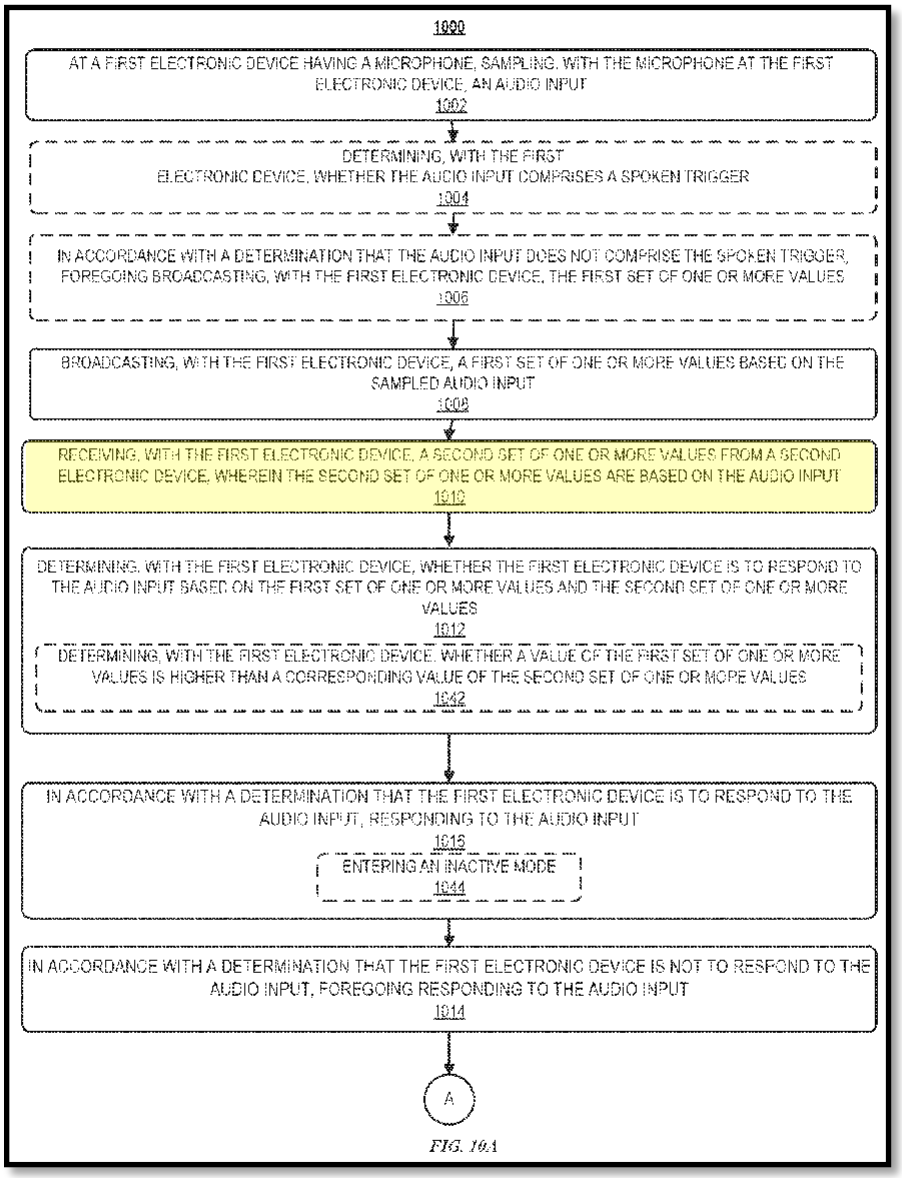


FIG. 10A

Id., FIG. 10A (annotations added).

148. Piersol explains that an electronic device receives one or more values or a score from another electronic device by virtue of the other electronic device broadcasting its one or more values or score:

> Turning to FIG. 8B, user 800 provides audio input 812, which includes only a spoken trigger ("Hey Siri") and does not specify a task. In response, ***each of electronic devices 802-808*** generates and ***broadcasts***

> a set of one or more values, as described. In some examples, each of electronic devices 802-808 ***broadcasts*** an "energy level" value ***to each other***.

Piersol, ¶267.

> In some examples, each of electronic devices 802-808 ***broadcasts*** multiple values (e.g., an aggregated ***score*** and a "device type" value). Accordingly, electronic devices 802-808 may arbitrate in accordance with one or more predetermined algorithms. In some examples, each of electronic devices 802-808 ***broadcasts*** a single ***score***, which is calculated according to one or more functions using some or all of the above-discussed values. For example, ***a single score for broadcasting*** can be calculated based on a predetermined weighting of received audio quality (e.g., signal-to-noise ratio), type of device, ability to perform task, and device state.

Piersol, ¶261.

149. Thus, it is my opinion that Piersol discloses that *each of the electronics devices [is] programmed to:…* **[16.3]** *receive respective second values from other electronic devices of the plurality of electronic devices*.

e. [16.4] Compare the First Value and the Respective Second Values

150. Claim 16 recites that *each of the electronics devices [is] programmed to:…* **[16.4]** *compare the first value and the respective second values* ….

151. Piersol discloses this element. For instance, Piersol discloses embodiments where each electronic device is configured to compare its own determined one or more values or score with the one or more values or score received from another electronic device. *See* Piersol ¶256-59.

152. As one particular example, Piersol discloses an electronic device comparing its "'energy level' value" with the "'energy level' values broadcasted by other electronic devices":

> In some examples, each device determines whether to respond to the spoken instruction 834 based on the energy level values broadcasted by electronic devices 802-808. Each of the electronic devices 802-808 ***compares*** a respective "energy level" value with the "energy level" values broadcasted by other electronic devices. In some examples, an electronic device responds to the audio input if the electronic device has broadcasted the highest "energy level" value. As discussed above, a higher "energy level" value can be indicative of greater proximity to the user. Because a device closest to a user may be associated with a highest energy level value, it is beneficial to have an electronic device that has broadcasted the highest "energy level" value respond to the spoken instruction 934.

Piersol, ¶256; *see also id.*, ¶257 (comparing "'state' values"), ¶258 (comparing "'acoustic fingerprint' values"), ¶258 (comparing "'device type' values").

153. Thus, it is my opinion that Piersol discloses that *each of the electronics devices [is] programmed to:…* **[16.2]** *compare the first value and the respective second values*.

f. [16.5] In Accordance with a Determination that the First Value Is Higher than the Respective Second Values, Respond to the Voice Input

154. Claim 16 recites that *each of the electronics devices [is] programmed to:…* **[16.5]** *in accordance with a determination that the first value is higher than the respective second values, respond to the voice input*.

155. Piersol discloses this element. For instance, Piersol discloses, in connection with FIG. 10, that each electronic device is configured to (i) "determine[] whether the [] electronic device is to respond to the audio input by determining whether a value of the first set of one or more values is higher than a corresponding value of the second set of one or more values" from another electronic device and (ii) "respond[] to the audio input" "in accordance with a determination that the [] electronic device is to respond to the audio input":

> At block 1012, the first electronic device determines whether the first electronic device is to respond to the audio input based on the first set of one or more values and the second set of one or more values. In some examples, the first electronic device determines whether the first electronic device is to respond to the audio input by ***determining*** whether a value of the first set of one or more values is ***higher than*** a corresponding value of the second set of one or more values, at block 1042. At block 1014, ***in accordance with a determination*** that the first electronic device is to respond to the audio input, the first electronic device ***responds to the audio input***.

Piersol, ¶300-301.

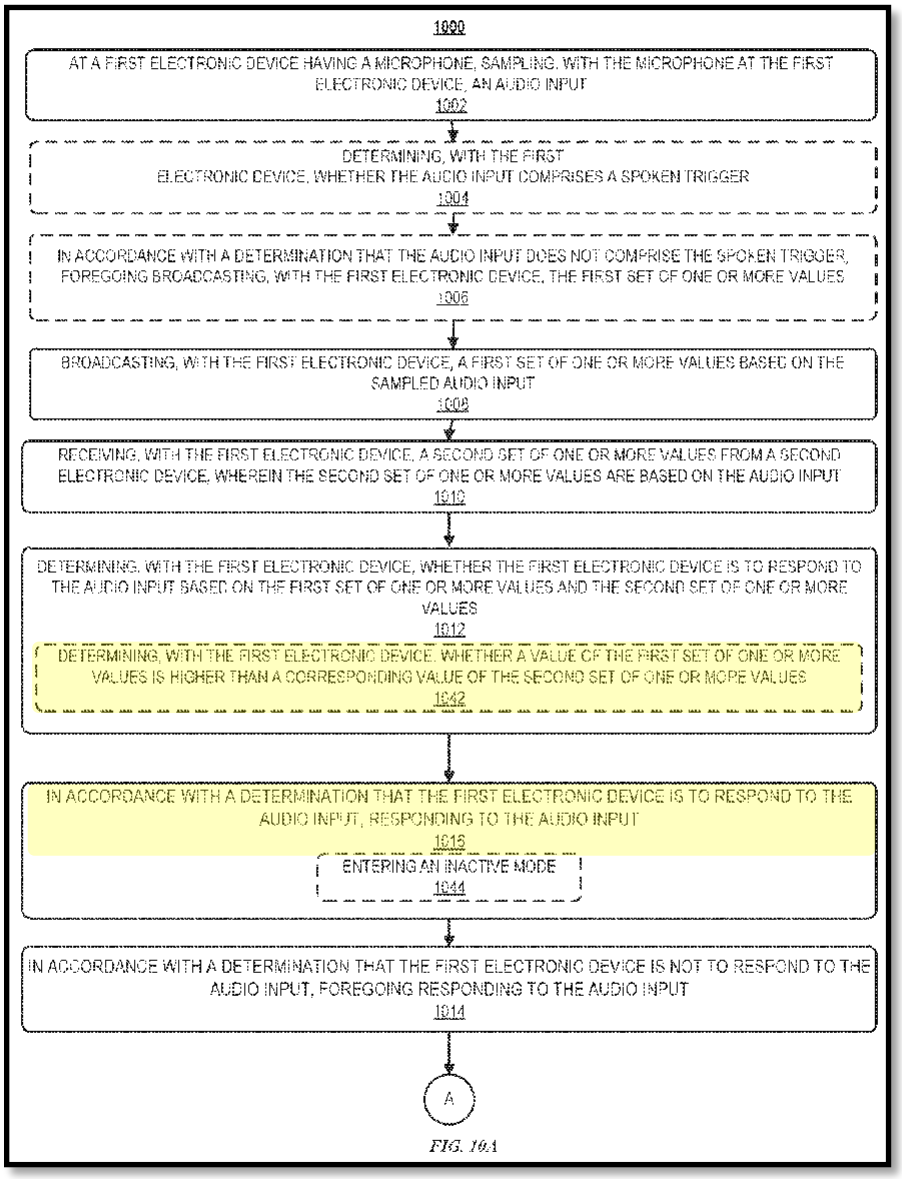


FIG. 10A

Id., FIG. 10A (annotations added); *see also, e.g.*, *id.*, ¶339 ("In some examples, determining whether the electronic device is ***to respond*** to the audio input comprises: ***determining*** (e.g., with determining unit 1214) whether a value of the first set of one or more values is ***higher than*** a corresponding value of the second set of one or more values (e.g., block 1042 of FIG. 10A).").

156. Piersol provides examples where an electronic device responds to a detected voice input in accordance with determining that it has the highest "energy level" value:

> In some examples, each device determines whether to respond to the spoken instruction 834 based on the energy level values broadcasted by electronic devices 802-808. Each of the electronic devices 802-808 compares a respective "energy level" value with the "energy level" values broadcasted by other electronic devices. In some examples, an ***electronic device responds*** to the audio input ***if*** the electronic device has broadcasted ***the highest "energy level" value***. As discussed above, a higher "energy level" value can be indicative of greater proximity to the user. Because a device closest to a user may be associated with a highest energy level value, it is beneficial to have an electronic device that has broadcasted the highest "energy level" value respond to the spoken instruction 934.

Piersol, ¶256.

> Turning to FIG. 8B, user 800 provides audio input 812, which includes only a spoken trigger ("Hey Siri") and does not specify a task. In response, each of electronic devices 802-808 generates and broadcasts a set of one or more values, as described. In some examples, each of electronic devices 802-808 broadcasts an "energy level" value to each other. In some examples, each of electronic devices 802-808 calculates a single value aggregating some or all of the exemplary values disclosed herein. In this example, electronic device 804 ***determines that it has*** broadcasted the ***highest "energy level" value***, indicating that electronic device 804 is closer to user 800 than the other electronic devices. ***Accordingly, electronic device 804 responds to the audio input 812***, and the rest of the electronic devices forego responding to the audio input.

Id., ¶267.

157. Piersol discloses that the electronic device with the highest value/score can respond to the user's voice input in various manners, including by outputting an audible and/or visual response:

> The electronic device can respond to an audio input by providing a ***visual output*** (e.g., a display notification or LED toggle), ***an auditory output***, a haptic output, or a ***combination thereof***, in some examples. For example, electronic device 804 can respond to the audio input with ***a visual output (e.g. displaying a transcript of the audio input) and an audio output (e.g., "Looking for episodes … ")***.

Piersol, ¶262.

> A satisfactory ***response*** to the user request can be a provision of the requested informational answer, a performance of the requested task, or a combination of the two. For example, a user can ask the digital assistant a question, such as "Where am I right now?" Based on the user's current location, ***the digital assistant can answer, "You are in Central Park near the west gate."*** The user can also request the performance of a task, for example, "Please invite my friends to my girlfriend's birthday party next week." In response, ***the digital assistant can acknowledge the request by saying "Yes, right away,"*** and then send a suitable calendar invite on behalf of the user to each of the user's friends listed in the user's electronic address book. During performance of a requested task, the digital assistant can sometimes interact with the user in a continuous dialogue involving multiple exchanges of information over an extended period of time. There are numerous other ways of interacting with a digital assistant to request information or performance of various tasks. ***In addition to providing verbal responses*** and taking programmed actions, ***the digital assistant can also provide responses in other visual or audio forms***, e.g., as text, alerts, music, videos, animations, etc.

Id., ¶38; *see also, e.g.*, *id.*, ¶240-41; *see also* dependent claim 8.

158. Thus, it is my opinion that Piersol discloses that *each of the electronics devices [is] programmed to:…* **[16.5]** *in accordance with a determination that the first value is higher than the respective second values, respond to the voice input*.

2. Independent Claims 1 & 10

159. As shown below, (i) independent claim 1 recites a *method performed at a first electronic device* where the recited method is broader than the recited functions of independent claim 16 that each *electronic device* of the *system* of claim 16 is programmed to perform and (ii) independent claim 10 recites a *first electronic device* provisioned with *one or more programs comprising instructions for* performing functions that are broader than the recited functions of independent claim 16 that each *electronic device* of the *system* of claim 16 is programmed to perform:

Claim 16	Claim 1	Claim 10
[16.0] A system comprising a plurality of [16.0(a)] electronic devices, each of the electronic devices including [16.0(b)] one or more microphones, [16.0(c)] one or more processors, and [16.0(d)] memory storing one or more programs for execution by the one or more processors, each of the electronic devices programmed to:	[1.0] A method performed at a first electronic device of a plurality of [1.0(a)] electronic devices, each of the plurality of electronic devices comprising [1.0(b)] one or more microphones, [1.0(c)] one or more processors, and [1.0(d)] memory storing one or more programs for execution by the one or more processors, the method comprising:	[10.0] A first electronic device of a plurality of electronic devices, [10.0(a)] the first electronic device comprising: [10.0(b)] one or more microphones; [10.0(c)] one or more processors; and [10.0(d)] memory storing one or more programs to be executed by the one or more processors, the one or more programs comprising instructions for:
[16.1] detect a voice input;	[1.1] detecting a voice input from a user;	[10.1] detecting a voice input from a user;
[16.2] determine a first value associated with the detected voice input;	[1.2] determining a first value associated with the voice input;	[10.2] determining a first value associated with the voice input;
[16.3] receive respective second values from other electronic devices of the plurality of electronic devices;	[1.3] receiving second values associated with the voice input from other electronic devices of	[10.3] receiving second values associated with the voice input from other electronic devices of

Claim 16	Claim 1	Claim 10
	the plurality of electronic devices; and	the plurality of electronic devices; and
[16.4] compare the first value and the respective second values; and		
[16.5] in accordance with a determination that the first value is higher than the respective second values, respond to the voice input.	[1.4] in accordance with a determination that the first value is higher than the second values, responding to the detected input.	[10.4] in accordance with a determination that the first value is higher than the second values, responding to the detected input.

160. As is evident, (i) **[1.0]**, **[1.1]**, **[1.2]**, **[1.3]**, and **[1.4]** map to **[16.0]**, **[16.1]**, **[16.2]**, **[16.3]**, and **[16.5]**, respectively, and (ii) **[10.0]**, **[10.1]**, **[10.2]**, **[10.3]**, and **[10.4]** map to **[16.0]**, **[16.1]**, **[16.2]**, **[16.3]**, and **[16.5]**, respectively.

161. Unlike claim 16, claim 1 and 10 are broader because they do not expressly require the step of **[16.4]** *compar[ing] the first value and the second values*.

162. Further, while claim 16 requires that each *electronic device* of *a plurality of electronic devices* be *programmed to* perform the recited functions, claim 1 only requires a *first electronic device* to perform the claimed method functions and claim 10 only requires a *first electronic device* to comprise *one or more programs comprising instructions* performing the recited functions.

163. Thus, Piersol discloses each of the elements of claims 1 and 10 for the same reasons that Piersol discloses each of the elements of claim 16. I therefore incorporate by reference here my analysis for claim 16. *Supra* ¶118-58.

B. Ground 1A: Piersol Anticipates Dependent Claims 2-3, 8-9, 11, 14-15, 17-20

164. As noted above, the Challenged Claims include (i) claims 2-3 and 8-9 that depend from independent claim 1, (ii) claims 11 and 14-15 that depend from independent claim 10, and (iii) claims 17-20 that depend from independent claim 16.

165. As I explained above, it is my opinion that Piersol anticipates each of independent claims 1, 10, and 16. Below, I address the additional elements of the aforementioned dependent Challenged Claims and explain why Piersol anticipates them as well.

1. Dependent Claim 2

166. Dependent claim 2 recites *[t]he method of claim 1, further comprising:* **[2.1]** *in accordance with a determination that the first value is not higher than the second scores, forgoing responding to the detected input*.[5]

167. Piersol discloses this additional element. As I explained above, Piersol discloses that each electronic device is configured to determine whether it is to respond to a detected voice input if its values/score is higher than any values/score received from another electronic device. *See* element **[16.5]**.

168. Piersol further discloses that each electronic device is configured such that, upon determining that its values/score is not higher than values/score received

[5] For the purposes of this IPR, I have been asked to assume that *the second scores* recited in claim 2 refers to *the second values* recited in independent claim 1.

from another electronic device, it forgoes responding to the detected voice input. *See* Piersol, ¶262 ("If the electronic device determines not to respond to the audio input, the electronic device can ***forego responding*** to the audio input by entering an inactive mode (e.g. sleep mode)."); *see also id.*, ¶266, 303 ("At block 1016 (FIG. 10A), in accordance with a determination that the first electronic device is not to respond to the audio input, the first electronic device ***foregoes responding*** to the audio input.").

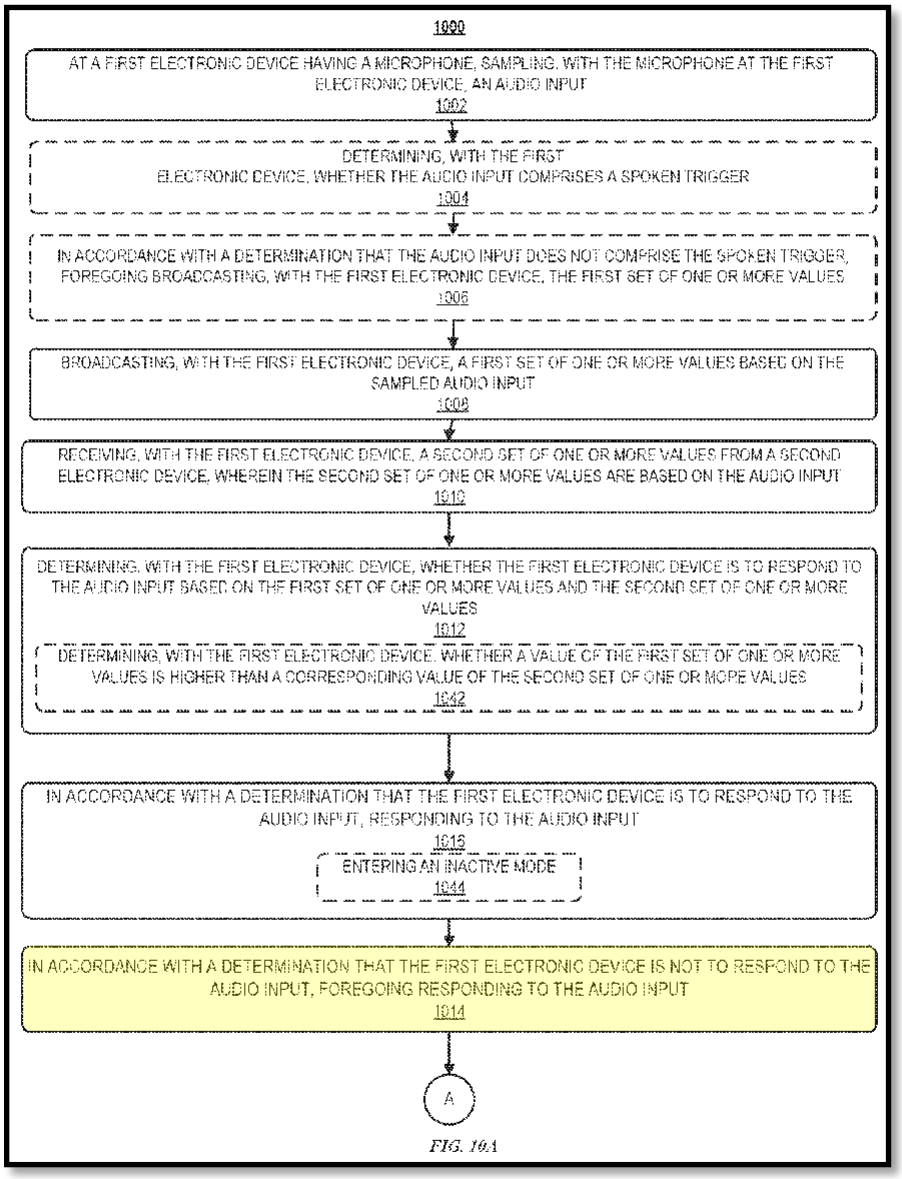


FIG. 10A

Id., FIG. 10A (annotations added).

169. Piersol's disclosure of **[2.1]** is demonstrated by an example described in connection with FIG. 8B where devices that ***do not*** determine that they have the "highest 'energy level' value" "forego responding to the audio input":

> Turning to FIG. 8B, user 800 provides audio input 812, which includes only a spoken trigger ("Hey Siri") and does not specify a task. In response, each of electronic devices 802-808 generates and broadcasts

> a set of one or more values, as described. In some examples, each of electronic devices 802-808 broadcasts an "energy level" value to each other. In some examples, each of electronic devices 802-808 calculates a single value aggregating some or all of the exemplary values disclosed herein. In this example, ***electronic device 804 determines that it has*** broadcasted ***the highest "energy level" value***, indicating that electronic device 804 is closer to user 800 than the other electronic devices. Accordingly, ***electronic device 804 responds*** to the audio input 812, and ***the rest of the electronic devices forego responding to the audio input***.

Id., ¶267.

170. Thus, it is my opinion that Piersol discloses **[2.1]**.

2. Dependent Claim 3

171. Dependent claim 3 recites *[t]he method of claim 2,* **[3.1]** *wherein forgoing responding to the detected input further comprises changing a state of operation of the first electronic device*.

172. Piersol discloses this additional element. As just discussed, Piersol discloses that each electronic device is configured such that, upon determining that its values/score is not higher than values/score received from another electronic device, it forgoes responding to the detected voice input.

173. Piersol further explains that an "electronic device can forego responding to the audio input ***by entering an inactive mode (e.g. sleep mode)***." Piersol, ¶262; *see also id.*, ¶266, 305, 336.

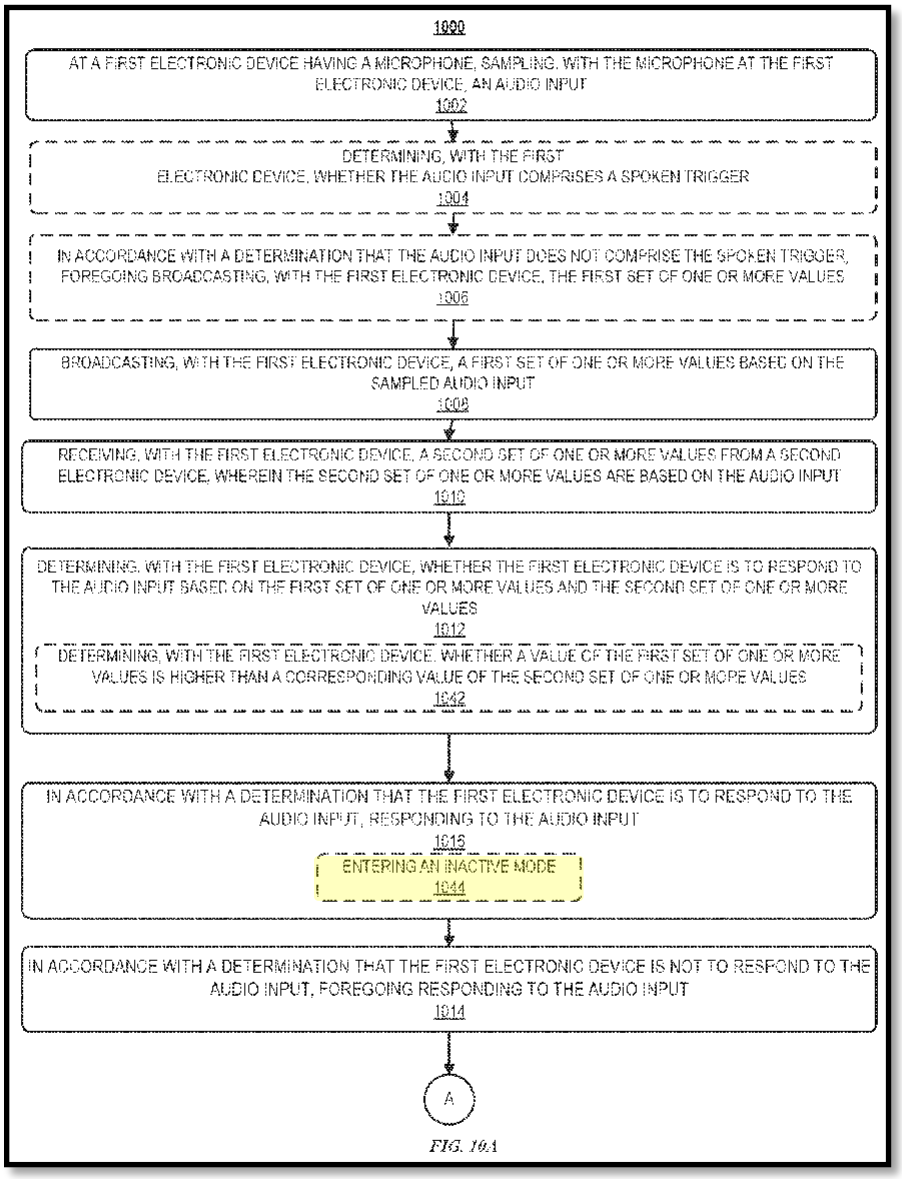


FIG. 10A

Id., FIG. 10A (annotations added).

174. A POSITA would understand that Piersol's electronic device "entering" such a mode/state of operation would involve it changing its current mode/state from an active or non-sleep mode/state to the "inactive mode (e.g. sleep mode)." Otherwise, Piersol's electronic device would simply ***remain*** in the "inactive mode."

175. Thus, it is my opinion that Piersol discloses **[3.1]**.

3. Dependent Claim 8

176. Dependent claim 8 recites *[t]he method of claim 1,* **[8.1]** *wherein responding to the detected input comprises outputting an audible and/or a visual response to the detected voice input*.

177. Piersol discloses this additional element. For instance, Piersol discloses that each electronic device is configured such that responding to a detected voice input can involve providing a visual and/or audible output. *See, e.g.*, Piersol, ¶307 ("In some examples, in accordance with the determination that the first electronic device is to respond to the audio input, the first electronic device provides ***a visual output, an auditory output***, a haptic output, or ***a combination thereof***."), ¶338; *see also* element **[16.5]**.

178. Thus, it is my opinion that Piersol discloses **[8.1]**.

4. Dependent Claim 9

179. Dependent claim 9 recites *[t]he method of claim 1,* **[9.1]** *wherein the first electronic device further comprises a speaker, and* **[9.2]** *responding to the detected input further comprises outputting an audible response*.

180. Piersol discloses these additional elements. As just discussed, Piersol discloses that each electronic device is configured such that responding to a detected voice input can involve providing an audible output.

181. Piersol discloses that each electronic device comprises a "speaker" to enable outputting such responses. *See, e.g.*, Piersol, ¶55:

> Audio circuitry 210, ***speaker 211***, and microphone 213 provide an audio interface between a user and device 200. Audio circuitry 210 receives audio data from peripherals interface 218, converts the audio data to an electrical signal, and transmits the electrical signal to ***speaker 211***. ***Speaker 211*** converts the electrical signal to ***human-audible sound*** waves.

Id., ¶82:

> Digital assistant client module 229 can also be capable of ***providing output in audio (e.g., speech output),*** visual, and/or tactile forms through various output interfaces (e.g., ***speaker 211***, touch-sensitive display system 212, tactile output generator(s) 267, etc.) of portable multifunction device 200. For example, output can be provided as voice, sound, alerts, text messages, menus, graphics, videos, animations, vibrations, and/or combinations of two or more of the above.

182. Thus, it is my opinion that Piersol discloses **[9.1]** and **[9.2]**

5. Dependent Claim 11

183. Dependent claim 11 recites *[t]he first electronic device of claim 10, the one or more programs further comprising instructions for:* **[11.1]** *recognizing a command in the voice input; and* **[11.2]** *in accordance with a determination that a type of the command is related to the first electronic device, responding to the detected voice input.*

184. Piersol discloses these additional elements.

185. **First**, Piersol discloses that each electronic device is configured to recognize a command (or "specified task" to be performed) in the voice input.

186. For instance, Piersol discloses embodiments where "the functions of a digital assistant can be implemented as a standalone application installed on [an electronic] device." Piersol, ¶46; *see also id.*, ¶215 (describing "[Speech-to-text] SST processing module 730" with "one or more ASR systems: that "produce a recognition result" and explaining, "[i]n some examples, the speech input can be processed ***at least partially… on the [electronic] device*** (e.g., device 104, 200, 400, or 600) to produce the recognition result."), ¶217, 219-20.

187. Piersol discloses such functions of a digital assistant involve performing speech recognition on a voice input and deriving the user's intent using natural language understanding. *See, e.g.*, *id.*, ¶83 ("[U]ser data and models 231 can includes various models (e.g., ***speech recognition models***, statistical language models, ***natural language processing models***, ontology, task flow models, service models, etc.) for processing user input and ***determining user intent***.").

188. As explained in further detail below, Piersol discloses that "[d]evice type is particularly relevant to intelligent device arbitration when the ***user's input specifies a task to be performed***, such as 'Hey Siri, find me a TV episode' or 'Hey Siri, drive me there'." *Id.*, ¶259. Piersol explains that an electronic device can identify the user's specified task "by locally analyzing the audio input…." *Id.*, ¶260. In other words, "spoken instruction 834 can be processed and resolved into one or more tasks locally…." *Id.*, ¶264.

189. Thus, it is my opinion that Piersol discloses **[11.1]**.

190. **Second**, Piersol discloses that each electronic device is configured to respond to the voice input upon determining that a type of the command (or "specified task") is related to the given electronic device. *See, e.g.*, Piersol, ¶260 ("In some examples, the electronic device obtains a list of acceptable ***device types*** that can ***handle the specified task*** (e.g., by locally analyzing the audio input and/or by receiving the list from one or more servers), and ***determines whether the electronic device is to respond to*** the audio input based on the broadcasted 'device type' values and the list of acceptable device types.").

191. Piersol provides an example where "each of the electronic devices 802-808 determines whether to respond to the spoken instruction based on the device type values broadcasted by electronic devices 802-808":

> Each of the electronic devices 802-808 ***compares a respective "device type" value(s)*** with the values broadcasted by other electronic devices. Device type is particularly relevant to intelligent device arbitration ***when the user's input specifies a task to be performed***, such as "Hey Siri, find me a TV episode" or "Hey Siri, drive me there". In some examples, ***the electronic device is to respond to the audio input if***, based on the broadcasted values, ***the electronic device is the only device of a type that can handle the task specified in the audio input***.

Piersol, ¶259.

192. Piersol explains that its arbitration process of determining whether to respond can look at "device type" separate from another score/value of the electronic device or use "device type" as part of determining a single, aggregated score/value:

> It will be appreciated that the arbitration processes described herein are exemplary, and that, using one or more numerical and/or logical functions and algorithms, some or all of the values above can be factored, alone or in combination, into the determination of whether an electronic device is to respond to an audio input. In some examples, each of electronic devices 802-808 broadcasts ***multiple values (e.g., an aggregated score and a "device type" value)***. Accordingly, electronic devices 802-808 may arbitrate in accordance with one or more predetermined algorithms. In some examples, each of electronic devices 802-808 broadcasts a single score, which is calculated according to one or more functions using some or all of the above-discussed values. For example, ***a single score*** for broadcasting can be calculated based on ***a predetermined weighting of received audio quality (e.g., signal-to-noise ratio), type of device***, ability to perform task, and device state.

Piersol, ¶261.

193. In connection with FIG. 10B, Piersol discloses **[11.2]:**

> In some examples, the audio input comprises an additional input indicative of a task. At block 1018 (FIG. 10B), the first electronic device additionally ***determines whether a type of the first electronic device meets a requirement of the task***…. At block 1022, ***in accordance with a determination that the type of the first electronic device meets the requirement of the task, the first electronic device responds to the audio input***.

Id., ¶301-302.

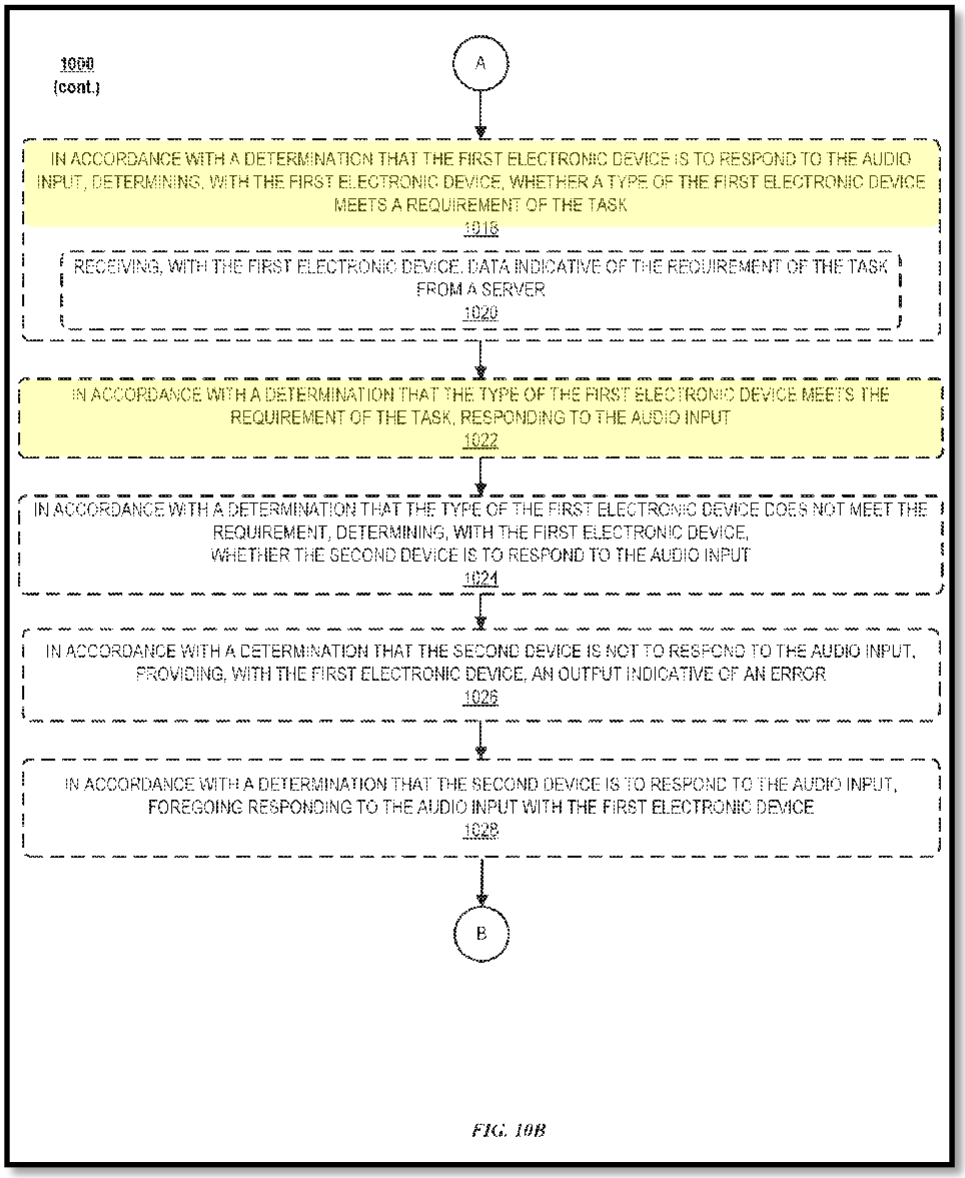


FIG. 10B

Id., FIG. 10B (annotations added); *see also id.*, ¶303-304, 331.

194. Thus, it is my opinion that Piersol discloses **[11.2]**.

195. Piersol provides an example that demonstrates how it discloses the additional elements of claim 11 where "electronic device 806" responds to a voice input despite not having the highest "'energy level' value" because it determines that it is of a device type that can handle the task the user requested be performed:

With reference to FIG. 8A, each of electronic device 802 (implemented as a smart watch), electronic device 804 (implemented as a mobile phone), electronic device 808 (implemented as a speaker dock), and electronic device 806 (implemented as a television) generates and broadcasts a set of one or more values after sampling spoken instruction 834 from user 800. In this example, each set of one or more values includes an "energy level" value and a "device type" value. By comparing the broadcasted "energy level" values, ***electronic device 804 determines that it has broadcasted the highest "energy level" value*** among electronic devices 802-808, indicating that electronic device 804 is in relatively close proximity to user 800.

However, ***electronic device 804 further analyzes the broadcasted "device type" values in light of the task specified in the spoken instruction 834***. As discussed above, spoken instruction 834 can be processed and resolved into one or more tasks locally or remotely. For example, electronic device 804 can resolve the spoken instruction 834 into a task locally, or receive the task along with a list of acceptable device types for handling the task from one or more servers. In this example, ***electronic device 804 determines that it is not of an acceptable (or preferred) device type for handling the task specified in the audio input*** (i.e., playback of a video). As such, electronic device 804 foregoes responding to the audio input.

* * *

Similarly, electronic device 806 (implemented as a television) receives respective sets of one or more values from electronic devices 802 (implemented as a smart watch), 808 (implemented as a speaker dock), and 804 (implemented as a mobile phone). By analyzing the broadcasted sets of values, ***electronic device 806 determines that it has not broadcasted the highest "energy level" value***. But, ***electronic device 806 further determines that it is of a device type that can handle the task of video playback***. In accordance with the determination, ***electronic device 806 makes an additional determination of whether it is to respond to the audio input despite not having broadcasted the highest "energy level" value***. For example, if electronic device 806 determines that none of the electronic devices that have broadcasted higher "energy level" values (e.g. electronic device 804) are of acceptable device types for handling the task, ***electronic device 806 responds to the audio input***. On the other hand, if electronic device 806 determines that there is at least one electronic device that has

> broadcasted a higher "energy level" value and is of an acceptable device type for handling the task, electronic device 806 foregoes responding to the audio input.

Piersol, ¶263-64, 266.

6. Dependent Claim 14

196. Dependent claim 14 recites *[t]he first electronic device of claim 10, wherein:* **[14.1]** *the plurality of electronic devices is communicatively coupled through a local network; and* **[14.2]** *the receiving is performed through the local network.*

197. Piersol discloses these additional elements.

198. **First**, Piersol discloses that the electronic devices are communicatively coupled through a local network, such as a Wi-Fi or Bluetooth local network. *See, e.g.*, Piersol, ¶42 ("Examples of communication network(s) 110 can include ***local area networks (LAN)***…."), ¶44 ("User device 104 can be configured to communicatively couple to second user device 122 via a direct communication connection, such as Bluetooth, NFC, BTLE, or the like, or via a wired or wireless network, such as a ***local Wi-Fi network***."), ¶247:

> The sets of one or more values can be broadcasted by electronic devices 802-808 via ***any unidirectional broadcast communications standards, protocols, and/or technologies*** known now or in the future. In some examples, one or more sets of one or more values are broadcasted using ***Bluetooth Low Energy (BTLE)*** advertising mode. Other communication methods can be used, including but not limited to ***WiFi (e.g., any 802.11 compliant communication)***, NFC, infrared, sound wave, etc., or any combination thereof.

199. Thus, it is my opinion that Piersol discloses **[14.1]**.

200. **Second**, Piersol discloses that each electronic device is configured to receive one or more values or a score broadcasted by another electronic device through the local network:

> The sets of one or more values can be ***broadcasted*** by electronic devices 802-808 via any unidirectional ***broadcast*** communications standards, protocols, and/or technologies known now or in the future. In some examples, one or more sets of one or more values are ***broadcasted*** using ***Bluetooth Low Energy (BTLE)*** advertising mode. Other communication methods can be used, including but not limited to ***WiFi (e.g., any 802.11 compliant communication)***, NFC, infrared, sound wave, etc., or any combination thereof.

Piersol, ¶247.

201. Thus, it is my opinion that Piersol discloses **[14.2].**

7. Dependent Claim 15

202. Dependent claim 15 recites *[t]he first electronic device of claim 10,* **[15.1]** *wherein the first electronic device is configured to be awakened by a plurality of affordances including a voice-based affordance; and* **[15.2]** *wherein detecting the voice input is in response to the awakening of the first electronic device*.

203. Piersol discloses these additional elements.

204. **First**, Piersol discloses that each electronic device is configured to be awakened by a plurality of mechanisms including a voice-based affordance, a touch-based affordance, a user-proximity-based affordance, and a device-position based affordance, among other affordances. *See, e.g.*, Piersol, ¶246 ("In some examples,

one or more of electronic devices 802-808 begin to sample audio input ***in response to detecting proximity of user 800***. In some examples, one or more of electronic devices 802-808 begin to sample audio input ***in response to a particular input by user 800, such as a spoken trigger***."), ¶156 ("Push button 306 is, optionally, ***used to turn the power on/off*** on the device by depressing the button and holding the button in the depressed state for a predefined time interval[.]"), ¶245 ("In some examples, electronic device 802 is a wearable electronic device, such as a smart watch, and is, optionally, ***powered off*** when in ***a lowered position***, as illustrated."), ¶269 ("[U]ser 800 lifts electronic device 802 (implemented as a smart watch) into a ***raised position*** and then provides audio input 814 ('Hey Siri').").

205. Thus, it is my opinion that Piersol discloses **[15.1]**.

206. **Second**, Piersol discloses that each electronic device is configured such that detecting a voice input is responsive to the awakening of the electronic device. *See, e.g.*, Piersol ¶246 ("In some examples, one or more of electronic devices 802-808 ***begin to sample audio input*** in response to detecting proximity of user 800. In some examples, one or more of electronic devices 802-808 ***begin to sample audio input*** in response to a particular input by user 800, such as a spoken trigger."), ¶269 ("[U]ser 800 lifts electronic device 802 (implemented as a smart watch) into a ***raised position and then provides audio input*** 814 ('Hey Siri').").

207. Thus, it is my opinion that Piersol discloses **[15.2]**.

8. Dependent Claim 17

208. Dependent claim 17 recites *[t]he system of claim 16,* **[17.1]** *wherein the first value comprises a first quality score of the voice input.*

209. Piersol discloses this additional element. As explained before, Piersol discloses that each electronic device is configured to determine a "score" (or one or more "values") that can be "calculated based on a predetermined weighting of received ***audio quality*** (e.g., ***signal-to-noise ratio***), type of device, ability to perform task, and device state." Piersol, ¶261; *see also, e.g., id.*, ¶248 ("One exemplary ***value indicates an energy level*** of the audio input as sampled on the electronic device…. In some examples, ***the energy level is measured using known metrics for audio quality***, such as ***signal-to-noise ratio***, sound pressure, or a combination thereof."), ¶297 ("At block 1008, the first electronic device broadcasts a first set of one or more values based on the sampled audio input. In some examples, a value of the first set of values is based on ***a signal to noise ratio of speech*** of the audio input sampled with the first electronic device.").

210. Claim 19 expressly recites that the *first quality score comprises a signal-to-noise rating of detection of the voice input*, and the '311 Patent explains that "signal-to-noise ratio" is one example of *a signal-to-noise rating*. '311 Patent, 30:47-50.

211. Thus, it is my opinion that Piersol discloses **[17.1].**

9. Dependent Claim 18

212. Dependent claim 18 recites *[t]he system of claim 17,* **[18.1]** *wherein the first quality score comprises a confidence level of detection of the voice input.*

213. As noted above, Google broadly contends *a confidence level of detection* "is a measure reflecting the confidence that the audio input is voice." Ex.1020, p.6 (document page 89). In other words, Googles broadly contends *a confidence level of detection* "is a measure of the likelihood that an audio signal includes speech, and indicates a level of confidence that a voice input was detected." Ex.1021, p.6 (document page 70).

214. Piersol discloses this additional element, especially under Google's interpretation. For instance, Piersol discloses that each electronic device is configured to determine a "score" (or one or more "values") that can comprise "a confidence value indicative of a likelihood that the audio input was provided by a particular user":

> Another exemplary value indicates an acoustic fingerprint of the audio input, that is, whether the audio input is ***likely*** provided by a particular user. In some examples, the electronic device analyzes the audio input and calculates ***a confidence value*** that expresses the likelihood that the audio input originates from the particular user (e.g., a user that has enrolled in the virtual assistant service on the electronic device). In some examples, the electronic device determines whether the audio input is from an authorized user by comparing the confidence value with a predetermined threshold value, and broadcasts a value based on the comparison. In some examples, the set of one or more values may include ***a value corresponding to the confidence value*** as well.

Piersol, ¶249.

> At block 1008, the first electronic device broadcasts a first set of one or more values based on the sampled audio input.… In some examples, the first electronic device identifies a ***confidence value indicative of a likelihood that the audio input was provided by a particular user***, and a value of the first set of values is based on the confidence value.

Id., ¶297.

215. A POSITA would understand that Piersol's "confidence value" is a "measure ***reflecting*** the confidence that the audio input is voice." Ex.1020, p.6 (document page 89). In this respect, Piersol's "confidence value" provides a likelihood that audio input was provided by a particular user, which by definition ***reflects*** a likelihood that what was detected was voice, as opposed to some non-voice sound. If Piersol's electronic device detected non-voice, Piersol's "confidence value" would be extremely low or would not even be calculated, whereas if Piersol's electronic device detects voice, Piersol's "confidence value" would be relatively higher and extremely higher when that voice matches a voice model corresponding to a particular user's voice. In this way, Piersol's "confidence value" is a ***specific*** type of a "measure ***reflecting*** the confidence that the audio input is voice."

216. I note that the "confidence value" disclosed by Piersol is no different than the only example of a "confidence level" disclosed in the '311 Patent.

> Based on the comparing, the electronic device 190 determines (1106) that ***the first voice input corresponds to a first user*** of the plurality of users. For example, the user identification module 560 identifies a voice model in voice models data 550 that best matches the first voice input,

> and in accordance with the identification of the match determines that the user speaking the first voice input is the user to which the matching voice model corresponds. In some implementations, the user identification module 560 also determines ***a confidence level*** or some other similar measure of the quality or closeness of the match between a voice model and the voice input, and identifies a match only if the match is the best and the confidence level is above a predefined threshold.

'311 Patent, 24:35-48.

217. Thus, it is my opinion that Piersol discloses **[18.1]**.

10. Dependent Claim 19

218. Dependent claim 19 recites *[t]he system of claim 17,* **[19.1]** *wherein the first quality score comprises a signal-to-noise rating of detection of the voice input*.

219. For the reasons I explained for dependent claim 17, Piersol discloses this additional element. *Supra* ¶208-11.

11. Dependent Claim 20

220. Dependent claim 20 recites *[t]he system of claim 16,* **[20.1]** *wherein each of the respective second values is indicative of a quality of the voice input detected by a respective one of the other electronic devices.*

221. For the reasons I explained for dependent claim 17, Piersol discloses this that its score or one or more values *is indicative of a quality of the voice input detected* by a given electronic device. *Supra* ¶208-11.

222. And for the reasons I explained for **[16.3]**, Piersol discloses that each device is configured to receive a respective score or one or more values *indicative*

of a quality of the voice input detected from another electronic device. *Supra* ¶145-49.

223. Thus, it is my opinion that Piersol discloses **[20.1]**.

C. Ground 1B: Piersol + Meyers Render Obvious Independent Claims 1, 10, 16 and Dependent Claims 2-3, 8-9, 11-12, 14-15, 17-20

224. Piersol combined with Meyers renders obvious independent claims 1, 10, and 16 and dependent claims 2-3, 8-9, 11-12, 14-15, and 17-20.

225. In particular, for the independent claims, Piersol combined with Meyers is presented to the extent that Google argues that Piersol does not adequately disclose **[1.1]**, **[10.1]**, **[16.1]** *detect[ing] a voice input*. Other than Meyers making up for this hypothetical deficiency of Piersol with regard to **[1.1]**, **[10.1]**, **[16.1]**, the combination of Piersol and Meyers invalidates independent claims 1, 10, and 16 and dependent claims 2-3, 8-9, 11, 14-15, and 17-20 for the same reasons I explained before in **Ground 1A**.

226. Additionally, Piersol alone, or combined with Meyers, invalidates independent claim 10, and Piersol combined with Meyers renders obvious dependent claim 12.

227. Lastly, while Piersol itself discloses the additional limitation of dependent claim 18, Piersol combined with Meyers renders this additional limitation obvious as well. Thus, Piersol alone, or combined with Meyers, invalidates independent claim 16 and dependent claim 17, and Piersol combined with Meyers

renders obvious dependent claim 18 that ultimately depends on both claims 16 and 17.

1. Piersol + Meyers Render Obvious *Detecting a Voice Input*

228. As explained below, Meyers teaches *detect[ing] a voice input*, and a POSITA would have found it obvious to incorporate such teachings with Piersol's teachings.

a. Meyers Discloses Detecting a Voice Input

229. Meyers discloses an electronic device **[1.1]**, **[10.1]**, **[16.1]** *detect[ing] a voice input*.

230. In particular, Meyers discloses a system of "speech interface devices" (*see, e.g.*, Meyers, FIG. 1), each of which is configured to "detect" voice input containing "a user request 106." *See, e.g.*, *id.*, ¶15 ("In the described embodiments, each speech interface device ***detects*** that a user is speaking a command…."), ¶21 ("The devices 102 may be located near enough to each other so that both of the devices 102 may ***detect*** an utterance of the user 104.").

231. Meyers further explains that "the user request 106 may be prefaced by a wakeword or other trigger expression that is spoken by the user 104 to indicate that subsequent user speech is intended to be received and acted upon by one of the devices 102." *Id.*, ¶23. Meyers continues by explaining "[t]he device 102 may ***detect*** the wakeword and interpret subsequent user speech as being directed to the

device 102" where "[a] wakeword in certain embodiments may be a reserved keyword that is ***detected*** locally by the speech interface device 102." *Id.*

232. In particular, Meyers' speech interface device includes a "voice activity detector 922" that performs "voice activity detection" and "wake word detector 920" that performs "wakeword detection":

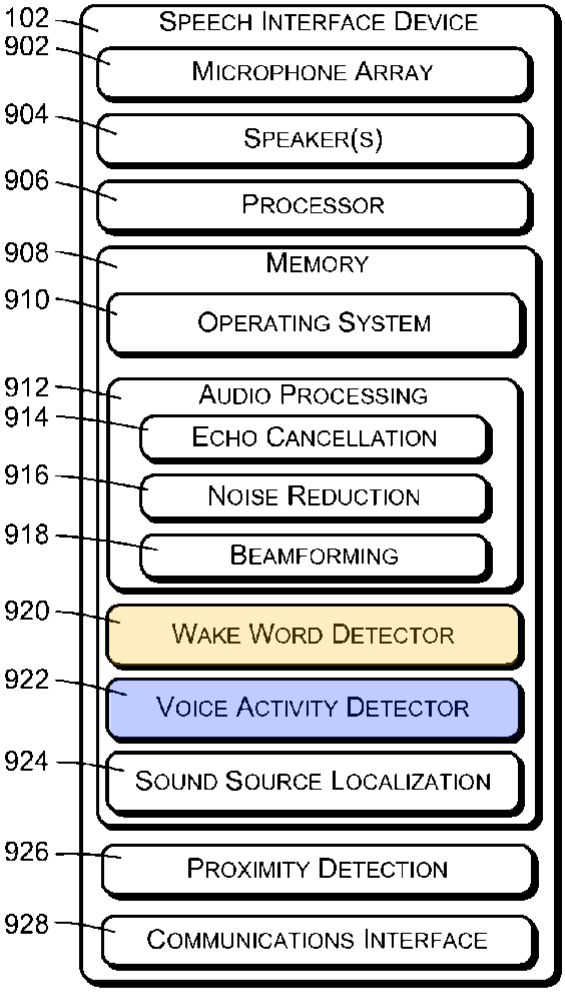


FIG. 9

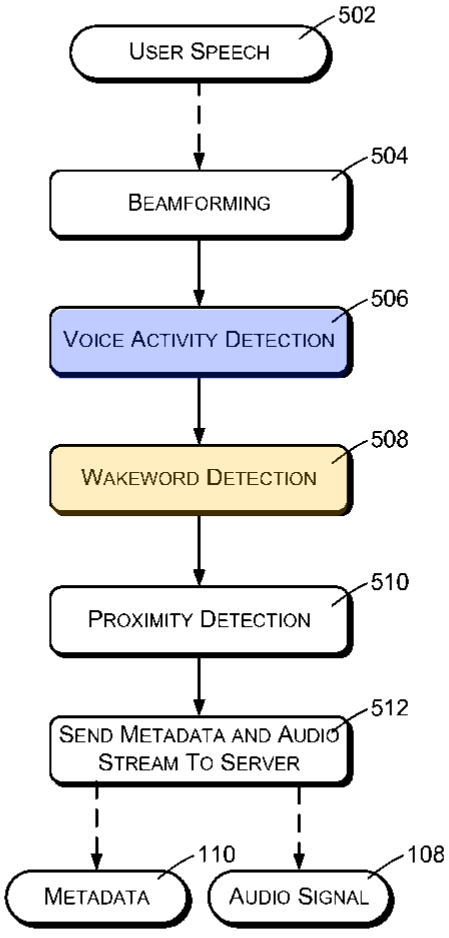


FIG. 5

Id., FIGs. 5, 9 (annotations added).

233. With respect to the voice activity detector and voice activity detection (VAD), Meyers explains:

> VAD ***determines*** the level of ***voice presence*** in an audio signal by analyzing a portion of the audio signal to evaluate features of the audio signal such as signal energy and frequency distribution. The features are quantified and compared to reference features corresponding to reference signals that are known to contain human speech. The comparison produces a score corresponding to the degree of similarity between the features of the audio signal and the reference features. The score is used as an indication of the detected or likely level of speech presence in the audio signal.

Id., ¶102; *see also id.*, ¶120.

234. With respect to the wake word detector (or "expression detector") and wakeword detection, Meyers explains:

> In certain implementations, each device 102 may have an expression detector that analyzes an audio signal produced by a microphone of the device 102 to ***detect the wakeword***, which generally may be a predefined word, phrase, or other sound. Such an expression detector may be implemented using keyword spotting technology, as an example. A keyword spotter is a functional component or algorithm that evaluates an audio signal to ***detect the presence*** a predefined word or expression in the audio signal. ***Rather than producing a transcription*** of the words of the speech, ***a keyword spotter generates a true/false*** output to indicate whether or not the predefined word or expression was represented in the audio signal.

Id., ¶24; *see also id.*, ¶104, 119.

235. Meyers also explains that wakeword detection may be performed on an audio signal "within which voice activity has been detected…." *Id.*, ¶103.

236. Upon detecting a voice signal, Meyers discloses that each speech interface device is configured to determine "metadata" associated with the detected

voice signal that includes "one or more signal attributes." *See, e.g., id.*, ¶16, 40. Meyers discloses numerous signal attributes, including (i) "the amplitude of the audio signal," (ii) "the signal-to-noise ratio of the audio signal," (iii) "the level of voice presence detected in the audio signal," and (iv) "the confidence level with which a wakeword was detected in the audio signal," among other examples. *Id.*, ¶40; *see also id.*, ¶70, 108.

237. Thus, it is my opinion that Meyers discloses **[1.1]**, **[10.1]**, **[16.1]**.

b. A POSITA Would Have Been Motivated to Combine Piersol with Meyers

238. In my opinion, a POSITA would have been motivated to combine Piersol with Meyers before the effective priority date of the '311 Patent (Oct. 3, 2016) such that Piersol's electronic device would detect a voice input utilizing Meyers's "voice activity detector" and/or local "wake word detector" (Meyers, FIG. 9), which in turn would "initiate [Piersol's] arbitration process to identify (e.g., determine) an electronic device for responding to the spoken instruction [] from user 800." Piersol, ¶247. In this way, a POSITA would achieve **[1.1]**, **[10.1]**, **[16.1]** (again, assuming Piersol does not already disclose these elements). It is my opinion that a POSITA would have been motivated to make this combination for a variety of reasons.

239. **First**, Piersol and Meyers are in the same field of endeavor, relate to voice-enabled devices, 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.,* Piersol, ¶3:

> Many contemporary electronic devices offer virtual assistant services to perform various tasks in response to spoken user inputs. In some circumstances, ***multiple electronic devices*** having virtual assistant services may concurrently operate in a shared environment. As a result, a user input may cause each of the ***multiple electronic devices to respond*** when the user input contains a trigger phrase or command recognized by each of the virtual assistant services on the electronic devices. This in turn may result in ***a confusing experience for the user***, as multiple electronic devices may simultaneously begin to listen and/or prompt for additional input. Further, ***the multiple electronic devices may perform duplicative or conflicting operations*** based on the same user input.

Meyers, ¶14:

> In situations in which multiple speech interface devices are near each other, such as within a single room or in adjoining rooms, each of the speech interface devices may receive a user utterance and each device may independently attempt to process and respond to the user utterance as if it were two separate utterances. The following disclosure relates to techniques for ***avoiding such duplicate efforts and responses***, among other things.

240. **Second**, Piersol and Meyers have interrelated teachings, both disclosing very similar "arbitration" processes to solve the duplicate response problem. *See, e.g.,* Piersol, Title ("Intelligent device ***arbitration*** and control"), ¶274 ("In response to detecting spoken instruction 834, electronic devices 802-808 initiate an ***arbitration process*** to identify (e.g., determine) an electronic device for responding to the spoken instruction 834 from user 800."); Meyers, Abstract ("***[A]rbitration*** is employed to select one of the multiple speech interface devices to

respond to the user utterance."), ¶66 ("The device whose audio signal 108 has the strongest set of attributes is selected as the winner of the ***arbitration***.").

241. Indeed, Piersol and Meyers both disclose arbitration techniques where electronic devices determine, receive, and compare signal-to-noise ratio (SNR) information associated with a detected voice input and select the electronic device with the highest SNR to respond to the user. *See, e.g.*, Piersol, ¶248, 256; Meyers, ¶44, 70. A POSITA studying Piersol would have been motivated to seek out other references that disclose similar arbitration techniques, like Meyers, to obtain additional details regarding such techniques.

242. **Third**, Piersol already discloses **[1.1]**, **[10.1]**, **[16.1]** *detect[ing] a voice input*. *Supra* ¶129-35. But Piersol does not provide implementation details regarding this functionality. A POSITA studying Piersol would have been motivated to seek out other references that disclose implementation details for **[1.1]**, **[10.1]**, **[16.1]**, like Meyers, to achieve a fuller understanding, especially when the POSITA was implementing a system in accordance with Piersol's teachings.

243. **Fourth**, Piersol and Meyers both disclose techniques where arbitration is initiated based on detecting a voice input, and Meyers's methods of performing this function (e.g., via VAD and/or hotword/wakeword detection) were well known ways to detect a voice input before the effective priority date of the '311 Patent (Oct. 3, 2016). *Supra* ¶229-37.

244. For instance, other prior art such as Basye filed on December 11, 2012 discloses a "speech detection module 108" configured to "determine whether [an] audio input includes speech" utilizing various techniques such as "voice activity detection (VAD) techniques" (Basye, ¶21) and a "speech processing module 110" configured to "detect a keyword in the speech" (*id.*, ¶21), which is described as being separate and distinct from "an automatic speech recognition engine" responsible for recognizing speech responsive to "speech processing module 110" detecting a keyword (*id.*, ¶54).

245. **Fifth**, a POSITA would have appreciated that applying Meyers' disclosures of **[1.1]**, **[10.1]**, **[16.1]** to Piersol would have been nothing more than applying known methods to similar electronic devices to achieve predictable results.

246. Notably, I see nothing in the '311 specification that would suggest to a POSITA that *detect[ing] a voice input* under Google's narrow interpretation was anything more than a design choice among other choices already in the prior art. In fact, I see no discussion in the '311 specification of Google's narrow interpretation, much less any discussion in the '311 specification of *detect[ing] a voice input* under Google's narrow interpretation conferring some significant advantage to the '311 claims.

247. In this way, a POSITA would have had at least a reasonable expectation of success, especially because Meyers itself already discloses using such techniques for detecting a voice input in performing arbitration in a very similar way as Piersol.

248. For at least these reasons, it is my opinion that a POSITA would have been motivated to combine Piersol with Meyers before the effective priority date of the '311 Patent (Oct. 3, 2016).

2. Piersol + Meyers Render Obvious Dependent Claim 12

249. Piersol alone, or combined with Meyers, invalidates independent claim 10, and Piersol combined with Meyers renders obvious dependent claim 12.

250. In this regard, as explained below, Meyers discloses dependent claim 12, and a POSITA would have found it obvious to incorporate such teachings with Piersol's teachings.

a. Meyers Discloses [12.1]

251. Dependent claim 12 recites *[t]he first electronic device of claim 10, the one or more programs further comprising instructions for:* **[12.1]** *communicating to the other devices of the plurality of electronic devices to forgo responding to the detected output.*[6]

[6] For the purposes of this IPR, I have been asked to assume that *the detected* ***output*** in claim 10 refers to *the detected* ***input*** in independent claim 10.

252. Meyers discloses a "speech interface device" capable of performing **[12.1]**. In particular, as explained, Meyers discloses that a "speech service 112" performs the disclosed arbitration techniques and that one or more speech interface devices may include or provide the speech service 112. *Supra* ¶97.

253. Meyers further explains that, when the speech service 112 determines that a given speech interface device loses arbitration, the speech service 112 sends a message to that speech interface device that cancels any response that otherwise might have been given by the speech interface device:

> If the first device 102(a) wins the arbitration, an action 214 is performed of processing and responding to the first audio signal 108(a), including producing an appropriate response by the first device 102(a) to the user command represented by the first audio signal 108(a). An action 216 comprises canceling the processing of the second audio signal 108(b) and ***canceling any response*** that might otherwise have been provided based on the second audio signal 108(b), including ***any response that might have otherwise been given by the device 102(b)***. In some implementations, ***a message is sent to the device 102(b)*** informing the device 102(b) not to expect a further response from the speech service 112.

Meyers, ¶67.

254. In some embodiments, Meyers explains that such a message may cause a device that lost arbitration to enter a "listening mode" and/or "take some… action indicating that the device is not going to respond to the user request":

> When operation of a processing pipeline instance is aborted, the aborted pipeline instance does not provide a response 114 to the corresponding device 102. The aborted pipeline instance may also ***provide*** an [*sic*] ***message to the device 102, indicating that the device 102 will not be used to provide a response to the user request***. In response, the device

> may stop providing the audio signal 108 to the speech service 112. As an example, the message or other indication may comprise data including an instruction that causes or results in the device entering a ***listening mode***.… In some cases, the device may be ***instructed to*** play a tone, produce an LED illumination, or take ***some other action indicating that the device is not going to respond to the user request***.

Meyers, ¶53.

255. In this way, Meyers informs a POSITA that there could be instances where the "speech service 112" is implemented on a "speech interface device" and that "speech interface device" wins arbitration for a given voice input, which would cause that "speech interface device" to send cancellation messages to all the other "speech interface devices" informing them to forgo responding to the voice input.

256. Thus, it is my opinion that Meyers discloses **[12.1]**.

b. A POSITA Would Have Been Motivated to Combine Piersol with Meyers

257. In my opinion, a POSITA would have been motivated to combine Piersol with Meyers before the effective priority date of the '311 Patent (Oct. 3, 2016) such that Piersol's electronic device (alone or as modified by Meyers' teachings regarding **[10.1]**) would be configured to communicate to other electronic devices in a system to forgo responding to detected voice input upon the electronic device determining it won arbitration, as taught by Meyers. In this way, a POSITA would achieve **[12.1]**. It is my opinion that a POSITA would have been motivated to make this combination for a variety of reasons.

258. **First**, as explained before, Piersol and Meyers are in the same field of endeavor, relate to voice-enabled devices, and seek to address the same problem – namely, the duplicate response problem when multiple voice-enabled devices detect the same user utterance. *Supra* ¶239.

259. **Second**, as also explained, Piersol and Meyers have interrelated teachings, both disclosing very similar "arbitration" processes to solve the duplicate response problem. *Supra* ¶240-41. Again, a POSITA studying Piersol would have been motivated to seek out other references that disclose similar arbitration techniques, like Meyers, to obtain additional details regarding such techniques.

260. **Third**, a POSITA would have been motivated to incorporate Meyers' cancellation message when implementing a system in accordance with Piersol's teachings on a Wi-Fi network (*see, e.g.*, Piersol, ¶247) because such networks were known to have message/packet loss whereby transmitted messages/packets did not reach their intended recipient. Ex.1014, p.1 ("Packet loss is [a] condition in which data packets are transmitted correctly at source end, but never arrive at the destination end. This might be because of network conditions are poor or… because of internet congestion."); Ex.1015, p.1 ("Packet loss represents the percentage of packets that do not make it to their destination or those that arrive with errors and are therefore useless.... On wireless networks, interference and collision can be significant contributors to packet loss.").

261. In this regard, Piersol's arbitration winner would communicate (e.g., periodically for the duration of the interaction with user) with the other electronic devices for them to forgo responding to detected voice input (e.g., by using Meyers' cancellation message) to account for noise or other interference in the Wi-Fi network causing Piersol's winner's score/values to not be received by each of the other electronic devices in the first instance. In this way, even if the winner's score/values was not received by each other electronic device (which would mean one other electronic device would believe it was the arbitration winner despite having a lower score than the true winner), the cancellation message would still achieve Piersol's stated objective of eliminating user confusion and excessive power consumption resulting from "redundant responses." *See, e.g.*, Piersol, ¶3, 33.

262. **Fourth**, a POSITA would have appreciated that applying Meyers' disclosures of **[12.1]** to Piersol would have been nothing more than applying known methods to similar electronic devices to achieve predictable results.

263. Notably, I see nothing in the '311 specification that would suggest to a POSITA that *communicating to the other devices of the plurality of electronic devices to forgo responding to the detected [input]* was anything more than a design choice already in the prior art. In fact, I see no discussion in the '311 specification itself about this limitation at all, much less any discussion in the '311 specification of it conferring some significant advantage to the '311 claims.

264. In this way, a POSITA would have had at least a reasonable expectation of success, especially because Meyers itself already discloses using such techniques in performing arbitration in a similar way as Piersol.

265. For at least these reasons, it is my opinion that a POSITA would have been motivated to combine Piersol with Meyers before the effective priority date of the '311 Patent (Oct. 3, 2016).

3. Piersol + Meyers Render Obvious Dependent Claim 18

266. Piersol alone, or combined with Meyers, invalidates independent claim 16 and dependent claim 17, and Piersol combined with Meyers renders obvious dependent claim 18.

267. As explained before, Piersol discloses a *first quality score* that can comprise a variety of information including "a confidence value indicative of a likelihood that the audio input was provided by a particular user." *Supra* ¶214-16; Piersol ¶297. As explained below, Meyers also discloses a quality score comprising *a confidence level of detection of the voice input*, and a POSITA would have found it obvious to incorporate such teachings with Piersol's teachings.

a. Meyers Discloses [18.1]

268. As explained, Meyers discloses that each "speech interface device" is configured such that, upon detecting a voice signal, it determines "metadata" associated with the detected voice signal that includes "one or more signal

attributes," such as (i) "the level of voice presence detected in the audio signal" and (ii) "the confidence level with which a wakeword was detected in the audio signal," among other examples. *Supra* ¶104.

269. With regard to "the level of voice presence detected in the audio signal" attribute, Meyers describes this attribute as a likelihood of detection of speech:

> VAD determines the level of voice presence in an audio signal by analyzing a portion of the audio signal to evaluate features of the audio signal such as signal energy and frequency distribution. The features are quantified and compared to reference features corresponding to reference signals that are known to contain human speech. The comparison produces ***a score corresponding to the degree of similarity*** between the features of the audio signal and the reference features. The score is used as ***an indication of the detected or likely level of speech presence in the audio signal***.

Meyers, ¶102; *id.*, ¶108 ("[T]he VAD 506 may produce a voice presence level, ***indicating the likelihood a person is speaking*** in the vicinity of the device 102.").

270. With regard to "the confidence level with which a wakeword was detected in the audio signal" – or "the level of confidence with which a wakeword or other trigger expression was detected" (*id.*, ¶70) – attribute, Meyers explains "[t]he wakeword [detector] may produce ***a wakeword confidence level***, corresponding to ***the likelihood that the user 104 has uttered the wakeword***." *Id.*, ¶108; *see also id.*, ¶27, 106, 119. In this way, Meyers describes this attribute as a likelihood of detection of speech.

271. Each of Meyers' likelihood attributes – i.e., (i) "the level of voice presence detected in the audio signal" and (ii) "the confidence level with which a wakeword was detected in the audio signal" – satisfies *a confidence level of detection of [a] voice input* under Google's broad interpretation of this claim limitation.

b. A POSITA Would Have Been Motivated to Combine Piersol with Meyers

272. In my opinion, a POSITA would have been motivated to combine Piersol with Meyers before the effective priority date of the '311 Patent (Oct. 3, 2016) such that Piersol's electronic device would determine and include either or both of Meyers' likelihood attributes as part of Piersol's arbitration score/values used in Piersol's arbitration techniques. In this way, a POSITA would achieve Piersol's electronic device (alone or as modified by Meyer's teachings regarding **[16.1]**) configured to **[16.2]** *determine a first value associated with the detected voice input*, **[17.1]** *wherein the first value comprises a first quality score of the voice input*, and **[18.1]** *wherein the first quality score comprises a confidence level of detection of the voice input* (again, assuming Piersol does not already disclose **[18.1]**). It is my opinion that a POSITA would have been motivated to make this combination for a variety of reasons.

273. **First**, as explained before, Piersol and Meyers are in the same field of endeavor, relate to voice-enabled devices, and seek to address the same problem –

namely, the duplicate response problem when multiple voice-enabled devices detect the same user utterance. *Supra* ¶239.

274. **Second**, as also explained, Piersol and Meyers have interrelated teachings, both disclosing very similar "arbitration" processes to solve the duplicate response problem. *Supra* ¶240-41. Again, a POSITA studying Piersol would have been motivated to seek out other references that disclose similar arbitration techniques, like Meyers, to obtain additional details regarding such techniques.

275. **Third**, Piersol itself explains that its arbitration score/values can take various forms and include a variety of information. *See, e.g.*, Piersol, ¶253, 261. In fact, as noted, Piersol already discloses that an arbitration score/values could include a "confidence value." *Supra* ¶214.

276. In this way, Piersol itself provides a POSITA a motivation to modify its arbitration score/values, and Piersol and Meyers make clear that use of a "confidence level" for a detected audio input in voice applications was well known before the effective priority date of the '311 Patent (Oct. 3, 2016). As such, a POSITA would have appreciated that applying Meyers' disclosures of **[18.1]** to Piersol would have been nothing more than applying known measures to similar electronic devices to achieve predictable results.

277. **Fourth**, a POSITA would have had at least a reasonable expectation of success of modifying Piersol with Meyers' teachings regarding confidence levels of

detection of a voice input. Indeed, Meyers itself discloses using such attributes in its own arbitration process. This demonstrates that the proposed modification to Piersol was well within a POSITA's skill.

278. Notably, the '311 Patent itself says nothing about any added benefit of using a score comprising *a confidence level of detection of the voice input* over any of the other various forms of information that it could comprise. *See, e.g.*, '311 Patent, 30:44-50. Instead, the '311 Patent merely informs a POSITA that *a confidence level of detection of the voice input* is but one design choice picked from the prior art to be used in a predictable way.

279. For at least these reasons, it is my opinion that a POSITA would have been motivated to combine Piersol with Meyers before the priority date of the '311 Patent (Oct. 3, 2016).

D. Ground 2A: Masahide Anticipates Independent Claims 1, 10, & 16

280. As explained in detail below, it is my opinion that Masahide anticipates independent claims 1, 10, and 16. As I explained before, independent claim 16 is representative of the independent claims. *Supra* ¶159-62. Thus, for the same reasons that I explain in detail below as to why Masahide anticipates claim 16, it is also my opinion that Masahide anticipates independent claims 1 and 10.

1. Masahide Anticipates Claim 16

281. As explained below, it is my opinion that Masahide discloses each element of claim 16 and thus, anticipates claim 16.

282. Claim 16 recites in full (with bracketed numerals inserted for ease of reference):

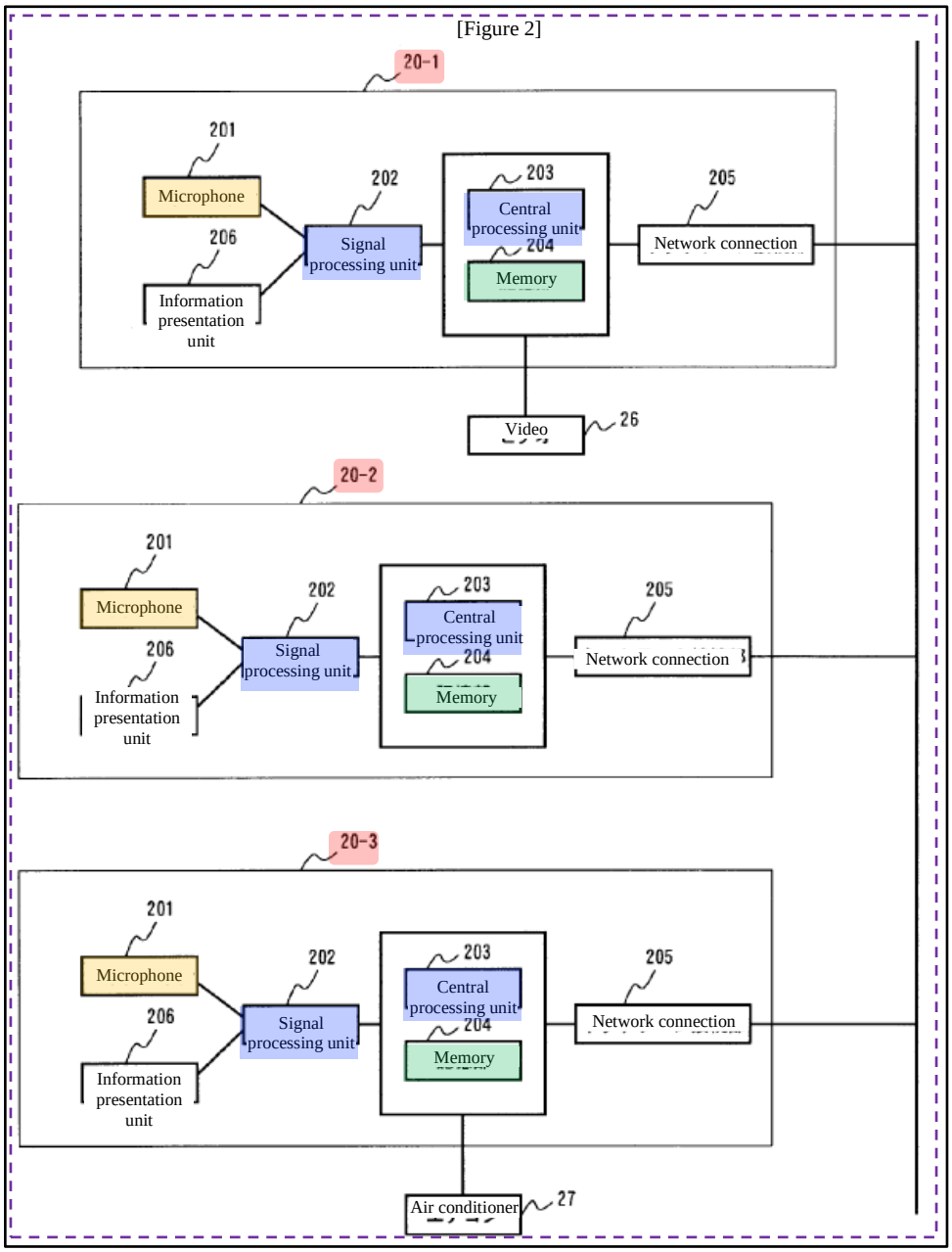
> **[16.0]** *A system comprising a plurality of* **[16.0(a)]** *electronic devices, each of the electronic devices including* **[16.0(b)]** *one or more microphones,* **[16.0(c)]** *one or more processors, and* **[16.0(d)]** *memory storing one or more programs for execution by the one or more processors, each of the electronic devices programmed to:*
>
> **[16.1]** *detect a voice input;*
>
> **[16.2]** *determine a first value associated with the detected voice input;*
>
> **[16.3]** *receive respective second values from other electronic devices of the plurality of electronic devices;*
>
> **[16.4]** *compare the first value and the respective second values; and*
>
> **[16.5]** *in accordance with a determination that the first value is higher than the respective second values, respond to the voice input.*

a. [16.0] Preamble

283. The preamble of claim 16 recites a *system comprising a plurality of electronic devices, each of the electronic devices including one or more microphones, one or more processors, and memory storing one or more programs for execution by the one or more processors, each of the electronic devices programmed to* perform the recited functions. To the extent that the preamble is limiting, it is my opinion that Masahide discloses the elements of the preamble.

284. In this regard, Masahide discloses and illustrates in FIG. 2 a voice input system that includes "voice input devices (20-1 through 20-3)" that "are connected to network 21," where "[e]ach voice input device (20-1 through 20-3) consists of a microphone 201, a signal processing unit 202, a central processing unit 203, a storage unit 204, a network connection unit 205, and an information presentation unit 206, respectively." Masahide, ¶13-14.

[Figure 2]



Id., FIG. 2 (annotations added).

285. Masahide explains that "[c]entral processing unit 203 controls the processing of the entire voice input device. Central processing unit 203 controls the state of the voice input device and sends ***instructions*** to each processing mechanism as necessary. The content to be controlled can be determined ***based on information***

from signal processing unit 202, network connection unit 205 and memory unit 204." *Id.*, ¶21; *see also id.*, ¶28 ("Figure 3 shows how central processing unit 203 processes voice based on voice signals from signal processing unit 202, information from network connection unit 205, and ***information possessed*** by memory unit 204.").

286. In this way, a POSITA would understand that Masahide's memory/storage unit 204 includes one or more programs for execution by one or more of Masahide's processing units to perform the functions disclosed in Masahide.

287. Thus, it is my opinion that Masahide discloses **[16.0]** a *system comprising a plurality of* **[16.0(a)]** *electronic devices, each of the electronic devices including* **[16.0(b)]** *one or more microphones*, **[16.0(c)]** *one or more processors, and* **[16.0(d)]** *memory storing one or more programs for execution by the one or more processors, each of the electronic devices programmed to* perform the recited functions, as discussed in further detail below.

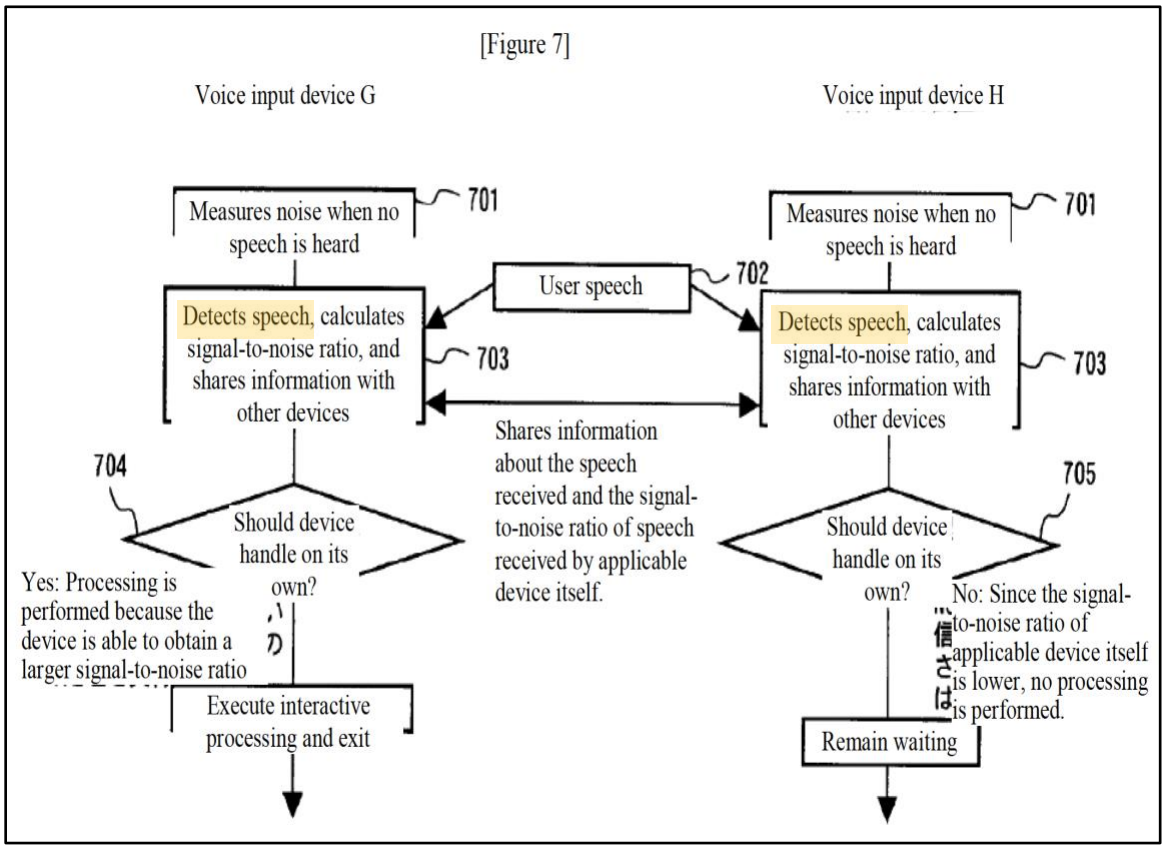
b. [16.1] Detect a Voice Input

288. Claim 16 recites that *each of the electronics devices [is] programmed to:* **[16.1]** *detect a voice input* …. As noted above, Google contends *detect[ing] a voice input* (i) refers "to a process for identifying that an audio input includes voice," (ii) "refers to a process separate from performing speech recognition (e.g., speech to

text) on the voice input," and (iii) can be satisfied by "voice activity detection" or "hotword/wakeword detection." *Supra* §VII.A.

289. Masahide discloses this element. In this regard, Masahide states that each voice input device is configured to "***detect[]*** voice from the user captured by microphone 201 at signal processing unit 202…." Masahide, ¶55; *see also, e.g.*, *id.*, ¶30 ("First, when a user speaks to voice input device A and voice input device B (step 301), ***voice is detected*** from a user captured by microphone 201 and signal is processed in signal processing unit 202 of each voice input device (step 302)."), ¶66, 73, 88, 114 ("When the user says [video] and [play], the voice input device connected to the air conditioner and the standalone voice input device ***detect the user's voice***.").

290. Masahide illustrates two voice input devices detecting speech/voice in many of its figures:



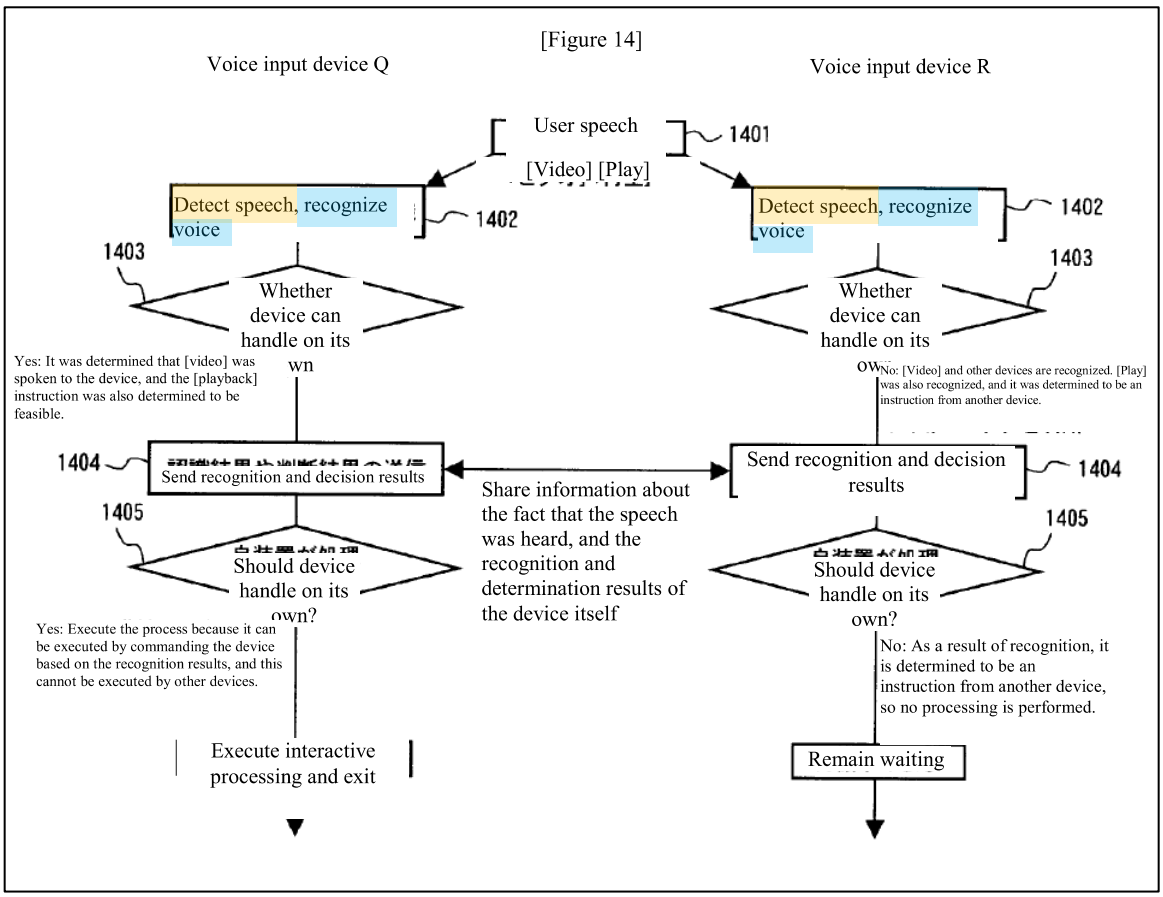
Id., FIG. 7 (annotations added); *see also id.*, FIGs. 3-6, 8-10, 12, 14.

291. Masahide repeatedly states that each voice input device's ability to "detect" voice is separate and distinct from its ability to "recognize" voice. *See, e.g.*, Masahide, ¶66 ("In the first instance, when the user says [play] as associated with voice input device K (step 901[])], each voice input device ***detects and recognizes*** the voice (step 902)."") (brackets around "play" original), ¶88 ("When the user says 'video' 'playback' as shown in the flow diagram in Figure 14 (step 1401), voice input device Q and voice input device R ***detect and recognize*** voice (step 1402).").

292. Masahide explains that (i) each voice input device recognizes voice by performing "speech recognition" (or "voice recognition") using various methods,

"such as HMM and DP matching using mixed normal distribution as models," (ii) the device's memory unit can include "the HMM and language model," and (iii) the voice input devices in the system might have a "common" "vocabulary" or different vocabularies. Masahide, ¶64; *see also, e.g., id.*, ¶87-91.

293. Masahide illustrates two voice input devices detecting speech/voice separately from performing speech recognition in many of its figures:



Id., FIG. 14 (annotations added); *see also id.*, FIGs. 9, 12.

294. While Masahide discloses "an example of the present invention mainly describes the human voice of a user," Masahide explains the "sound of machinery

and animals, may, depending on the purpose, be acceptable." Masahide, ¶12; *see also id.*, ¶70, 76. This disclosure does not change my opinion that Masahide discloses **[16.1]** *detect[ing] a voice input*.

295. In fact, Google's interpretation of **[16.1]** expressly encompasses voice activity detection (VAD), and a POSITA would understand that VAD often detected sounds other than human voice. *See, e.g.*, Ex.1016, p.2 ("We describe an autonomous system for robust detection of ***acoustic events*** in various practically relevant acoustic situations that benefits from a ***voice activity detection*** inspired pre-processing mechanism…. As a specific use case, we address ***coughing*** as an acoustic event of interest…."), p. 4 ("[A]n energy-based VAD algorithm that does not include a particular speech model could classify such events as a *voice activity* as well, such that ***its applicability*** is ***not*** only ***limited to speech signals***.") (italics alone original); Ex.1017, p.1 ("Important signal characteristics of ***elephant vocalisations*** were identified from spectrograms and a technique, based on the principles of existing ***voice activity detection*** algorithms, was developed to exploit these."), p. 3 ("[E]xisting techniques used for ***voice activity detection*** of human speech may also be ***suitable for elephant rumble detection***."); Ex.1018, p. 3 ("Advanced techniques employed in ***voice activity detection*** of human speech can also be used to improve the detection robustness and performance of ***cetacean [i.e., marine mammal]*** call detection in noisy environments.").

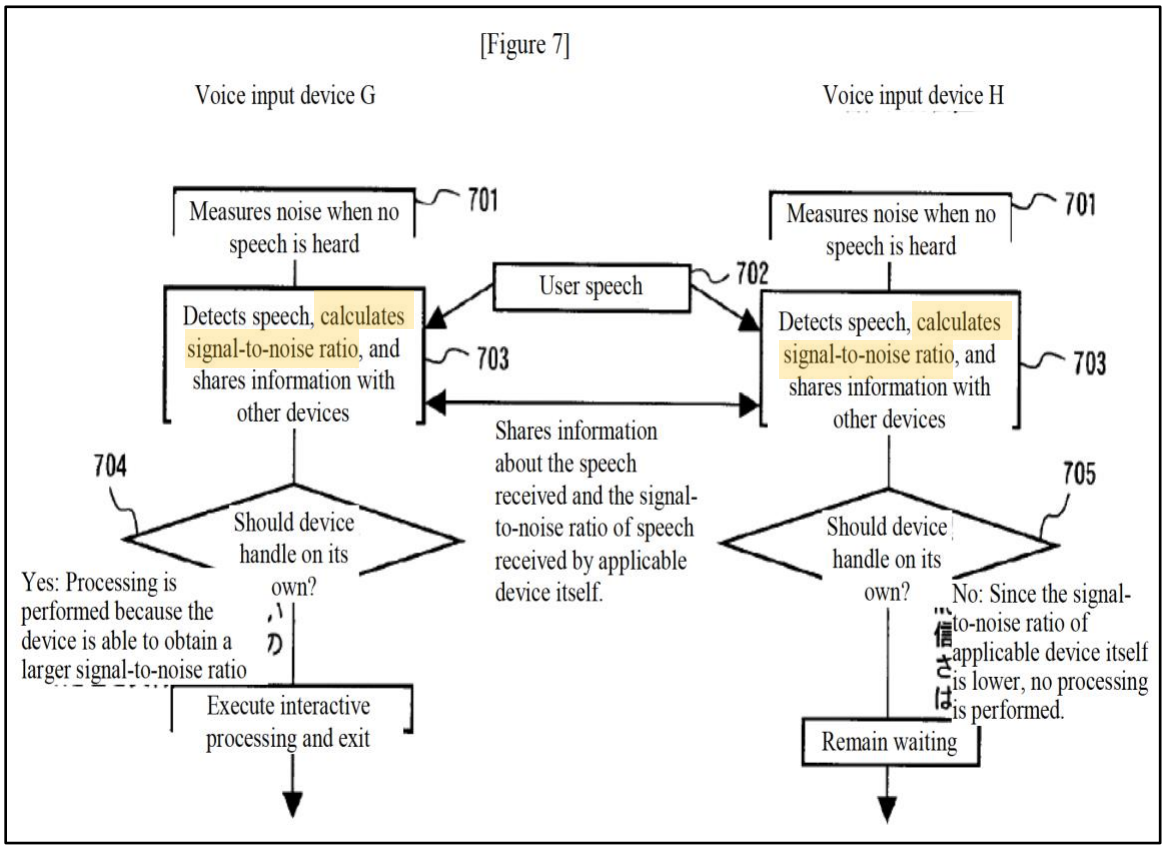
296. Thus, it is my opinion that Masahide discloses that *each of the electronics devices [is] programmed to:* **[16.1]** *detect a voice input*.

c. [16.2] Determine a First Value Associated with the Detected Voice Input

297. Claim 16 recites that *each of the electronics devices [is] programmed to:…* **[16.2]** *determine a first value associated with the detected voice input ….*

298. Masahide discloses this element. For instance, Masahide discloses that each voice input device is configured to determine (or calculate) "determinative information," such as "signal-to-noise ratio" (SNR) information, that is associated with the detected voice input. *See, e.g.*, Masahide, ¶4-6, 51-57, 116, FIG. 7, cls. 1, 3-6.

299. Masahide illustrates two voice input devices calculating SNR associated with detected voice input:



Id., FIG. 7 (annotations added).

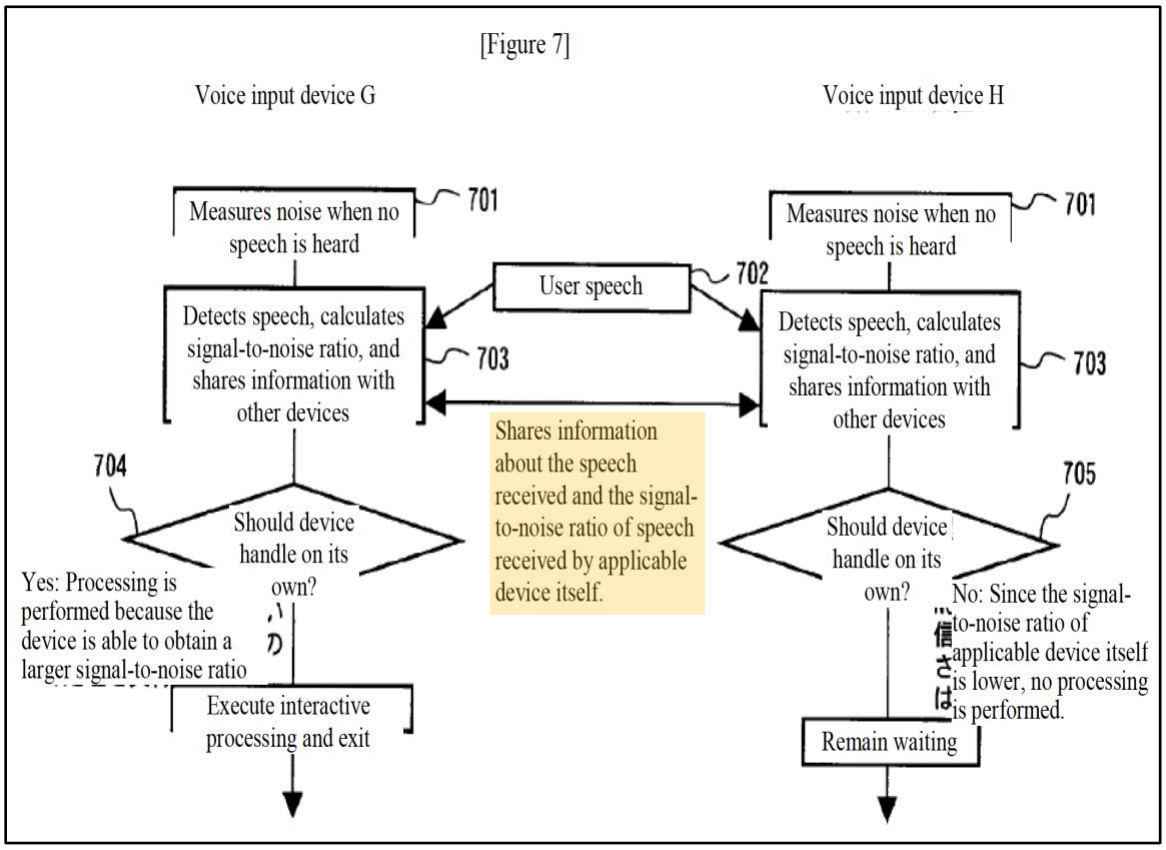
300. Thus, it is my opinion that Masahide discloses that *each of the electronics devices [is] programmed to:…* **[16.2]** *determine a first value associated with the detected voice input*.

d. [16.3] Receive Respective Second Values from Other Electronic Devices of the Plurality of Electronic Devices

301. Claim 16 recites that *each of the electronics devices [is] programmed to:…* **[16.3]** *receive respective second values from other electronic devices of the plurality of electronic devices ….*

302. Masahide discloses this element. For instance, Masahide discloses that each voice input device is configured to receive "determinative information," such as SNR information, from one or more other voice input devices. *See, e.g.*, Masahide, ¶25 ("Network connection unit 205 is a mechanism for sending and ***receiving information between voice input devices*** through network 21, and can be achieved by inter-device communication technology such as network connection over LAN and wireless technology such as Bluetooth. Network connection over LAN is used here."), ¶55 ("Next, the user speaks to voice input device G and voice input device H (step 702), and each voice input device detects voice from the user captured by microphone 201 at signal processing unit 202, calculates ***the signal-to-noise ratio*** based on the noise information when the user's voice is captured from the microphone, and ***communicates this to other voice input devices*** on the network (step 703).").

303. Masahide illustrates "voice input device G" receiving SNR information from "voice input device H":



Id., FIG. 7 (annotations added).

304. Thus, it is my opinion that Masahide discloses that *each of the electronics devices [is] programmed to:…* **[16.3]** *receive respective second values from other electronic devices of the plurality of electronic devices*.

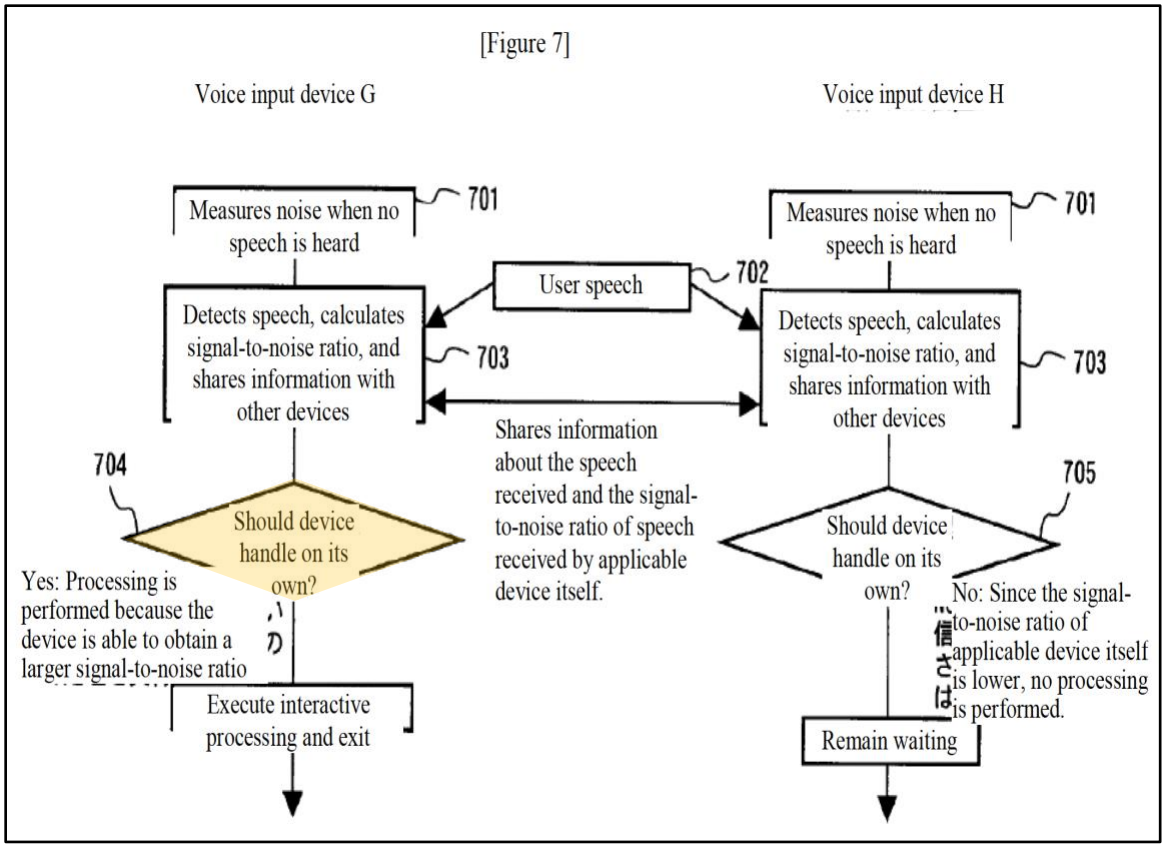
e. [16.4] Compare the First Value and the Respective Second Values

305. Claim 16 recites that *each of the electronics devices [is] programmed to:…* **[16.4]** *compare the first value and the respective second values ….*

306. Masahide discloses this element. For instance, Masahide discloses embodiments where each voice input device is configured to compare its own

"determinative information," such as SNR, with "determinative information," such as SNR, received from another voice input device. *See, e.g.*, Masahide, ¶56 ("Next, the signal-to-noise ratio information from other voice input devices that detected sound ***is compared*** with the signal-to-noise ratio information of the voice input device, and if the voice input device is the largest, the voice is processed (step 704)).").

307. Masahide illustrates "voice input device G" comparing its SNR with SNR from "voice input device H":



Id., FIG. 7 (annotations added).

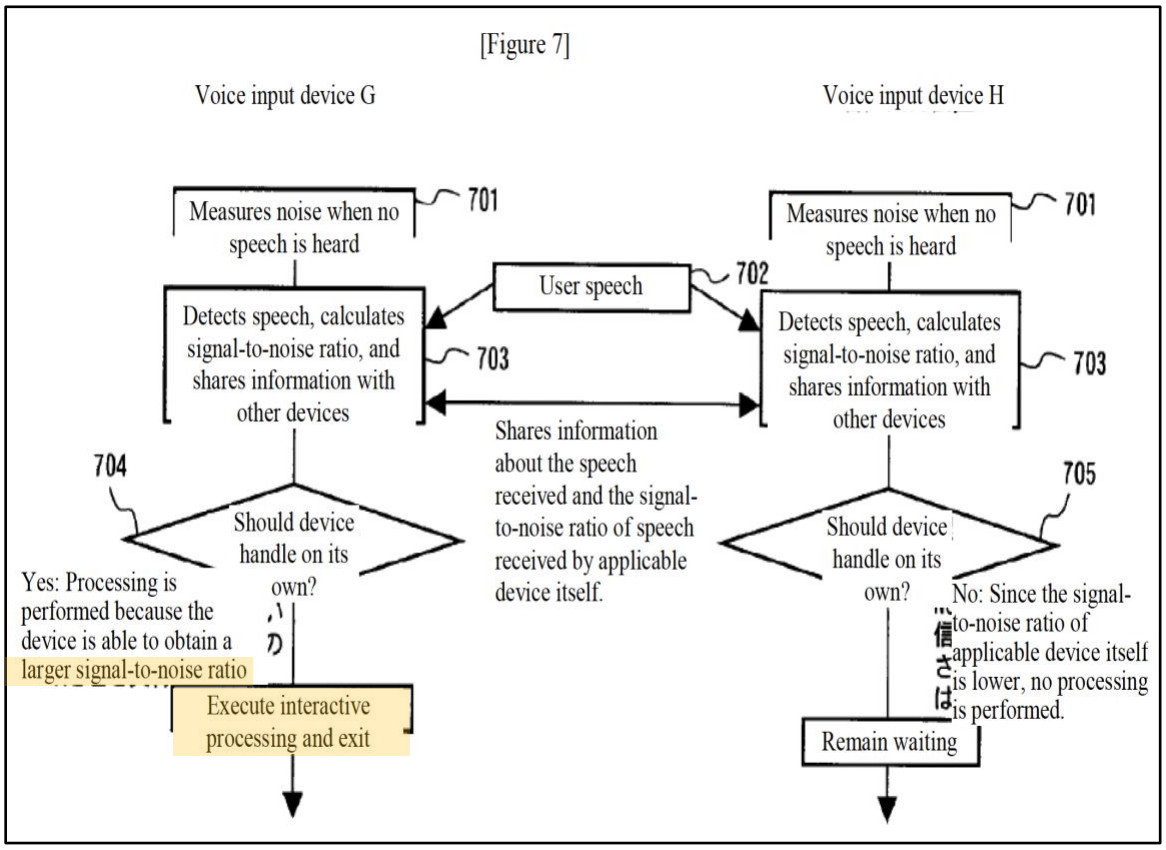
308. Thus, it is my opinion that Masahide discloses that *each of the electronics devices [is] programmed to:…* **[16.2]** *compare the first value and the respective second values*.

f. [16.5] In Accordance with a Determination that the First Value Is Higher than the Respective Second Values, Respond to the Voice Input

309. Claim 16 recites that *each of the electronics devices [is] programmed to:…* **[16.5]** *in accordance with a determination that the first value is higher than the respective second values, respond to the voice input*.

310. Masahide discloses this element. For instance, Masahide discloses that each voice input device is configured to respond to voice input in accordance with a determination that its "determinative information," such as SNR, is higher than the "determinative information," such as SNR, from another voice input device. *See, e.g.*, Masahide, ¶56 ("Next, the signal-to-noise ratio information from other voice input devices that detected sound is compared with the signal-to-noise ratio information of the voice input device, and if the voice input device is ***the largest***, the ***voice is processed*** (step 704)).").

311. Masahide illustrates "voice input device G" responding to the voice input by "execut[ing] interactive processing" because it had a "larger" SNR than "voice input device H":



Id., FIG. 7 (annotations added).

312. Masahide discloses to a POSITA that the reason for evaluating "determinative information" (e.g., SNR) is to determine which device will perform "processing and execution ***of [the] voice information***" (*id.*, ¶4). In this regard, Masahide discloses that, once a voice input device determines to handle the detected voice input on its own, the device "execute[s] interactive processing" and exits the routine. *Id.*, FIG. 3-10.

313. In connection with the first time it illustrates "execute interactive processing and exit" in FIG. 3, Masahide discloses that "central processing unit 202 of voice input device B ***processes the voice*** captured according to the function of the

voice input device, ***operates the device according to the content of the voice***, and waits again after the interaction ends (step 304)." *Id.*, ¶31.

314. FIGs. 4-10 each include a step with the same language –"execute interactive processing and exit"— as Masahide uses in step 304 of FIG. 3. A POSITA would understand that Masahide intended the same words from step 304 to have the same meaning in FIGs. 4-10.

315. Masahide provides examples of a voice input device responding to a voice input, including starting playback of a video and/or outputting a synthetic audible voice "playback started" or "starting playback." *Id.*, ¶89-90, 114-19. Masahide includes FIG. 13 identifying example ways by which a voice input device can respond to a voice input:

[Figure 13]

Recognition vocabulary	Target device	Processing details
[Video]	Video	Specify video
[Air conditioner]	Air conditioner	Specify air conditioner
[Power ON]	Shared Instructions	Turn on the target device
[Power OFF]	Shared Instructions	Turn off the target device
[Play]	Video	Play video
[Stop]	Video	Stop playing video
[Turn up temperature]	Air conditioner	Operates by increasing the temperature setting of the air conditioner
[Turn down temperature]	Air conditioner	Operation by lowering the air conditioner temperature setting

Id., FIG. 13 (annotations added); *see also* element **[2.1]**.

316. Thus, it is my opinion that Masahide discloses that *each of the electronics devices [is] programmed to:…* **[16.5]** *in accordance with a determination that the first value is higher than the respective second values, respond to the voice input*.

E. Ground 2A: Masahide Invalidates Dependent Claims 2-3, 8-9, 11, 14, 17, 19-20

317. As noted above, the Challenged Claims include (i) claims 2-3 and 8-9 that depend from independent claim 1, (ii) claims 11 and 14 that depend from independent claim 10, and (iii) claims 17 and 19-20 that depend from independent claim 16.

318. As I explained above, it is my opinion that Masahide anticipates each of independent claims 1, 10, and 16. Below, I address the additional elements of these dependent Challenged Claims and explain why Masahide invalidates them as well.

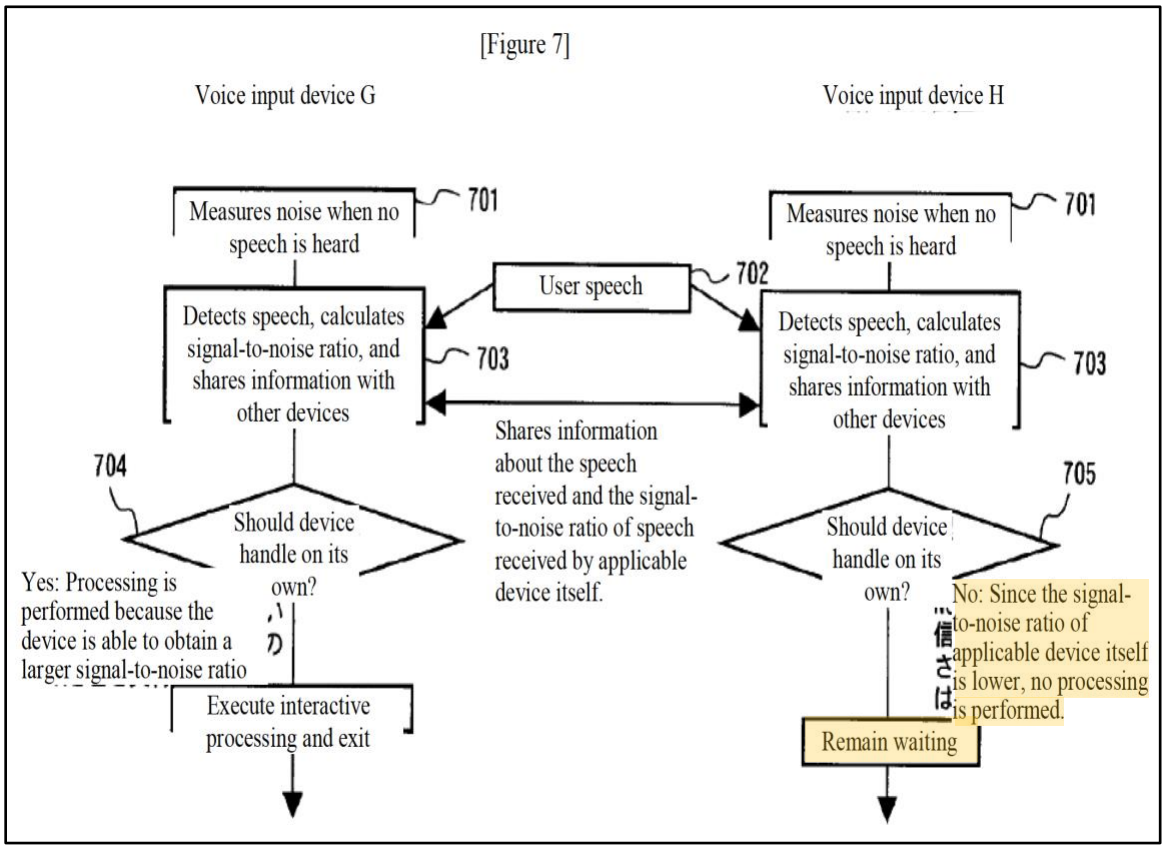
1. Dependent Claim 2

319. Dependent claim 2 recites *[t]he method of claim 1, further comprising:* **[2.1]** *in accordance with a determination that the first value is not higher than the second scores, forgoing responding to the detected input*.[7]

[7] For the purposes of this IPR, I have been asked to assume that *the second scores* recited in claim 2 refers to *the second values* recited in independent claim 1.

320. Masahide discloses this additional element. As I explained above, Masahide discloses that each voice input device is configured to determine whether it is to respond to a detected voice input if its "determinative information," such as SNR, is higher than any "determinative information," such as SNR, received from another voice input device. *See* element **[16.5]**.

321. Masahide further discloses that each voice input device is configured such that, upon determining that its "determinative information," such as SNR, is not higher than "determinative information," such as SNR, received from another electronic device, it forgoes responding to the detected voice input. *See, e.g.*, Masahide, ¶56 ("[T]he signal-to-noise ratio information from other voice input devices that detected sound is compared with the signal-to-noise ratio information of the voice input device, and if the voice input device is the largest, the voice is processed (step 704). ***Otherwise***, a decision is made that ***the voice is not to be processed*** (step 705)."); *see also, e.g.*, *id.*, ¶47, 51.



Id., FIG. 7 (annotations added).

322. Thus, it is my opinion that Masahide discloses **[2.1]**.

2. Dependent Claim 3

323. Dependent claim 3 recites *[t]he method of claim 2,* **[3.1]** *wherein forgoing responding to the detected input further comprises changing a state of operation of the first electronic device*.

324. Masahide discloses this additional element, and/or at a minimum, renders this additional element obvious in view of Masahide itself.

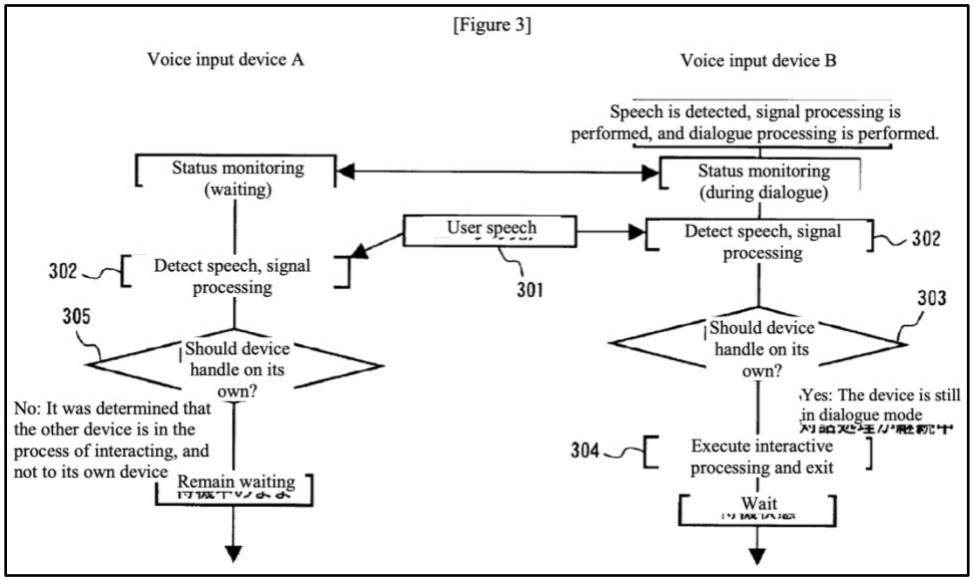
325. To start, Masahide discloses multiple arbitration approaches, each involving different types of information for making the arbitration decision. *See,*

e.g., Masahide, ¶29-34, FIG. 3 (disclosing arbitration based on processing state), ¶44-48, FIG. 5 (disclosing arbitration based on voice detection time), ¶49-51, FIG. 6 (disclosing arbitration based on volume of detected voice input), ¶52-57, FIG. 7 (disclosing arbitration based on SNR of detected voice input).

326. Masahide also discloses that multiple, different types of information can be used for making the arbitration decision. *See*, *e.g.*, Masahide, ¶104-107. In this regard, Masahide explains that, "***in addition to*** information such as detection and recognition results when voice is detected" (e.g., voice detection time, volume, SNR, etc.), "the voice input device can ***also*** exchange processing status, processing performance, recognizable vocabulary, and processing details of other voice processing systems…." *Id.*, ¶104. In other words, Masahide explains "the means for determining input for voice input devices as described conventionally ***may be combined*** with those described above" and provides an example where (i) voice detection time and (ii) (a) volume of detected voice input or (b) (1) speech recognition likelihood and (2) device rank information are used to make the arbitration decision. *Id.*

327. In this respect, in connection with FIG. 3 showing voice input devices exchanging respective status information, Masahide explains that, "if voice input device B is in the process of interacting with the user ***and other systems are not in***

an interactive state, voice input device B makes a choice to process the content spoken by the user." *Id.*, ¶31.



328. From these disclosures, a POSITA would readily understand that Masahide discloses (or at least renders obvious) a situation where multiple voice input devices are in "an interactive state" (e.g., in "dialogue mode") such that that state/mode alone does not control the arbitration decision and other information, such as a detection result like SNR, would also need to be evaluated. Otherwise, multiple voice input devices would respond to the detected voice input, which would undesirably cause user confusion and/or redundant interactive processing. *See, e.g.*, *id.*, ¶1-3, 11, 51, 57, 116. In such a situation, a POSITA would appreciate that the device with the lower SNR would lose arbitration and forgo responding to the detected voice input by "enter[ing] a waiting state" (*see id.*, ¶32), thereby changing its mode of operation from the "interactive state" (or "dialogue mode").

329. Thus, in a scenario where Masahide's voice input devices consider both processing status (e.g., whether in an interactive state with the user, which is also referred to as being in a "dialogue mode") and a detection result like SNR, a voice input device that loses arbitration is configured such that *forgoing responding to the detected input further comprises changing a state of operation of the* [voice input device].

330. Thus, it is my opinion that Masahide discloses **[3.1]** or at least renders it obvious in view of Masahide alone.

3. Dependent Claim 8

331. Dependent claim 8 recites *[t]he method of claim 1,* **[8.1]** *wherein responding to the detected input comprises outputting an audible and/or a visual response to the detected voice input*.

332. Masahide discloses this additional element. For instance, Masahide discloses that voice input devices are configured such that responding to a detected voice input can involve providing a visual and/or audible output. *See, e.g.*, Masahide, ¶35 ("[T]he dialogue here is not limited to ***one-on-one voice interactions between humans and systems***, but rather may include the return of unilateral speech from humans to systems, ***visual responses*** from the system side, or ***responses from the system to any person***, and the same applies to the dialogue used in the following explanations."), ¶113 ("The information presentation unit of ***each voice input device***

in this example is ***equipped with a speaker***, and it is possible to ***return the voice of any text synthesized*** by the central processing unit and the signal-processing unit to the user."), ¶119 ("In this way, a response instruction sent from a video voice input device to a user can be interpreted, and a standalone voice input device can convey to the user a message of ***'started playing' with a synthetic voice***."); *see also* element **[16.5]**; Masahide, ¶17.

333. Thus, it is my opinion that Masahide discloses **[8.1]**.

4. Dependent Claim 9

334. Dependent claim 9 recites *[t]he method of claim 1,* **[9.1]** *wherein the first electronic device further comprises a speaker, and* **[9.2]** *responding to the detected input further comprises outputting an audible response*.

335. Masahide discloses these additional elements. As just discussed, Masahide discloses that voice input devices are configured such that responding to a detected voice input can involve providing an audible output. *See* element **[8.1]**

336. Masahide discloses that voice input devices can comprise a "speaker" to enable outputting such responses. *See, e.g.*, Masahide, ¶113 ("The information presentation unit of ***each voice input device*** in this example is ***equipped with a speaker***, and it is possible to ***return the voice of any text synthesized*** by the central processing unit and the signal-processing unit to the user.").

337. Thus, it is my opinion that Masahide discloses **[9.1]** and **[9.2]**

5. Dependent Claim 11

338. Dependent claim 11 recites *[t]he first electronic device of claim 10, the one or more programs further comprising instructions for:* **[11.1]** *recognizing a command in the voice input; and* **[11.2]** *in accordance with a determination that a type of the command is related to the first electronic device, responding to the detected voice input.*

339. Masahide discloses these additional elements.

340. **First**, Masahide discloses that each voice input device is configured to recognize a command in the voice input.

341. For instance, Masahide discloses that each voice input device is configured to perform "speech recognition" (or "voice recognition") to recognize voice commands/instructions:

> The information from the signal-processing unit is processed by a ***voice recognition*** mechanism and the result is passed to the central processing unit. ***Speech recognition*** performed at this time may be handled by the central processing unit.
>
> Also, the method used for ***speech recognition*** can be a method generally utilized, such as ***HMM and DP matching*** using mixed normal distribution as models, and the HMM and language model used at this time can be in the memory unit. The vocabulary of voice recognition may be different for each voice input device or common to all such devices. Further, the vocabulary can be matched with control instructions to enable ***voice commands***.

Masahide, ¶63-64; *id.*, ¶82 ("[I]f the user says [video] [power ON] (step 1201), voice input device O and voice input device P ***recognize the device name and device***

instructions using the matching method used for the aforementioned voice recognition (step 1202), and ***determine whether they are instructions to their system*** or can be processed (step 1203)."); *see also id.*, ¶8, 43, 65-69, 79-91, 114-18, FIGs. 9, 12, 13, 14.

342. Thus, it is my opinion that Masahide discloses **[11.1]**.

343. **Second**, Masahide discloses that each voice input device is configured to respond to the voice input upon determining that a type of the command is related to the voice input device. For instance, Masahide discloses an example where multiple voice input devices recognize the same command(s) in a user's utterance and only one device responds when it determines the command is an instruction to its system:

> In this case, if the user says [video] [power ON] (step 1201), voice input device O and voice input device P ***recognize the device name and device instructions*** using the matching method used for the aforementioned voice recognition (step 1202), and ***determine whether they are instructions to their system*** or can be processed (step 1203).
>
> The results can be communicated to other voice input devices and controllable devices on the network (step 1204), and ***these results and results from other voice input devices*** can be ***used to determine if the voice input device should process these*** (step 1205) and then processing corresponding to the control instructions can be performed.

Masahide, ¶82-83, FIG. 12.

344. Masahide provides another example in connection with "Figure 13 [that] represents the concept of linking [a] recognition word to the target device and

processing content." *Id.*, ¶87. In particular, in connection with FIG. 14, Masahide discloses:

> When the user says "video" "playback" as shown in the flow diagram in Figure 14 (step 1401), voice input device Q and voice input device R detect and recognize voice (step 1402).
>
> Furthermore, using the concept shown in Fig. 13, the content of speech is determined (step 1403), the results are transmitted to other voice input devices on the network (step 1404), and based on these results and the results sent by other voice input devices, it is determined whether the speech sent by another voice input device should have been processed (step 1405), and the processing corresponding to the control instructions is performed.
>
> In the case of [video] and [playback] as described above, ***it can be determined that the voice of both voice input devices was an instruction for playback of the video*** by linking the recognition word to the target device and the processing content as shown in Figure 13. Furthermore, the information transmitted over the network can be used to uniquely interpret the voice, and the voice input device can perform processing corresponding to the recognition results.

Id., ¶88-90, FIGs. 13-14.

345. Masahide discloses that its arbitration process of determining whether to respond can look at recognized voice commands and/or other "determinative information" such as SNR. *See, e.g.*, cls. 1, 8, ¶104-107.

346. Thus, it is my opinion that Masahide discloses **[11.2]**.

6. Dependent Claim 14

347. Dependent claim 14 recites *[t]he first electronic device of claim 10, wherein:* **[14.1]** *the plurality of electronic devices is communicatively coupled*

through a local network; and **[14.2]** *the receiving is performed through the local network.*

348. Masahide discloses these additional elements.

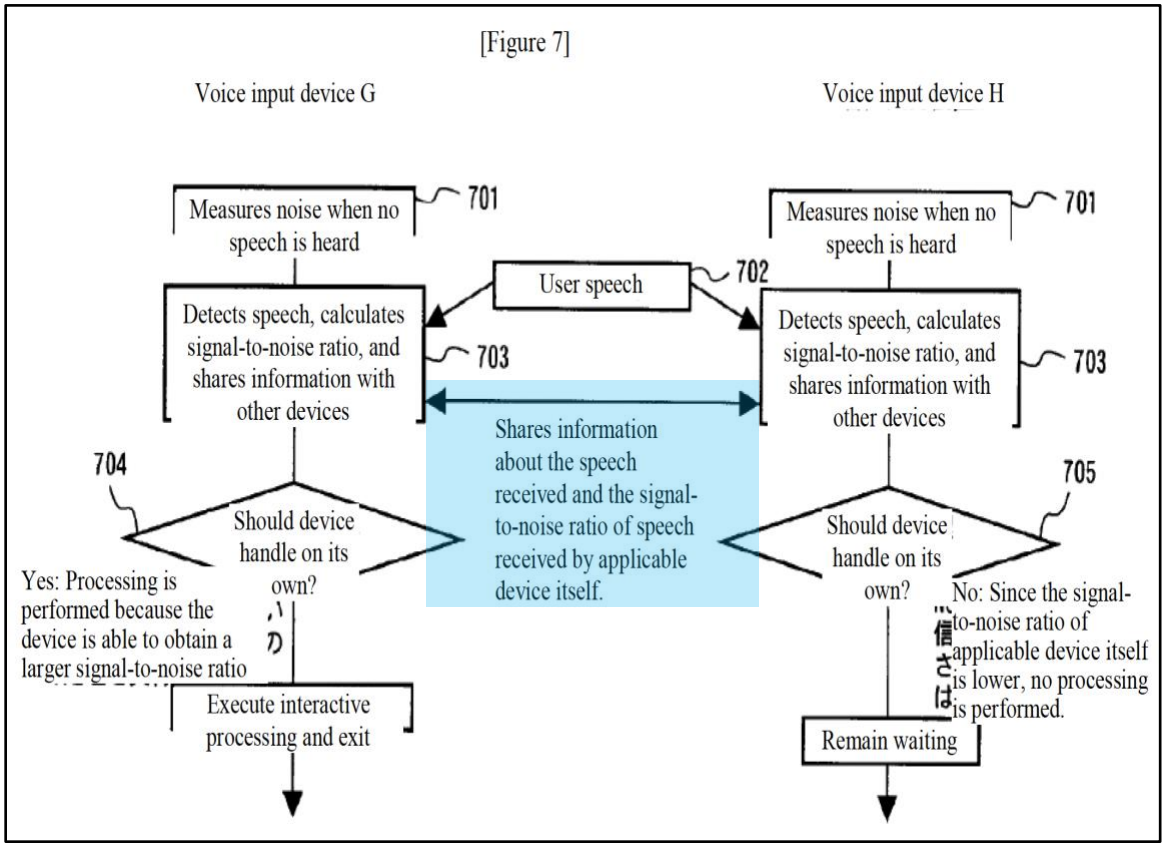
349. **First**, Masahide discloses that the voice input devices are communicatively coupled through a local area network (LAN), such as using Bluetooth. *See, e.g.*, Masahide ¶8 ("In the voice input system of the present invention, a device 103 having multiple single voice input devices 101 and 102 ***connected to network 104***…."), ¶14 ("Each voice input device (20-1 through 20-3) consists of a microphone 201, a signal processing unit 202, a central processing unit 203, a storage unit 204, ***a network connection unit 205***, and an information presentation unit 206, respectively."), ¶25 ("***Network connection unit 205*** is a mechanism for sending and receiving information between voice input devices through ***network*** 21, and can be achieved ***by inter-device communication technology such as network connection over LAN and wireless technology such as Bluetooth.*** Network connection over ***LAN is used here***.").

350. Thus, it is my opinion that Masahide discloses **[14.1]**.

351. **Second**, Masahide discloses that each voice input device is configured to receive "determinative information," such as SNR, from another voice input device through the local network. *See, e.g.*, *id.*, ¶25 ("Network connection unit 205 is a mechanism for ***sending and receiving information between voice input devices***

through ***network*** 21, and can be achieved by inter-device communication technology such as network connection over LAN and wireless technology such as Bluetooth. Network connection over LAN is used here."), ¶55 ("[T]he user speaks to voice input device G and voice input device H (step 702), and each voice input device detects voice from the user captured by microphone 201 at signal processing unit 202, calculates the ***signal-to-noise ratio*** based on the noise information when the user's voice is captured from the microphone, and ***communicates this to other voice input devices on the network*** (step 703).").

352. A voice input device receiving "determinative information" (e.g., SNR) from another voice input device through the LAN is shown below:



Id., FIG. 7 (annotations added); *see also, e.g.*, *id.*, FIGs. 1, 11, 15, 16, 17, 18, 19.

353. Thus, it is my opinion that Masahide discloses **[14.2].**

7. Dependent Claim 17

354. Dependent claim 17 recites *[t]he system of claim 16, wherein the first value comprises a first quality score of the voice input.*

355. Masahide discloses this additional element. As explained before, Masahide discloses that its "determinative information" associated with a detected voice input can include SNR information. *See* element **[16.2]**.

356. Claim 19 expressly recites that the *first quality score comprises a signal-to-noise rating of detection of the voice input* and the '311 Patent explains that "signal-to-noise ratio" is one example of *a signal-to-noise rating*. '311 Patent, 30:47-50.

357. Thus, it is my opinion that Masahide discloses **[17.1].**

8. Dependent Claim 19

358. Dependent claim 19 recites *[t]he system of claim 17,* **[19.1]** *wherein the first quality score comprises a signal-to-noise rating of detection of the voice input*.

359. For the reasons I explained for dependent claim 17, Masahide discloses this additional element. *Supra* ¶354-57.

9. Dependent Claim 20

360. Dependent claim 20 recites *[t]he system of claim 16,* **[20.1]** *wherein each of the respective second values is indicative of a quality of the voice input detected by a respective one of the other electronic devices.*

361. For the reasons I explained for dependent claim 17, Masahide discloses this that its "determinative information" *is indicative of a quality of the voice input detected* by a given voice input device at least when it includes SNR information. *Supra* ¶354-57.

362. And for the reasons I explained for **[16.3]**, Masahide discloses that each voice input device is configured to receive "determinative information" containing SNR information, which is *indicative of a quality of the voice input detected,* from other voice input devices. *Supra* ¶301-304.

363. Thus, it is my opinion that Masahide discloses **[20.1]**.

F. Ground 2B: Masahide + Jang Render Obvious Independent Claims 1, 10, 16 and Dependent Claims 2-3, 8-9, 11, 14, 17, 19-20

364. **Ground 2B** – Masahde combined with Jang – is presented to the extent that Google argues that Masahide does not adequately disclose **[1.1]**, **[10.1]**, **[16.1]** *detect[ing] a voice input*.

365. As explained below, Jang teaches *detect[ing] a voice input*, and a POSITA would have found it obvious to incorporate such teachings with Masahide's teachings. **Ground 2B** is otherwise the same as **Ground 2A**.

1. Jang Discloses Detecting a Voice Input

366. Jang discloses an electronic device **[1.1]**, **[10.1]**, **[16.1]** *detect[ing] a voice input*. In particular, Jang discloses a system of "electronic devices" (*see, e.g.*, Jang, ¶30, 53, FIGs. 1-2), each of which is configured to "detect" voice input:

> The voice recognition unit 182 can carry out voice recognition upon voice signals input through the microphone 122 of the electronic device 100 or the remote control 10 and/or the mobile terminal shown in FIG. 1; the voice recognition unit 182 can then obtain at least one recognition candidate corresponding to the recognized voice. For example, the voice recognition unit 182 can recognize the input voice signals by ***detecting voice activity from the input voice signals***, carrying out sound analysis thereof, and recognizing the analysis result as a recognition unit.

Id., ¶110.

367. In this way, a POSITA would understand Jang discloses that each electronic device's "voice recognition unit 182" performs voice activity detection (VAD) prior to performing speech/voice recognition. *See also, e.g., id.*, ¶98-99 (describing Jang's "recognition dictionary"), ¶109-110.

368. Thus, it is my opinion that Jang discloses **[1.1]**, **[10.1]**, **[16.1]**.

2. A POSITA Would Have Been Motivated to Combine Masahide with Jang

369. In my opinion, a POSITA would have been motivated to combine Masahide with Jang even before the earliest possible priority date of the '311 Patent (Oct. 9, 2014) such that Masahide's voice input device would detect a voice input utilizing Jang's "voice recognition unit 182… detecting voice activity from the input

voice signals" (Jang, ¶110), which in turn would initiate Masahide's arbitration process to "solve[] the question of what to do with each voice input device when the voice of the user is input to a plurality of networked voice input devices." Masahide, ¶11. In this way, a POSITA would achieve **[1.1]**, **[10.1]**, **[16.1]** (again, assuming Masahide does not already disclose these elements). It is my opinion that a POSITA would have been motivated to make this combination for a variety of reasons.

370. **First**, Masahide and Jang are in the same field of endeavor, relate to voice-enabled devices, 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.*, Masahide, ¶11 ("The present invention solves the question of what to do with each voice input device when the voice of the user is input to a ***plurality*** of networked voice input devices."), Summary; 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.").

371. **Second**, Masahide and Jang have interrelated teachings, both disclosing similar arbitration processes to solve the duplicate response problem. *See, e.g.*, Masahide, FIGs. 6-7; Jang, FIGs. 11-12.

372. Indeed, Masahide and Jang both disclose arbitration techniques where electronic devices are configured to determine, receive, and compare values associated with a detected voice input (e.g., magnitude (gain value) or volume of voice signal) and select the electronic device with the highest value to respond to the user. *See, e.g.*, Masahide, ¶49-51; Jang, ¶122, 127-136, FIG. 11. A POSITA studying Masahide would have been motivated to seek out other references that disclose similar arbitration techniques, like Jang, to obtain additional details regarding such techniques.

373. **Third**, Masahide already discloses **[1.1]**, **[10.1]**, **[16.1]** *detect[ing] a voice input*. *Supra* ¶288-96. But Masahide does not provide implementation details regarding this functionality. A POSITA studying Masahide would have been motivated to seek out other references that disclose implementation details for **[1.1]**, **[10.1]**, **[16.1]**, like Jang, to achieve a fuller understanding, especially when the POSITA was implementing a system in accordance with Masahide's teachings.

374. **Fourth**, Masahide and Jang both disclose techniques where arbitration is initiated based on detecting a voice input, and Jang's methods of performing this function (e.g., via VAD) were well known ways to detect a voice input before the earliest possible priority date of the '311 Patent (Oct. 9, 2014). *Supra* ¶366-67. For instance, other prior art such as Basye filed on December 11, 2012 discloses a "speech detection module 108" configured to "determine whether [an] audio input

includes speech" utilizing various techniques such as "voice activity detection (VAD) techniques" (Basye, ¶21) and a "speech processing module 110" configured to "detect a keyword in the speech" (*id.*, ¶21), which is described as being separate and distinct from "an automatic speech recognition engine" responsible for recognizing speech responsive to "speech processing module 110" detecting a keyword (*id.*, ¶54).

375. **Fifth**, a POSITA would have appreciated that applying Jang's disclosures of **[1.1]**, **[10.1]**, **[16.1]** to Masahide would have been nothing more than applying known methods to similar electronic devices to achieve predictable results.

376. As explained before, I see nothing in the '311 specification that would suggest to a POSITA that *detect[ing] a voice input* under Google's narrow interpretation was anything more than a design choice among other choices already in the prior art. In fact, I see no discussion in the '311 specification of Google's narrow interpretation, much less any discussion in the '311 specification of *detect[ing] a voice input* under Google's narrow interpretation conferring some significant advantage to the '311 claims.

377. In this way, a POSITA would have had at least a reasonable expectation of success, especially because Jang itself already discloses using such techniques for detecting a voice input in performing arbitration in a very similar way as Masahide.

378. For at least these reasons, it is my opinion that a POSITA would have been motivated to combine Masahide with Jang before the earliest possible priority date of the ’311 Patent (Oct. 9, 2014).

XII. LACK OF SECONDARY CONSIDERATIONS

379. I am unaware of any secondary considerations of non-obviousness relating to the Challenged Claims of the ’311 Patent.

380. 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 priority date of the ’311 Patent. This is because the case for obviousness in view of primary references Piersol and Masahide is so clear and compelling.

XIII. CONCLUSION

381. 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 4, 2023 in Lexington, Kentucky

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 Engineerin*g: (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，郭璁，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 Clemins, 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. Clemins 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 Clemins 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, Thessolonica, Greece, September, 2003.

130. Patrick Clemins 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 Clemins 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-Artictulatory 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. Clemins, “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. Ricke, “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 Mugganthaler (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: Koninklikjke 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