**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

| | | |
|---|---|---|
| In re Patent of: | Petit *et al*. | |
| U.S. Patent No.: | 8,326,611 | Attorney Docket No.: 50095-0094IP1 |
| Issue Date: | December 4, 2012 | |
| Appl. Serial No.: | 12/606,140 | |
| Filing Date: | October 26, 2009 | |
| Title: | ACOUSTIC VOICE ACTIVITY DETECTION (AVAD) FOR ELECTRONIC SYSTEMS | |

# DECLARATION OF DR. THOMAS W. KENNY

## Declaration

I hereby declare that all statements made of my own knowledge are true and that all statements made on information and belief are believed to be true. I further declare that these statements 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 the Title 18 of the United States Code.

Dated: June 3, 2022 By: _____

Thomas W. Kenny, Ph.D.

## TABLE OF CONTENTS


I. QUALIFICATIONS AND BACKGROUND INFORMATION ..... 1

II. OVERVIEW OF THE CONCLUSION FORMED ..... 9

III. LEVEL OF ORDINARY SKILL IN THE ART ..... 10

IV. LEGAL STANDARDS ..... 11

- A. Terminology ..... 11
- B. Legal Standards for Anticipation ..... 11
- C. Legal Standards for Obviousness ..... 12

V. THE ’611 PATENT ..... 17

- A. Overview of the ’611 Patent ..... 17
- B. Prosecution History of the ’611 Patent ..... 21

VI. SUMMARY OF THE PRIOR ART ..... 22

- A. Overview of Avendano ..... 22
- B. Overview of Visser ..... 29
- C. Overview of Bisgaard ..... 34
- D. Overview of Hou ..... 37
- E. Overview of Byrne ..... 40
- F. Overview of Burnett ..... 43
- G. Overview of Berglund ..... 45

VII. GROUND 1: CLAIMS 1-7 AND 25-28 ARE RENDERED OBVIOUS BY AVENDANO IN VIEW OF VISSER ..... 46

- A. Combination of Avendano and Visser ..... 46
- B. Claim 1 ..... 55
- C. Claim 2 ..... 74
- D. Claim 3 ..... 75
- E. Claim 4 ..... 77
- F. Claim 5 ..... 78
- G. Claim 6 ..... 79
- H. Claim 7 ..... 80
- I. Claim 25 ..... 82
- J. Claim 26 ..... 83
- K. Claim 27 ..... 84
- L. Claim 28 ..... 85

VIII. GROUND 2: CLAIMS 8-16, 23, AND 24 ARE RENDERED OBVIOUS BY AVENDANO IN VIEW OF VISSER AND BISGAARD ..... 86

A. Combination of Avendano, Visser, and Bisgaard .....86
B. Claim 8 .....90
C. Claim 9 .....92
D. Claim 10 .....93
E. Claim 11 .....94
F. Claim 12 .....96
G. Claim 13 .....96
H. Claim 14 .....97
I. Claim 15 .....99
J. Claim 16 .....100
K. Claim 23 .....103
L. Claim 24 .....107

IX. GROUND 3: CLAIMS 17-19 ARE RENDERED OBVIOUS BY AVENDANO IN VIEW OF VISSER, BISGAARD, AND HOU .....109

A. Combination of Avendano, Visser, Bisgaard, and Hou .....109
B. Claim 17 .....115
C. Claim 18 .....118
D. Claim 19 .....120

X. GROUND 4: CLAIMS 20-22 ARE RENDERED OBVIOUS BY AVENDANO IN VIEW OF VISSER, BISGAARD, HOU, AND FREQUENCY ART (BYRNE, BURNETT, AND/OR BERGLUND) .....122

A. Combination of Avendano, Visser, Bisgaard, Hou, and Frequency Art (Byrne, Burnett, and/or Berglund) .....122
B. Claim 20 .....126
C. Claim 21 .....129
D. Claim 22 .....137

XI. CONCLUSION .....145

## I. QUALIFICATIONS AND BACKGROUND INFORMATION

1. My education and experience are described more fully in the attached curriculum vitae (APPLE-1004). For ease of reference, I have highlighted certain information below.

2. My academic and professional background is in Physics, Mechanical Engineering, Sensing, and Robotics, with a research specialization focused on microfabricated physical sensors, and I have been working in those fields since the completion of my Ph.D. more than 32 years ago. The details of my background and education and a listing of all publications I have authored in the past 37 years are provided in my curriculum vitae, APPLE-1004. Below I provide a short summary of my education and experience which I believe to be most pertinent to the opinions that I express here.

3. I received a B.S. in Physics from University of Minnesota, Minneapolis in 1983, and a Ph.D. in Physics from University of California at Berkeley in 1989. I was educated as a Physicist specializing in sensors and measurement. My Physics Ph.D. thesis involved measurements of the heat capacity of monolayers of atoms on surfaces, and relied on precision measurements of temperature and power using time-varying electrical signals, and also on the design and construction of miniature sensor components and associated electrical circuits for conditioning and conversion to digital format.

4. After completion of my Ph.D. in Physics at U.C. Berkeley in 1989, I joined the Jet Propulsion Laboratory (JPL) in Pasadena, CA, as a staff scientist, and began working on miniature sensors and instruments for small spacecraft. This work involved the use of silicon microfabrication technologies for miniaturization of the sensors, and served as my introduction to the field of micro-electromechanical systems (MEMS), or the study of very small mechanical sensors powered by electricity and used for detection of physical and chemical signals.

5. While at JPL, we developed accelerometers, uncooled infrared sensors, magnetometers, seismometers, force and displacement sensors, soil chemistry sensors, miniature structures for trapping interstellar dust, and many other miniature devices. Some of these projects led to devices that were launched with spacecraft headed for Mars and for other interplanetary missions. Much of this work involved the use of physical sensors for detection of small forces and displacements using micromechanical sensors.

6. I am presently the Richard Weiland Professor at the Department of Mechanical Engineering at Stanford University, where I have taught for the past 28 years. I am also currently the Senior Associate Dean of Engineering for Student Affairs at Stanford.

7. In my role at Stanford, I supervise Ph.D. students carrying out research on design, fabrication and characterization of micromechanical devices and sensors.

In this research, we have developed devices for measurement of inertial forces, pressure, light, temperature, and acoustic signals, as well as techniques for capturing and processing of signals and noise from sensors, with a detailed listing of the resulting publications and patents included in my CV.

8. I have also been active in teaching courses at Stanford, including ME210 Introduction to Mechatronics, and ME220 Introduction to Sensors. In ME220, I provide a survey of mechanisms used to transform physical phenomena into electric signals, followed by a discussion of many different categories of sensors such as pressure sensors, accelerometers, thermometers, microphones, gyroscopes, as well as sensors for chemical parameters, light, and electromagnetic phenomena. I have taught this course more than 25 times over my career at Stanford, and am currently offering the Spring 2022 version. In ME210, we provide a broad introduction to the concepts and components needed for design of autonomous mobile robots, which includes capture and processing of signals from sensors.

9. I have advised 73 Ph.D. students that have completed Ph.D. degrees and many more M.S. and B.S. students in Engineering during my time at Stanford. The research carried out with these Ph.D. students routinely includes measurement of signals which contain information and noise from the background, and require development of methods for identifying the information and noise components of

the signal and then implementing filters, subtractions, correlations, etc., as needed to produce clear signals as needed for the analysis.

10. I have published over 250 technical papers in refereed journals and conferences in the field of sensors, MEMS, and measurements. I have further presented numerous conference abstracts, posters, and talks in my field. I am a named inventor on 50 patents in my areas of work.

11. I served as General Chair or Technical Program Committee Chair for the International Conference on Solid State Sensors and Actuators ("Transducers") in 2003 and 2015, for the Hilton Head Workshop on Solid State Sensors and Actuators in 2002 and 2004, for the IEEE Sensors Conference in 2009 and 2010, and on technical program committees for several other conferences. These conferences included technical sessions on capture of signals in the presence of environmental noise sources, and processing to remove the portion of the signal that comes from the environmental sources. I currently serve as President of the Transducers Research Foundation, a private foundation that oversees technical conferences in the area of micromechanical sensors, actuators and systems, energy harvesting, microtechnologies for medicine and biology, and raises funds to support growth of the community, primarily by providing travel assistance for students to attend technical meetings.

12. I received the Office of the Secretary of Defense Award for Exceptional Public Service in 2010, the IEEE Sensors Council Technical Achievement Award in 2011, and was named Fellow of the ASME in 2014. I was awarded the 2018 IEEE Daniel Noble Award for Emerging Technologies. In 2022, I was elected to the National Academy of Engineering.

13. Among the publications listed in my CV, there are several which are specifically relevant to the subject matter of this declaration. For example, in the paper titled "Measurement of the Noise of a Geophone in the Presence of Large Background Signals" (*Rev. Sci. Instrum*. 69, 2767-2772 (1998)), we present a method for using a pair of sensitive vibration sensors ("geophones") operated at the same time to enable a determination of the background vibration signals detected by both sensors and the internal fundamental noise in each sensor. Since both sensors are mounted on the same surface and observing the same seismic background signals, the portion of the signal coming from the background is highly correlated, whereas the portion of the signal associated with the fundamental noise in each sensor is uncorrelated. We use measurements of the correlation between the signals from the two sensors to distinguish the seismic background signal from the sensor noise signal. To extract the sensor noise signal, we utilize the correlation coefficients to subtract out the correlated portion due to true ground motion, allowing clearer identification of the inherent noise of each sensor, even in

cases where the ground motion signals are much larger than the noise of the sensors. A similar technique is used in "A High-Precision, Wide-Bandwidth Micromachined Tunneling Accelerometer" (*JMEMS*, 10,425 (2001)), for example.

14. In the paper titled "Ultraminiature Encapsulated Accelerometers as a Fully-Implantable Sensors for Implantable Hearing Aids", (*Biomedical MicroDevices* 9, 939 (2007)), we designed and developed a novel ultraminiature accelerometer which could be mounted on the eardrum, allowing electrical measurement of acoustic signals using the mechanical elements of the ear, and enabling signal delivery to the cochlea that bypassed the mechanical elements in the middle ear. In general, the publications presented in my CV describe careful measurements of small signals from micromechanical devices, followed by amplification, filtering and processing to suppress or remove contributions from uninteresting background sources.

15. I have previously served as an expert on a patent infringement case involving the mounting and use of pressure sensors on guidewire catheters for cardiovascular procedures that included a number of sensing aspects, such as recording static and dynamic pressure signals, and compensating for electrical and mechanical errors. I have also previously served as an expert on a patent infringement case involving the design and use of miniature inertial sensors. That case involved the design and operations of micromechanical sensors, and

particularly the use of inertial sensors for detection of states of movement and rest. I have also served as an expert in a patent infringement case involving the use of sensors on athletic shoes for determining athletic performance. More recently, I served as an expert in a patent infringement case involving optical proximity sensors in smartphones and in a case involving technologies related to physiological sensors for measurement of parameters such as heart rate. My CV, APPLE-1004, includes a full listing of all cases in which I have testified at deposition or trial in the preceding four years.

16. I have been retained on behalf of Apple Inc. to offer technical opinions relating to U.S. Patent No. 8,326,611 ("the '611 patent") and prior art references relating to its subject matter. I have reviewed the '611 patent, relevant excerpts of the prosecution history of the '611 patent. I have also reviewed the following prior art references:

| **Prior Art Reference** |
| --- |
| U.S. Patent No. 8,194,880 B2 ("Avendano") |
| U.S. Patent No. 7,464,029 B2 ("Visser") |
| U.S. Patent No. 7,155,019 B2 ("Hou") |
| Byrne, D, et al, "An international comparison of long-term average speech spectra," 1994 Oct; *J. Acoust. Soc. Am.*; 96(4): 2108-2120 ("Byrne"). |
| U.S. Publication No. US 2011/0103626 A1 ("Bisgaard") |
| U.S. Provisional App. No. 60/816,244 ("the Bisgaard Provisional") |

| |
|---|
| U.S. Publication No. US 2002/0198705 A1 ("Burnett")<br><br>Berglund, B, *et al*, "Source and effects of low-frequency noise," 1996 May; *J. Acoust. Soc. Am*; 99(5): 2985-3002 ("Berglund"). |

17. Counsel has informed me that I should consider these materials through the lens of one of ordinary skill in the art related to the '611 patent at the time of the earliest possible priority date of the '611 patent, and I have done so during my review of these materials. The application leading to the '611 patent was filed on October 26, 2009 and claims the benefit of priority to a provisional application filed October 24, 2008 ("Critical Date"). Further, application leading to the '611 patent is a continuation-in-part of two applications filed May 25, 2007 and June 13, 2008, respectively. Counsel has informed me that the Critical Date represents the earliest possible priority date to which the challenged claims of '611 patent are entitled, and I have therefore used that Critical Date in my analysis below.

18. I have no financial interest in the party or in the outcome of this proceeding. I am being compensated for my work as an expert on an hourly basis. My compensation is not dependent on the outcome of these proceedings or the content of my opinions.

19. In writing this declaration, I have considered the following: my own knowledge and experience, including my work experience in the fields of

mechanical engineering, computer science, biomedical engineering, and electrical engineer; my experience in teaching those subjects; and my experience in working with others involved in those fields. In addition, I have analyzed various publications and materials, in addition to other materials I cite in my declaration.

20. My opinions, as explained below, are based on my education, experience, and expertise in the fields relating to the '611 patent. Unless otherwise stated, my testimony below refers to the knowledge of one of ordinary skill in the fields as of the Critical Date, or before. Any figures that appear within this document have been prepared with the assistance of Counsel and reflect my understanding of the '611 patent and the prior art discussed below.

## II. OVERVIEW OF THE CONCLUSION FORMED

21. This declaration explains the conclusions that I have formed based on my analysis. To summarize those conclusions, based upon my knowledge and experience and my review of the prior art publications listed above, I believe that:

- Claims 1-7 and 25-28 are rendered obvious by Avendano in view of Visser.
- Claims 8-16, 23, and 24 are rendered obvious by Avendano in view of Visser and Bisgaard.
- Claims 17-19 are rendered obvious by Avendano in view of Visser, Bisgaard, and Hou.

- Claims 20-22 are rendered obvious by Avendano in view of Visser, Bisgaard, Hou, and Frequency Art (Byrne, Burnett, and/or Berglund).

## III. LEVEL OF ORDINARY SKILL IN THE ART

22. In my opinion, one of ordinary skill in the art relating to, and at the time of, the invention of the ’611 patent would have been someone with at least a bachelor of science in electrical engineering, computer engineering, computer science, mechanical engineering, or a related discipline, with at least two years of relevant experience in a field related to acoustics, speech recognition, speech detection, or signal processing. A greater amount of education, *i.e.*, a doctorate in electrical engineering, computer engineering, computer science, mechanical engineering, or a related discipline would also qualify for the hypothetical person of ordinary skill in the art in lieu of fewer years of work experience. Additional education or industry experience may compensate for a deficit in one of the other aspects of the requirements stated above.

23. Based on my experiences, I have a good understanding of the capabilities of one of ordinary skill. Indeed, I have taught, participated in organizations, and worked closely with many such persons over the course of my career. Based on my knowledge, skill, and experience, I have an understanding of the capabilities of one of ordinary skill. For example, from my industry experience, I am familiar with what an engineer would have known and found predictable in the art. From

teaching and supervising my post-graduate students, I also have an understanding of the knowledge that a person with this academic experience possesses. Furthermore, I possess those capabilities myself.

## IV. LEGAL STANDARDS

### A. Terminology

24. I have been informed by Counsel and understand that the best indicator of claim meaning is its usage in the context of the patent specification as understood by one of ordinary skill. I further understand that the words of the claims should be given their plain meaning unless that meaning is inconsistent with the patent specification or the patent's history of examination before the Patent Office. Counsel has also informed me, and I understand that, the words of the claims should be interpreted as they would have been interpreted by one of ordinary skill at the time of the invention was made (not today). Because I do not know at what date the invention as claimed was made, I have used the earliest possible priority date of the '611 patent as the point in time for claim interpretation purposes. That date was October 24, 2008, the Critical Date.

### B. Legal Standards for Anticipation

25. I have been informed by Counsel and understand that documents and materials that qualify as prior art can render a patent claim unpatentable as anticipated. I am informed by Counsel and understand that all prior art references are to be looked at from the viewpoint of a person of ordinary skill in the art.

26. I am informed by Counsel and understand that a challenged claim is unpatentable as "anticipated" under 35 U.S.C. § 102 if it is determined that all the limitations of the claim are described in a single prior art reference. I am informed by Counsel and understand that, to anticipate a claim, a prior art reference must disclose, either expressly or inherently, each and every limitation of that claim and enable one of ordinary skill in the art to make and use the invention.

27. I have been informed by Counsel and understand that in an *inter partes review*, "the petitioner shall have the burden of proving a proposition of unpatentability," including a proposition of anticipation, "by a preponderance of the evidence." 35 U.S.C. §316(e).

### C. Legal Standards for Obviousness

28. I have been informed by Counsel and understand that documents and materials that qualify as prior art can render a patent claim unpatentable as obvious. I am informed by Counsel and understand that all prior art references are to be looked at from the viewpoint of a person of ordinary skill in the art at the time of the invention, and that this viewpoint prevents one from using his or her own insight or hindsight in deciding whether a claim is obvious.

29. I have been informed by Counsel and understand that a claim is unpatentable for obviousness under 35 U.S.C. § 103 (in the pre-AIA form of that statute that applies to the '611 patent) "if the differences between the subject matter sought to

be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains." I am informed by Counsel and understand that obviousness may be based upon a combination of references. I am informed by Counsel and understand that the combination of familiar elements according to known methods is likely to be obvious when it does no more than yield predictable results. However, I am informed by Counsel and understand that a patent claim composed of several elements is not proved obvious merely by demonstrating that each of its elements was, independently, known in the prior art.

30. I am informed by Counsel and understand that when a patented invention is a combination of known elements, a court must determine whether there was an apparent reason to combine the known elements in the fashion claimed by the patent at issue by considering the teachings of prior art references, the effects of demands known to people working in the field or present in the marketplace, and the background knowledge possessed by a person having ordinary skill in the art.

31. I am informed by Counsel and understand that a patent claim composed of several limitations is not proved obvious merely by demonstrating that each of its limitations was independently known in the prior art. I am informed by counsel for the Patent Owner and understand that identifying a reason those elements would be combined can be important because inventions in many instances rely upon

building blocks long since uncovered, and claimed discoveries almost of necessity will be combinations of what, in some sense, is already known. I am informed by Counsel and understand that it is improper to use hindsight in an obviousness analysis, and that a patent's claims should not be used as a "roadmap."

32. I am informed by Counsel and understand that an obviousness inquiry requires consideration of the following factors: (1) the scope and content of the prior art; (2) the differences between the claims and the prior art; (3) the level of ordinary skill in the pertinent art; and (4) any objective indicia of non-obviousness, such as commercial success, long-felt but unresolved need, failure of others, industry recognition, copying, and unexpected results. I understand that the foregoing factors are sometimes referred to as the "Graham factors."

33. I have been informed by Counsel and understand that an obviousness evaluation can be based on a combination of multiple prior art references. I understand that the prior art references themselves may provide a suggestion, motivation, or reason to combine, but that the nexus linking two or more prior art references is sometimes simple common sense. I have been informed by Counsel and understand that obviousness analysis recognizes that market demand, rather than scientific literature, often drives innovation, and that a motivation to combine references may be supplied by the direction of the marketplace.

34. I have been informed by Counsel and understand that if a technique has been used to improve one device, and a person of ordinary skill at the time of invention would have recognized that it would improve similar devices in the same way, using the technique is obvious unless its actual application is beyond his or her skill.

35. I have been informed by Counsel and understand that practical and common sense considerations should guide a proper obviousness analysis, because familiar items may have obvious uses beyond their primary purposes. I have been informed by Counsel and understand that a person of ordinary skill looking to overcome a problem will often be able to fit together the teachings of multiple prior art references. I have been informed by Counsel and understand that obviousness analysis therefore takes into account the inferences and creative steps that a person of ordinary skill would have employed at the time of invention.

36. I have been informed by Counsel and understand that a proper obviousness analysis focuses on what was known or obvious to a person of ordinary skill at the time of invention, not just the patentee. Accordingly, I understand that any need or problem known in the field of endeavor at the time of invention and addressed by the patent can provide a reason for combining the elements in the manner claimed.

37. I have been informed by Counsel and understand that a claim can be obvious in light of a single reference, without the need to combine references, if the

elements of the claim that are not found explicitly or inherently in the reference can be supplied by the common sense of one of skill in the art.

38. I have been informed by Counsel and understand that secondary indicia of non-obviousness may include (1) a long felt but unmet need in the prior art that was satisfied by the invention of the patent; (2) commercial success of processes covered by the patent; (3) unexpected results achieved by the invention; (4) praise of the invention by others skilled in the art; (5) taking of licenses under the patent by others; (6) deliberate copying of the invention; (7) failure of others to find a solution to the long felt need; and (8) skepticism by experts. I understand that evidence of secondary indicia of non-obviousness, if available, should be considered as part of the obviousness analysis.

39. I have been informed by Counsel and understand that there must be a relationship between any such secondary considerations and the invention, and that contemporaneous and independent invention by others is a secondary consideration supporting an obviousness determination.

40. In sum, my understanding is that prior art teachings are properly combined where one of ordinary skill having the understanding and knowledge reflected in the prior art and motivated by the general problem facing the inventor, would have been led to make the combination of elements recited in the claims. Under this analysis, the prior art references themselves, or any need or problem known in the

field of endeavor at the time of the invention, can provide a reason for combining the elements of multiple prior art references in the claimed manner.

41. I have been informed by Counsel and understand that in an *inter partes review*, "the petitioner shall have the burden of proving a proposition of unpatentability," including a proposition of obviousness, "by a preponderance of the evidence." 35 U.S.C. §316(e).

## V. THE '611 PATENT

### A. Overview of the '611 Patent

42. Prior to the Critical Date of the '611 patent, numerous products, publications, and patents existed that implemented or described the functionality claimed in the '611 patent. Specifically, it was well established to use a pair of physical microphones to construct a pair of virtual microphones, such that one of the virtual microphones is more sensitive to speech than the other. Further, it was also well established to generate such virtual microphones by delaying and filtering signals received from the physical microphones, and forming various combinations of the delayed and filtered signals. Further still, it was also well established to detect voice activity by comparing the virtual microphones to one another to identify characteristics indicative of speech. Further still, it was also well established to use adaptive filters and other signal processing elements to implement these techniques. The methodology of the '611 patent was therefore

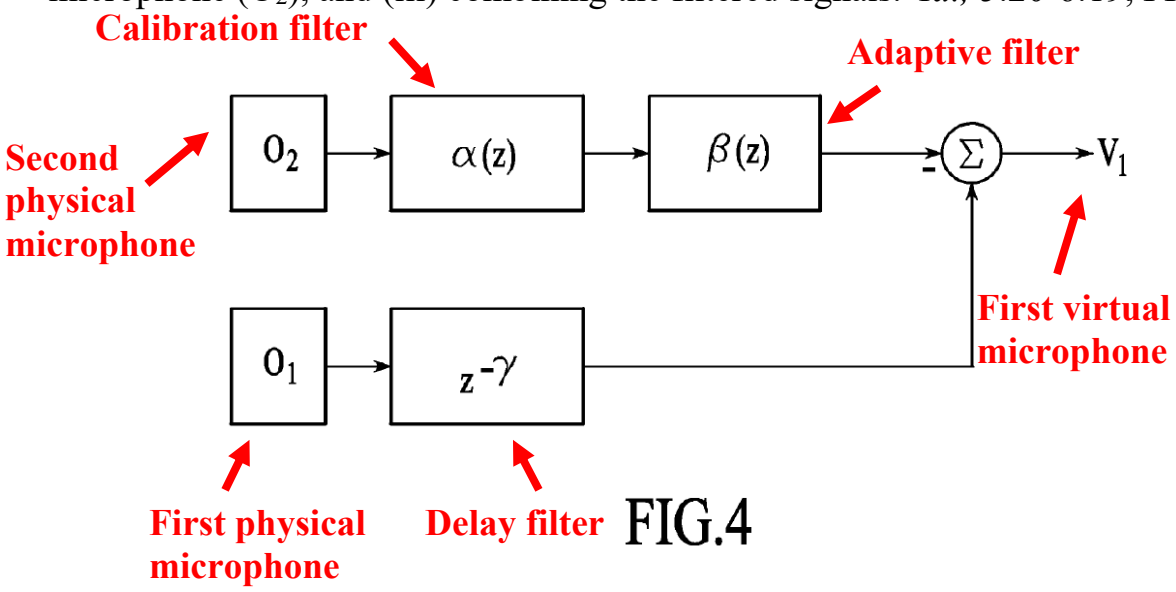
well-known in the prior art as of the Critical Date. Further, to the extent there was any problem to be solved in the ’611 patent, it had already been solved in the prior art systems before the Critical Date of the ’611 patent as I discuss below.

43. The ’611 patent describes the problem of pollution of the speech signal captured by a single microphone with noise from one or more noise sources in the environment, and describes that single-microphone systems can be improved through the use of additional non-acoustic sensors for determining when speech is being produced. APPLE-1001, 1:23-46. However, “an acoustic-only system is desired in some cases (e.g., for reduced cost, as a supplement to the non-acoustic sensor, etc.).” *Id.*, 1:44-46. As I discuss below, systems with two or more microphones that were capable of detecting the presence or absence of speech—and therefore providing improved performance relative to the single-microphone systems identified in the ’611 patent—were well-known in the prior art.

44. The ’611 patent “relates to noise suppression systems, devices, and methods for use in acoustic applications.” *Id.*, 1:16-18. In at least some of the embodiments described in the ’611 patent, a system constructs “virtual microphones ... using two or more omnidirectional microphones and associated signal processing,” and “detects voice activity by generating a ratio of the energies of the virtual microphones, and comparing that ratio to a threshold.” *Id.*, 3:55-7:7, 18:29-32.

45. In general, the ’611 patent describes generating each of a first virtual microphone ($V_1$) and a second virtual microphone ($V_2$) by obtaining signals from two physical microphones, applying various filters to the signals, and forming various combinations of the filtered signals. *Id.*, 5:20-6:19, FIGS. 3-4.

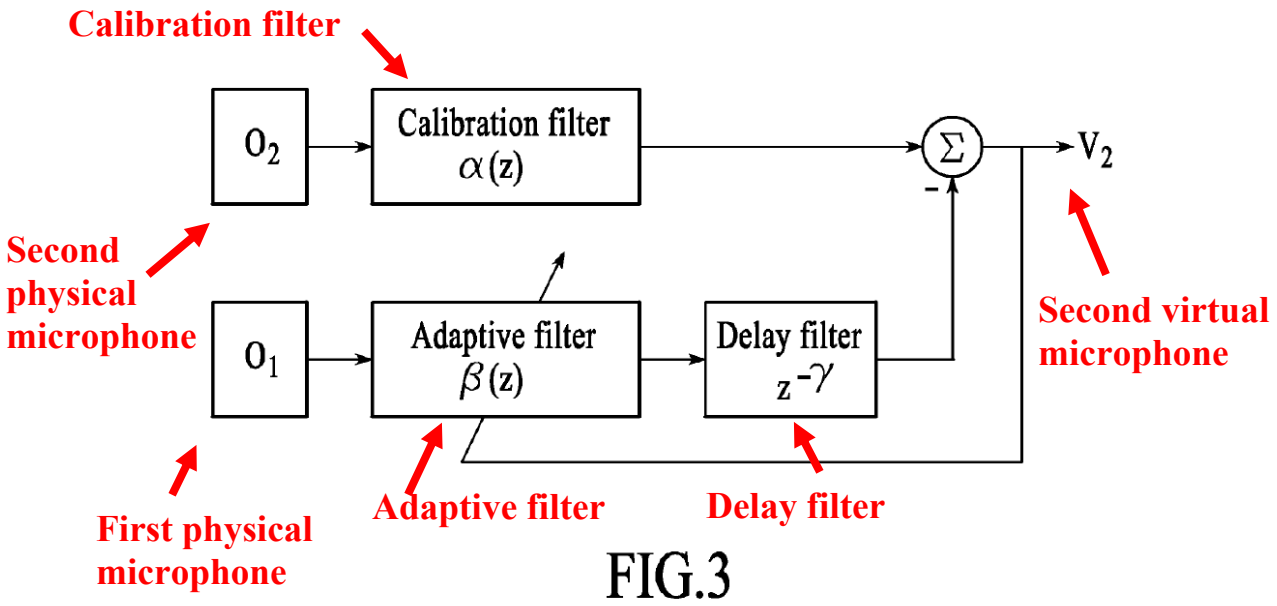
46. As an example, with reference to FIG. 4, the ’611 patent describes that first virtual microphone ($V_1$) is generated by (i) applying a delay filter ($z^{-\gamma}$) to a signal that is output from a first physical microphone ($O_1$), (ii) applying a calibration filter ($\alpha(z)$) and an adaptive filter ($\beta(z)$) to a signal that is output from a second physical microphone ($O_2$), and (iii) combining the filtered signals. *Id.,* 5:20-6:19, FIG. 4.




**APPLE-1001, FIG. 4 (annotated)**

47. As another example, with reference to FIG. 3, the ’611 patent describes that a second virtual microphone ($V_2$) is generated by (i) obtaining a first signal from a first physical microphone ($O_1$) and a second signal from a second physical

microphone ($O_2$), (ii) applying an adaptive filter ($\beta(z)$) and a delay filter ($z^{-\gamma}$) to the first signal, (iii) applying a calibration filter ($\alpha(z)$) to the second signal, and (iv) combining the filtered first signal and the filtered second signal. *Id.,* 5:20-6:19, FIG. 3.




**APPLE-1001, FIG. 3 (annotated)**

48. According to the '611 patent, a calibration filter ($\alpha(z)$), an adaptive filter ($\beta(z)$), and a delay filter ($z^{-\gamma}$) are used to construct a virtual microphone $V_2$ from two physical microphones $O_1$ and $O_2$. In particular, the calibration filter ($\alpha(z)$) compensates a microphone such that a response of the second physical microphone $O_2$ matches a response of the first physical microphone $O_1$. *Id.*, 5:20-34. Further, the adaptive filter ($\beta(z)$) "describes the relationship between $O_1$ and calibrated $O_2$

for speech." *Id.* Further, the delay filter ($z^{-\gamma}$) applies a fixed delay "that depends on the size of the [microphone] array." *Id*.

49. As described by the '611 patent, the ratio of energies of the first and second virtual microphones may be used "with any detection system . . . to determine when speech is occurring." *Id*., 6:20-10:8, FIGS. 5-11. In particular, the ratio is compared to a threshold value to determine the presence of acoustic voice activity. *Id.,* 6:47-51, 7:5-7, FIGS. 5-11. For instance, when the ratio is greater than a threshold value, this is indicative of a presence of acoustic voice activity. *Id.* In contrast, when the ratio is less than the threshold value, this is indicate of an absence of acoustic voice activity. *Id.*

### B. Prosecution History of the '611 Patent

50. Only one non-final Office Action was issued during the prosecution of the application from which the '611 patent issued. APPLE-1002, 193-198. The Office Action contained a single rejection—a provisional double patenting rejection of all of the pending claims over claims of a co-pending application in the family (*i.e.,* claims 1-42 of U.S. Patent Application No. 12/606,146, now U.S. Patent No. 8,321,213). *Id*., 195-196. A notice of allowance was issued in response to a terminal disclaimer. *Id.*, 218-219, 227-233.

## VI. SUMMARY OF THE PRIOR ART

### A. Overview of Avendano

51. Avendano relates to determining "inter-microphone level differences (ILD) ... based on energy level differences of a pair of omni-directional microphones," and "utilizing [ILD] to attenuate noise and enhance speech." APPLE-1005, 2:5-9.

52. Avendano discloses that an "audio device 104" includes at least a "primary microphone 106" and a "secondary microphone 108." *Id.*, 3:27-35, FIGS. 1a-1b. Each of Avendano's microphones may be an "omni-directional microphone." *Id.*

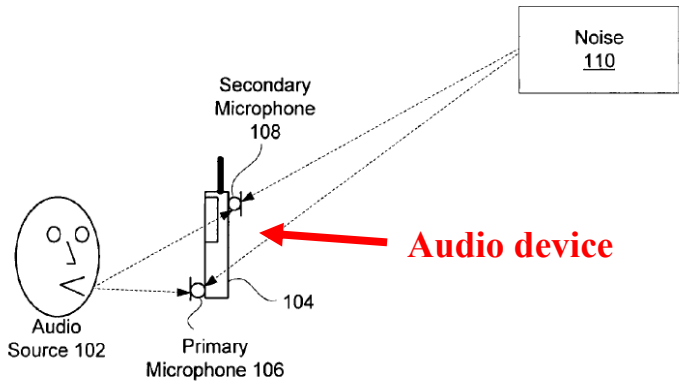



FIG. 1a

**APPLE-1005, FIG. 1a (annotated)**

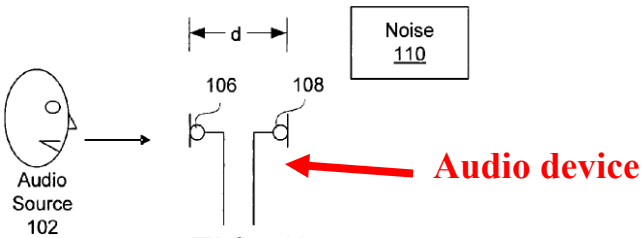



FIG. 1b

**APPLE-1005, FIG. 1a (annotated)**

53. Avendano describes that the "primary microphone 106" and a "secondary microphone 108" are arranged at different locations on the "audio device 104," such that the microphones detect speech from an "audio source 102" differently. *Id.*, 3:45-49. For example, with reference to Figures 1a and 1b, Avendano discloses that "the primary microphone 106 is much closer to [an] audio source 102 than the secondary microphone 108." *Id.* During to this difference in locations between the two microphones, "the intensity level is higher for the primary microphone 106 resulting in a larger energy level during a speech/voice segment." *Id*. Further, Avendano describes that, "because directional response of the primary microphone 106 is highest in the direction of the audio source 102 and directional response of the secondary microphone 108 is lower in the direction of the audio source 102, the level difference is highest in the direction of the audio source 102 and lower elsewhere." *Id.*, 3:49-55.

54. By combining signals from the pair of omni-directional microphones to form a pair of virtual directional microphones, Avendano provides the benefits of capturing speech signals and suppressing noise signals that would be available from use of a pair of directional microphones, but with smaller size and cost. *Id.*, 1:64-66.

55. As described by Avendano, this "level difference may then be used to discriminate speech and noise in the time-frequency domain." *Id.*, 3:55-57. For example, Avendano describes obtaining signals from the two microphones (*e.g.,* signals $x_1$ and $x_2$), and processing the signals using a "differential microphone array (DMA) module 302" to "create two different directional patterns around the audio device 104." *Id.*, 4:20-41. As Avendano explains, "[e]ach directional pattern is a region about the audio device 104 in which sounds generated by an audio source 102 within the region may be received by the microphones 106 and 108 with little attenuation," and "[s]ounds generated by audio sources 102 outside of the directional pattern may be attenuated." *Id.*, 4:41-46.

56. For example, the DMA module 302 generates two different processed signals, each having a different directional pattern for receiving sounds. *Id.*, 4:47-52, 5:25-35.

57. In particular, the DMA module 302 generates a first processed signal having a directional pattern for receiving sounds "within a front cardioid region around the audio device 104." *Id.* This first processed signal may be referred to as a "cardioid primary signal ($C_f$)." *Id.*

58. As another example, the DMA module 302 also generates a second processed signal having a directional pattern for receiving sounds "within a back

cardioid region around the audio device 104." *Id.* This second processed may be referred to as a "cardioid secondary signal ($C_b$)." *Id.*, 4:47-52, 5:25-35.

59. Avendano's front and back cardioid regions are shown in Figure 6. As shown in Figure 6, the front cardioid region primarily extends from an origin point towards a front of the audio device 104 (*e.g.,* in the 0° direction), and accordingly, in the direction of the audio source 102. *Id.*, 9:29-42, FIG. 6. In contrast, the back cardioid region primarily extends from the origin point towards a rear of the audio device 104 (*e.g.,* in the 180° direction), and accordingly, in the direction away from the audio source 102. *Id.*

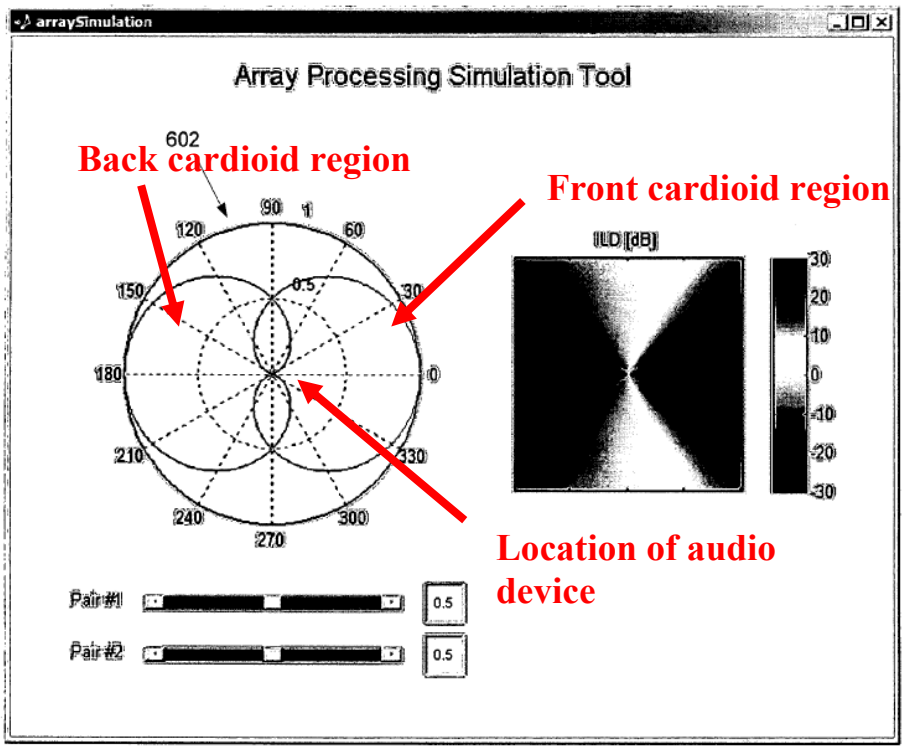



FIG. 6

**APPLE-1005, FIG. 6 (annotated)**

60. As shown in Figure 4a, Avendano discloses that the "cardioid primary signal ($C_f$)" is generated by combining (i) the signal $x_1$ from the primary microphone 106, and (ii) the signal $x_2$ from the secondary microphone 108 (*e.g.,* the signal $x_2$ having been filtered by a "delay node 404" and a "gain module 406"). *Id.*, 5:15-35. Further, Avendano discloses that the "cardioid secondary signal ($C_b$)" is generated by combining (i) the signal $x_2$ from the secondary microphone 108, and (ii) the signal $x_1$ from the primary microphone 106 (*e.g.,* the signal $x_1$ having been filtered by a "delay node 402"). *Id.* The "delay nodes" can be implemented using one or more filters (*e.g.,* "allpass filters"). *Id.,* 8:47-51.

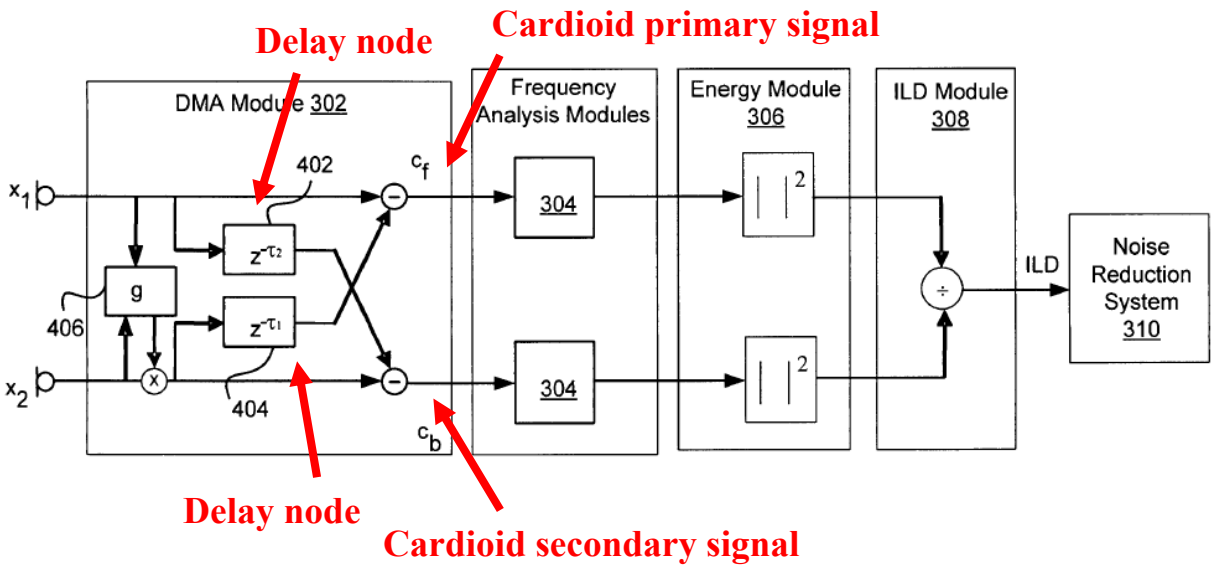



FIG. 4a

**APPLE-1005, FIG. 4a (annotated)**

61. As described by Avendano, the presence of speech can be detected by comparing the energies of the "cardioid primary signal ($C_f$)" and the "cardioid secondary signal ($C_b$)." *Id.*, 5:49-6:34. In particular, Avendano discloses that calculating the ratio between (i) the energy of the "cardioid primary signal ($C_f$)" and (ii) the energy of the "cardioid secondary signal ($C_b$)," and comparing the ratio of a threshold to determine a presence (or absence) of speech. *Id.*

62. For example, Avendano describes calculating an "energy level" ($E_f$) of a "cardioid primary signal ($C_f$)" based on the following equation, in which the primary signal is squared and integrated over a time frame:

$$E_f(t, \omega) = \int_{frame} |C_f(t', \omega)|^2 \, dt',$$

**APPLE-1005, 5:60**

63. Further, Avendano similarly describes calculating an "energy level" ($E_b$) of a "cardioid secondary signal ($C_b$)" based on the following equation, in which the secondary signal is squared and integrated over a time frame:

$$E_b(t, \omega) = \int_{frame} |C_b(t', \omega)|^2 \, dt'.$$

**APPLE-1005, 6:5**

64. Avendano describes determining the ratio between these two energy levels (ILD):

$$ILD(t, \omega) = \frac{\int |C_f(t', \omega)|^2 \, dt'}{\int_{frame} |C_b(t', \omega)|^2 \, dt'}.$$

**Energy level of $C_f$**

**Energy level of $C_b$**

**APPLE-1005, 6:16 (annotated)**

65. Avendano also describes determining the presence or absence of speech by comparing the ratio (ILD) to a "threshold":

$$\lambda_I(t, \omega) = \begin{cases} \approx 0 & \text{if} \quad ILD(t, \omega) < \text{threshold} \\ \approx 1 & \text{if} \quad ILD(t, \omega) > \text{threshold} \end{cases}$$

**APPLE-1005, 6:61**

66. As described by Avendano, if the ratio (ILD) is "smaller than a threshold value (e.g., threshold=0.5) above which speech is expected to be," a value $\lambda_1$ is set to zero to indicate an absence of speech. *Id.*, 6:58-7:3. In contrast, if the ratio (ILD) "starts to rise (e.g., because speech is present within the large ILD region), $\lambda_1$ increases" (*e.g.,* is set to one) to indicate a presence of speech. *Id.*.

67. Avendano's ratio (ILD) is used to process audio signals "through a noise reduction system 310" to "enhance the speech of the primary acoustic signal." *Id.*, Abstract, 6:35-36. *See also id.,* 6:36-8:23, 10:18-50, FIGS. 7-8. For example, as

described by Avendano, when speech is detected based on the ratio (ILD), a "noise estimate module 312 slows down [a] noise estimation process," such that "the speech energy does not contribute significantly to the final noise estimate." *Id.,* 7:4-6. This techniques enables a "filtering module 314" to generate a filter to "suppress noise/enhance noise" more effectively. *Id.,* 7:4-8:23, 10:26-49, 6:41-58.

### B. Overview of Visser

68. Visser relates to using a "voice activity detector" to "improv[e] the quality of a speech signal extracted from noisy acoustic environment." APPLE-1006, 6:57-60.

69. With reference to Figure 1, Visser describes a "speech separation process 100" for separating speech from "sound signals from microphones ... 102 and 104." *Id.*, 8:4-8. According to the "speech separation process 100," a "voice activity detector (VAD) 106 ... receives two input signals 105, with one of the signals defined to hold a stronger speech signal." *Id.*, 8:33-35. Based on the two "input signals 105," the "VAD 106 generates a control signal 107 ... to activate the signal separation process only when speech is occurring." *Id.*, 8:36-40.

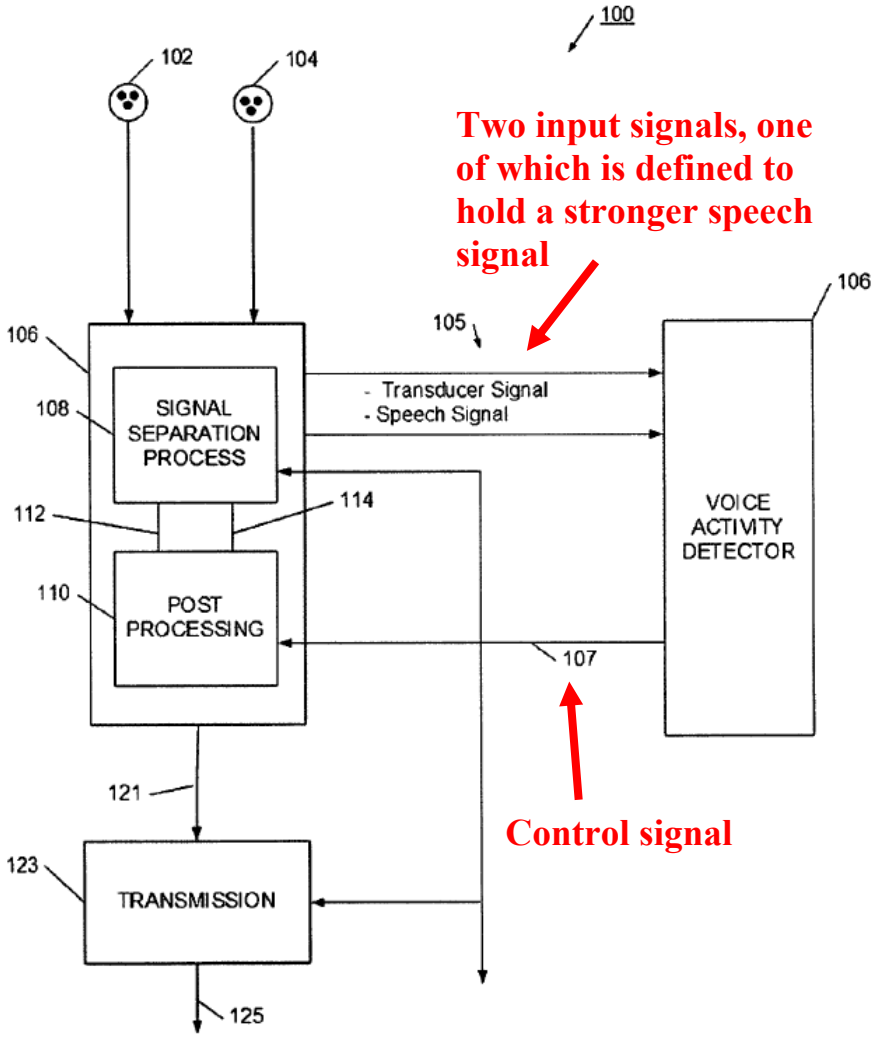



FIG. 1

**APPLE-1006, FIG. 1 (annotated)**

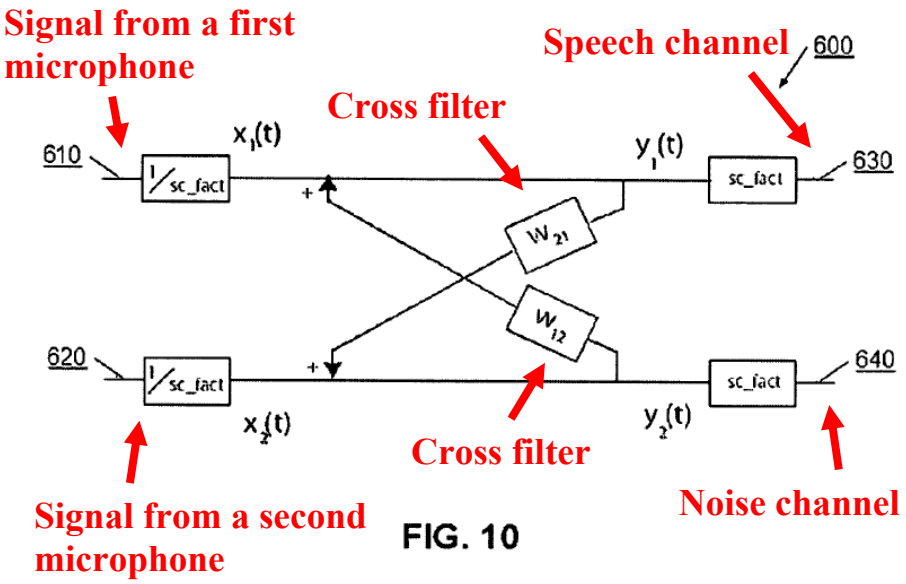
70. Visser describes that the "signal separation process 100" is performed based on signals generated by an "ICA [independent component analysis] or BSS [blind signal source] processing function." *Id.*, 8:16-18, 16:3-28, 17:29-30, FIG. 10. An example embodiment of an "ICA or BSS processing function" is shown in FIG. 10. *Id.*

71. According to this embodiment, the "ICA or BSS processing function" receives "signals $X_1$ and $X_2$ ... from channels 610 and 620." *Id.*, 17: 36-39. As Visser describes, each of these signals "typically ... would come from at least one microphone," such as the microphones 102 and 104. *Id.*

72. Further, the "ICA or BSS processing function" generates two channels based on the "signals $X_1$ and $X_2$," each having a different amounts of speech and noise. *Id.*, 17:39-44. In particular, the "ICA or BSS processing function" generates "a channel 630 of separated signals $U_1$" that "<u>contains predominantly desired signals,</u>" namely speech. *Id.*, 17:39-44 (emphasis added). The "channel 630" may also be referred to as a "speech channel." *Id.* Further, the "ICA or BSS processing function" generates "a channel 540 of separated signals $U_2$" that "<u>contains predominantly noise signals</u>." *Id* (emphasis added). The "channel 540" may also be referred to as a "noise channel." *Id.*

73. Visser's system generates the "speech channel" and the "noise channel" by applying "cross filters" to the signals, and forming various combinations of the filtered signals. *Id.*, 17:29-21:16, FIG. 10. For example, Visser's system generates the "speech channel" by (i) filtering the "input signal $X_2$" by a "cross filter $w_{12}$," and (ii) combining the filtered "input signal $X_2$" with the "input signal $X_1$." *Id.*, 17:29-50, FIG. 10. Further, Visser' system generates the "noise channel" by (i)

filtering the "input signal $X_1$" by a "cross filter $w_{21}$," and (ii) combining the filtered "input signal $X_1$" with the "input signal $X_2$." *Id.*




**APPLE-1006, FIG. 10 (annotated)**

74. The "speech channel" and the "noise channel" are provided as inputs to the "VAD 106" (*e.g.*, the "input signals 105") to determine "when speech is present." *Id.*, 8:33-36.

75. In general, Visser's cross filters $W_{21}$ and $W_{12}$ are configured to facilitate the separation of speech from noise. *Id.,* 16:3-27. As described by Visser, "[t]he cross filters $W_{21}$ and $W_{12}$ can have sparsely distributed coefficients over time to capture a long period of time delays." *Id.*, 17:56-57. For example, each of "the cross filters $W_{21}$ and $W_{12}$" can have "one filter coefficient per filter, for example a delay gain factor for the time delay between the output signal and the feedback input signal and an amplitude gain factor for amplifying the input signal," or "dozens, hundreds

or thousands of filter coefficients" to facilitate a speech separation process. *Id.*, 17:56-64.

76. Further, the "speech separation process ... may be adaptive and learn according to the specific acoustic environment," and to "adapt to particular microphone placement, the acoustic environment, or a particular user's speech." *Id.*, 9:8-12. For example, in an "ICA process," each filter "ha[s] an adaptable and adjustable filter coefficient." As Visser describes, "the coefficients are adjusted to improve separation performance ... and the new coefficients are applied. This continual adaptation of the filter coefficients enables the [speech separation] process ... to provide a sufficient level of separation, even in a changing acoustic environment." *Id.*, 16:65-17:4, FIG. 9.

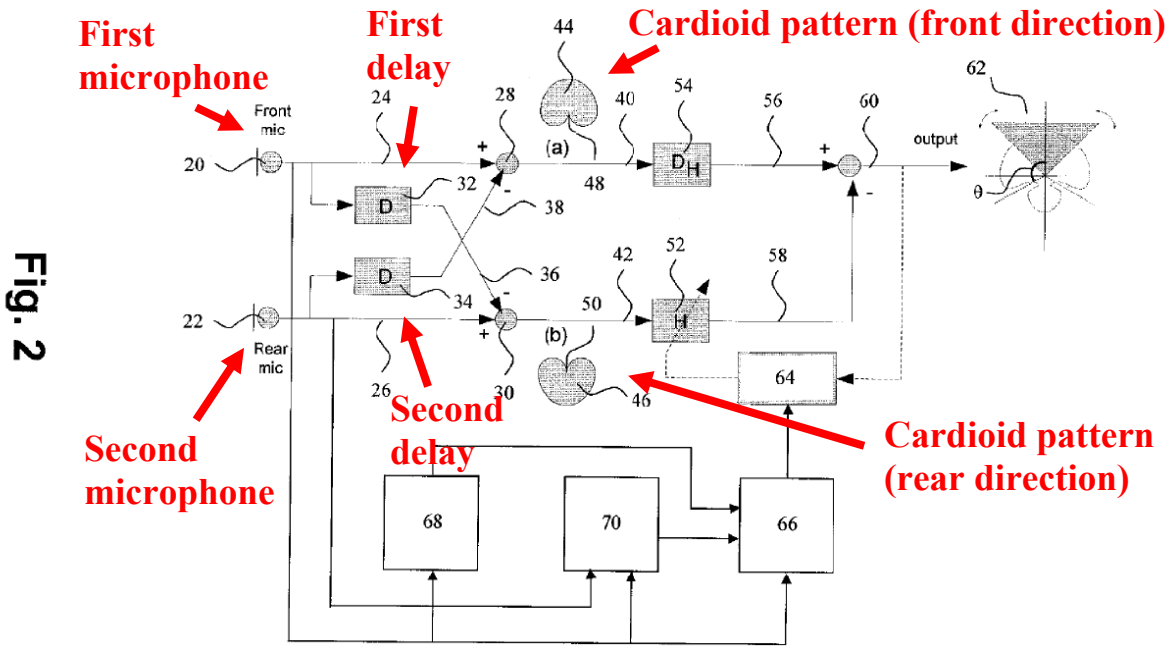
77. Visser uses these two signals as input to a Voice Activity Detection element. *Id.* 12:11-18. For example, as described by Visser, "[t]he signal separation process generates a noise-dominant signal as well as a signal having speech content as shown in block 252. The voice activity detector monitors the energy level in each of the signals, and compares the measured energy level, as shown in block 253. In one simple implementation, the signals are monitored for when the difference in energy levels between the signals exceeds a pre-defined threshold." *Id.*

### C. Overview of Bisgaard

78. Bisgaard relates to "a hearing instrument, such as a hearing aid, an implantable hearing prosthesis, a head set, a mobile phone, etc., with a signal processor for signal processing." APPLE-1010, [0002]. *See also* APPLE-1011, 1:3-5.

79. Bisgaard's Figure 2 "schematically illustrates the signal processing of a hearing instrument." APPLE-1010, [0039]. *See also* APPLE-1011, 5:1-2. According to this configuration, two microphone signals are obtained from two "microphones 20, 22," respectively, and are processed to obtain cardioid patterns extending in different directions. APPLE-1010, [0041]-[0042], FIG 2. *See also* APPLE-1011, 5:18-31, FIG. 2. For example, the two microphone signals are processed to obtain (i) a "cardioid pattern 44 ... ha[ving] <u>null</u> direction 48 at 180° azimuth, i.e. pointing in the rear direction of the user," and (ii) a "cardioid pattern 46 ... ha[ving] a <u>null</u> direction 50 at 0° azimuth, i.e. pointing in the front direction of the user." APPLE-1010, [0041]-[0042] (emphasis added), FIG 2. *See also* APPLE-1011, 5:18-31, FIG. 2. That is, the "cardioid pattern 44" is directed towards the front of the device (as the "null direction," or absence of the pattern, is directed towards the rear of the device). Further, the "cardioid pattern 46" is directed towards the rear of the device (as the "null direction," or absence of the pattern, is directed towards the front of the device).

80. Bisgaard describes that the "cardioid pattern 44" and the "cardioid pattern 46" are generated using "delay[s] 32, 34." APPLE-1010, [0042], FIG 2. *See also* APPLE-1011, 5:21-24, FIG. 2. For example, the "cardioid pattern 44" is generated by (i) applying a "delay 34" to the microphone signal obtained from the "microphone 22," and (ii) combining the delayed microphone signal with the microphone signal obtained from the "microphone 20." APPLE-1010, [0042], FIG 2. *See also* APPLE-1011, 5:21-24, FIG. 2. Further, the "cardioid pattern 46" is generated by (i) applying a "delay 32" to the microphone signal obtained from the "microphone 20," and (ii) combining the delayed microphone signal with the microphone signal obtained from the "microphone 22." APPLE-1010, [0042], FIG 2. *See also* APPLE-1011, 5:21-24, FIG. 2. As described by Bisgaard, each of the "delay[s] 32, 34 ... delays the digitized sound signal [received from the microphones 20, 22] by the amount of time used by a sound signal to propagate in the 0° azimuth direction from the front microphone 20 to the rear microphone 22. APPLE-1010, [0042], FIG 2. *See also* APPLE-1011, 5:21-24, FIG. 2.




**APPLE-1010, FIG. 2 (annotated)**

81. Bisgaard claims priority to U.S. 60/816,244 ("Bisgaard Provisional"). APPLE-1011, Cover. The Bisgaard Provisional supports at least Bisgaard's claim 1. For example, the Bisgaard Provisional discloses "[a] hearing instrument." APPLE-1011, claim 1, 1:3-5, 2:1-9, 2:15, 4:1-4, FIG. 1. Further, the Bisgaard Provisional discloses "at least two microphones for reception of sound and conversion of the received sound into corresponding electrical sound signals that are input to the signal processor." APPLE-1011, claim 1, 1:13-15, 2:1-9, 4:1-9, 5:1-20, FIGS. 1-2. Further, the Bisgaard Provisional discloses "wherein the signal processor is configured to process the electrical sound signals into a combined

signal with a directivity pattern with at least one adaptive null direction θ." APPLE-1011, claim 1, 2:1-9, 3:3-13, 5:21-6:6, 11:30-12:18, FIGS. 1-2. Further, the Bisgaard Provisional discloses "wherein the signal processor is further configured to prevent the at least one null direction θ from entering a prohibited range of directions, wherein the prohibited range is a function of a parameter of the electrical sound signals." APPLE-1011, claim 1, 2:6-9, 3:3-13, 6:18-19.

### D. Overview of Hou

82. Hou relates to "[i]mproved approaches to matching sensitivities of microphones in multi-microphone directional processing systems ... so that directional noise suppression is robust." APPLE-1008, Abstract.

83. As an example, Hou's Figure 5 shows "a two-microphone directional processing system 500" for "compensat[ing] (or correct[ing]) for the relative difference in sensitivity between ... mismatched first and second microphones" and producing an "output signal ... hav[ing] robust directionality despite a mismatch between the first and second microphones." *Id.*, 5:25-56, FIG. 5.

84. Hou's system compensates for the relative difference in sensitivity between microphones, at least in part, by processing signals output by the microphones and comparing the processing signals to one another. *Id.*, 5:25-6:22, FIG. 5.

85. For example, Hou's system receives "a first electronic sound signal" from a "microphone 502," and "estimates the minimum for the first electronic sound signal" using a "first minimum estimate unit 508." *Id.*, 5:27-41

86. Further, Hou's system receives "a second electronic sound signal" from a "microphone 504," delays the "second electronic sound signal" using a "delay unit 516," and "estimates the minimum for the second electronic sound signal" using a "second minimum estimate unit 510." *Id.*

87. Further, Hou's system compares the minimum for the first electronic sound signal and the second electronic sound signal to determine the degree of which one of the signals is scaled in order to compensate for a difference in sensitivity between the microphones. *Id.,* 5:42-6:2, FIG. 5.

88. In particular, in Hou's system, a "divide unit 512 produces a quotient by dividing the first minimum estimate by the second minimum estimate." *Id.*, 5:42-43. This "quotient represents a scaling amount that is sent to a multiplication unit 514." *Id*, 5:43-45. Further, in Hou's system, "[t]he second electronic sound signal is then multiplied with the scaling amount to produce a compensated sound signal." *Id.,* 5:45-47.

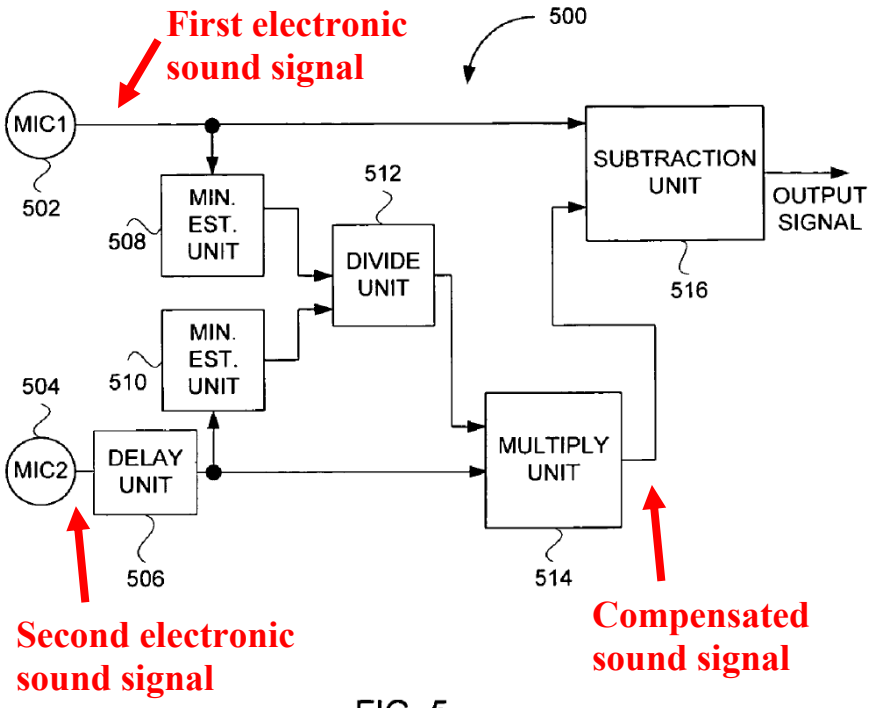



FIG. 5

**APPLE-1008, FIG. 5 (annotated)**

89. As described by Hou, "[t]he compensated sound signal is thus compensated (or corrected) for the relative difference in sensitivity between the mismatched first and second microphones 502 and 504." *Id.,* 5:47-50.

90. Further, in Hou's system, "[a] subtraction unit 516 then subtracts the compensated electronic sound signal from the first electronic sound signal to produce an output signal ... hav[ing] robust directionality despite a mismatch between the first and second microphones 502 and 504." *Id.,* 5:50-56.

91. As described by Hou, it is beneficial to generate a compensated sound signal in the context of a noise suppression system. *Id.* 9:28-41. For example, a noise

suppression system can generate such a compensated sound signal to ensure that "directional noise suppression is not affected by microphone mismatch, ... the drift of microphone sensitivity over time, ... [or] the non-uniform distribution of sound pressure in real-life application." *Id.* 9:28-38. Accordingly, "superior directionality and consistent Signal-to-Noise Ratio (SNR) [is achieved] across all conditions." *Id.*, 9:38-41.

### E. Overview of Byrne

92. Byrne describes a study that investigated "[t]he long-term average speech spectrum (LTASS) and some dynamic characteristics of speech ... for 12 languages," including "English (several dialects), Swedish, Danish, German, French (Canadian), Japanese, Cantonese, Mandarin, Russian, Welsh, Singhalese, and Vietnamese." APPLE-1009, Abstract. As Byrne describes, "LTASS is widely used in the prescription of hearing aid fittings." *Id.,* 2019, Abstract.

93. With reference to FIGS. 1, Byrne describes "[m]ale and female long-term average speech spectrum (LTASS) values for five samples of English." *Id.*, FIG. 1. *See also Id.,* 2112-2120. In particular, each of the plots shown in FIG. 1 indicates the average of the sound pressure level (SPL) of speech samples for frequencies in a range from approximately 63 Hz to approximately 16 kHz. *Id.* The average of the SPL corresponds to an intensity of sound (*e.g.,* as measured by a microphone).

94. According to Byrne, for speech samples obtained from English speakers from Australia, New Zealand, "USA Columbus," "USA Memphis," and the United Kingdom, the average of the SPL of those samples peaked at approximately 500 Hz. *Id.* This indicates that the most significant spectral components of speech for speakers from all these regions would generally be expected at or around a frequency of 500 Hz. For example, Byrne's FIG. 1 shows that these speech samples include significant spectral components in frequency ranges greater than 200 Hz, such as in a frequency range between 250 Hz and 1250 Hz, and in a frequency range between 200 Hz and 3000 Hz for all of these speakers.

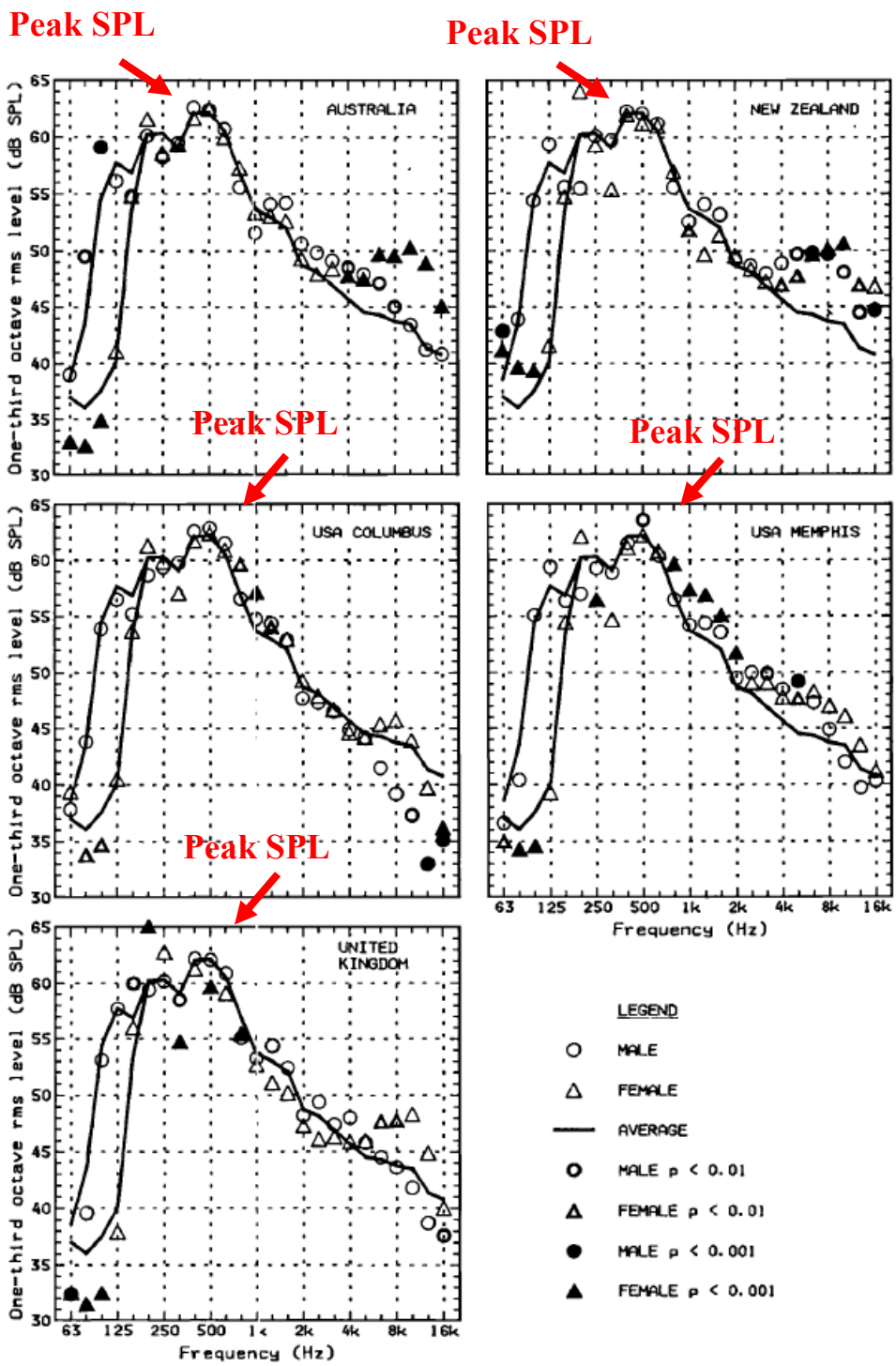



FIG. 1. Male and female long-term average speech spectrum (LTASS) values for five samples of English. Solid line shows LTASS average across 17 speech samples (all samples except Arabic), males and females separately for frequencies below 160 Hz, combined for higher frequencies.

**APPLE-1009, FIG. 1 (annotated)**

95. Further, Byrne discloses that the LTASS values vary for different types of speech. *Id.*, FIGS. 1-9, 2112-2120. For example, the LTASS value may vary depending on the language of the speech, the gender of the speaker, among other factors. *Id*.

### F. Overview of Burnett

96. Burnett describes "[s]ystems and methods ... for detecting voiced and unvoiced speech in acoustic signals having varying levels of background noise." APPLE-1012, Abstract. Burnett's system "receive[s] acoustic signals at two microphones, and generate[s] difference parameters between the acoustic signals received at each of the two microphones." These "difference parameters are representative of the relative difference in signal gain between portions of the received acoustic signals." *Id.* Further, Burnett's system (i) "identif[ies] information of the acoustic signals as unvoiced speech when the difference parameters exceed a first threshold," and (ii) "identif[es] information of the acoustic signals as voiced speech when the difference parameters exceed a second threshold." *Id.*

97. Burnett's system also "group[s] the utterances by their spectral characteristics," which would enable its system to "work better in noisy environments." *Id.,* [0063]. Specifically, Burnett discloses:

> While the systems described above have been described with reference to separating voiced and unvoiced speech from background acoustic noise, there are no reasons more complex classifications can not be made. For more in-depth characterization of speech, the system can bandpass the information from [two microphones] Mic 1 and Mic 2 so that it is possible to see which bands in the Mic 1 data are more heavily composed of noise and which are more weighted with speech. Using this knowledge, it is possible to group the utterances by their spectral characteristics similar to conventional acoustic methods; this method would work better in noisy environments.
>
> As an example, the "k" in "kick" has significant frequency content form 500 Hz to 4000 Hz, but a "sh" in "she" only contains significant energy from 1700-4000 Hz. Voiced speech could be classified in a similar manner. For instance, an /i/ ("ee") has significant energy around 300 Hz and 2500 Hz, and an /a/ ("ah") has energy at around 900 Hz and 1200 Hz. This ability to discriminate unvoiced and voiced speech in the presence of noise is, thus, very useful.

*Id.*, [0063]-[0064].

### G. Overview of Berglund

98. Berglund describes a study regarding "[t]he source of human exposure to low-frequency noise and its effects." APPLE-1013, Abstract.

99. In Figure 3, Berglund shows the spectral components of common noises experienced by passengers of "road transportation vehicles":

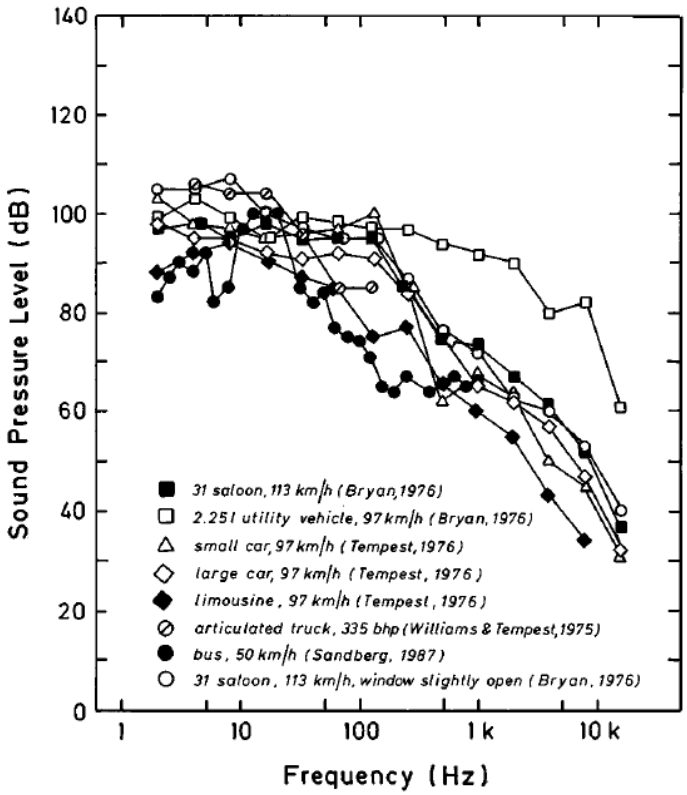



FIG. 3. Passenger noise exposure in road transportation vehicles as a function of frequency.

APPLE-1013, FIG. 3

100. Berglund discloses that the sound pressure level (SPL)—corresponding to an intensity of sound—for noise produced by "road transportation vehicles" is highest

in the "low frequency noise" range (*e.g.,* a range of frequencies less than approximately 250 Hz"). *Id.,* 2985-2987, FIGS. 1, 3.

## VII. GROUND 1: CLAIMS 1-7 AND 25-28 ARE RENDERED OBVIOUS BY AVENDANO IN VIEW OF VISSER

### A. Combination of Avendano and Visser

101. As detailed below, a POSITA would have found it obvious to combine Avendano and Visser.

102. First, both references come from the same field of endeavor of enhancing speech and attenuating noise based on the detection of voice activity. APPLE-1005, Abstract, 1:24-26, 3:13-26, 3:42-60; APPLE-1006, Abstract, 1:19-23, 6:57-7:25, 8:4-47.

103. Second, both references describe identifying voice activity by analyzing (i) a first processed signal representative of speech detected by two physical microphones, and (ii) a second processed signal representative of noise detected by the two physical microphones. APPLE-1005, Abstract, 1:24-26, 3:13-7:9; APPLE-1006, Abstract, 1:19-23, 6:57-7:25, 8:4-47, 17:29-50.

104. Third, modifying Avendano's system in view of Visser's disclosure would have improved Avendano's system by enabling the system to "adapt" and "learn" to separate speech "according to the specific acoustic environment," such that the system can accurately separate speech "even in a changing acoustic environment." APPLE-1006, 16:65-17:4, FIG. 9.

105. As I explain in Section VI(A) incorporated herein), Avendano describes that the presence of speech can be detected by comparing the energies of two processed signals having different directional patterns, namely (i) a first processed signal (*e.g.*, a "cardioid primary signal ($C_f$)" having a directional pattern directed to the front of the "audio device 104") and (ii) a second processed signal (*e.g.*, a "cardioid secondary signal ($C_b$)" having a directional pattern directed to the back of the "audio device 104"). APPLE-1005, 4:47-7:9.

106. For example, according to Avendano, a source of speech (*e.g.,* the "audio source 102") is positioned on a front side of the "audio device 104." *Id.,* FIGS. 1a-1b. Thus, the front-facing "cardioid primary signal ($C_f$)" would have a greater response to speech than the back-facing "cardioid secondary signal ($C_b$)." Accordingly, when the energy level of the front-facing "cardioid primary signal ($C_f$)" is sufficiently higher than the energy level of the back-facing "cardioid secondary signal ($C_b$)," this is indicative of a presence of speech activity by the "audio source 102" positioned at the front of the "audio device 104." *Id.,* 6:58-7:3. As Avendano describes, this comparison can be performed by determining the ratio ("inter-microphone level difference," ILD) between the energies of the "cardioid primary signal ($C_f$)" and the "cardioid secondary signal ($C_b$)," and comparing the ratio of a threshold value. *Id.,* 4:47-7:9.

107. As I also explained in Section VI(A), Avendano describes that each of the "cardioid primary signal ($C_f$)" and the "cardioid secondary signal ($C_b$)" is generated by applying a delay to a signal received from one of the physical microphones, and combining the delayed signal with a signal received from the other physical microphone. *Id.*, 5:15-35.

108. Visser discloses determining the presence of speech using principles similar to those taught by Avendano.

109. For example, as I explain in Section VI(B) (incorporated herein), Visser describes a "VAD 106" that determines the presence of speech based on two processed signals, "with one of the signals defined to hold a stronger speech signal." APPLE-1006, 8:33-35. In particular, Visser's processed signals include (i) a "speech channel" that "contains predominantly desired signals," and (ii) a "noise channel" that "contains predominantly noise signals" (*e.g.*, as generated by an "ICA or BSS processing function"). *Id.*, 17:36-44. This is similar to Avendano's processed signals, which includes one processed signal (*e.g.*, a front-facing "cardioid primary signal ($C_f$)") that would have a greater response to speech along a front of an audio device than the other. APPLE-1005, *Id.,* 6:58-7:3.

110. Further, like Avendano, Visser describes that these processed signals can be generated by applying a delay to a signal received from one of the physical microphones, and combining the delayed signal with a signal received from the

other physical microphone. APPLE-1006, 17:39-62, FIG. 10. For example, Visser describes that signals received from each of the physical microphones can be delayed using respective cross filters having "sparsely distributed coefficients over time to capture a long period of time delays," such as a cross filter having "a delay gain factor for the time delay between the output signal and the feedback input signal." *Id.*

111. Visser additionally describes that the separation between the "speech channel" and the "noise channel" can be further enhanced by generating and adapting the filters according to an "ICA process." *Id.,* 9:20-62, 17:29-25:9, FIGS. 10-13. For example, Visser describes that an "ICA process" enables a system to adapt to "the specific acoustic environment," including "[a] particular microphone placement, the acoustic environment, or a particular user's speech," in order "to provide a sufficient level of separation, even in a changing acoustic environment." *Id.,* 9:8-12, 16:65-17:4.

112. In view of Avendano's and Visser's disclosures, a POSITA would have been motivated to incorporate Visser's "ICA process" into Avendano's system. Such a modification would have further enhanced the differences in speech content in Avendano's two processed signals. Accordingly, speech can be detected with a greater degree of accuracy (*e.g.*, by comparing the energy levels of signals representing speech and noise), "even in a changing acoustic environment."

113. Specifically, a POSITA would have recognized that each of Avendano's system and Visser's system is configured to process signals from two physical microphones including the step of filtering one microphone signal, in order to generate a first processed signal representative of speech (*e.g.,* by combining the filtered signal with another microphone signal). APPLE-1005, 5:15-35, FIG. 4a; APPLE-1006, 17:36-44, FIG. 10.

114. For example, Avendano's system is configured to filter a microphone signal $x_2$ using a "delay node 404," and generate a first processed signal representative of speech (*e.g.,* a "cardioid primary signal ($C_f$)") by combining the filtered signal with another microphone signal $x_1$. APPLE-1005, 5:15-35, FIG. 4a.

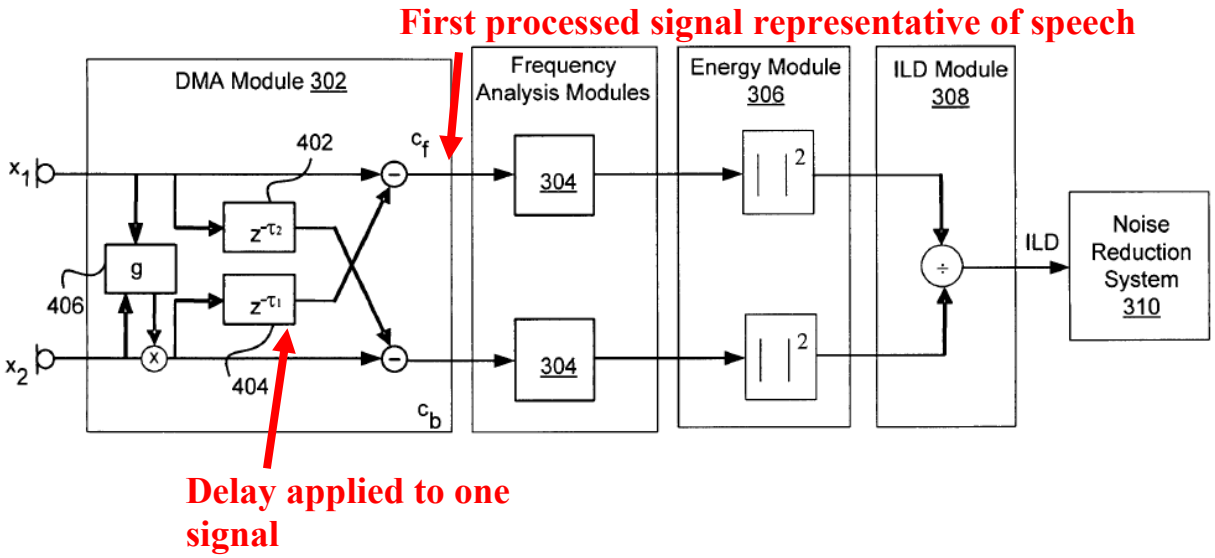



FIG. 4a

**APPLE-1005, FIG. 4a (annotated)**

115. Similarly, a POSITA would have recognized that, like Avendano's "delay node 404," Visser's "cross filter $w_{12}$" is configured to filter one microphone signal (*e.g.*, a "channel 620") in order to generate a first processed signal representative of speech (*e.g.,* a "speech channel, such as the "channel 630") by combining the filtered signal with another microphone signal (*e.g.,* a "channel 610"). APPLE-1006, 17:36-44, FIG. 10.

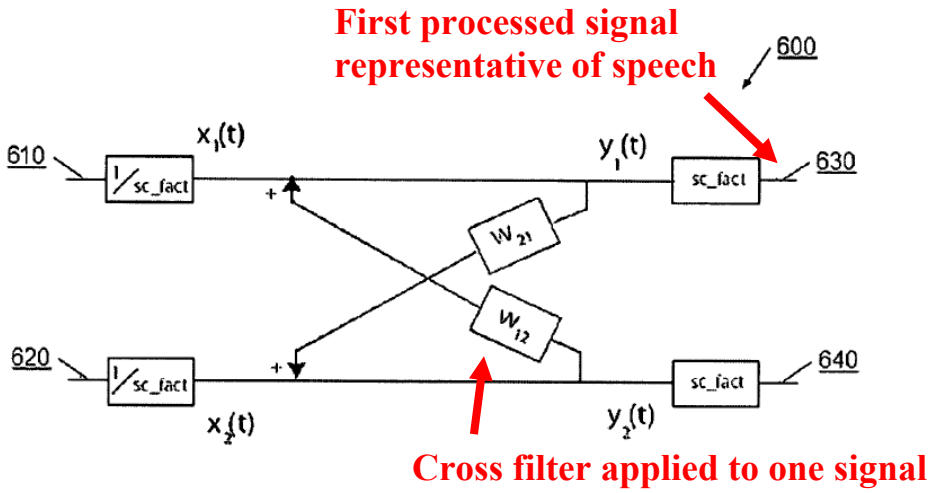



FIG. 10

**APPLE-1006, FIG. 10 (annotated)**

116. A POSITA would have recognized that each of Avendano's system and Visser's systems is configured to process signals from two physical microphones including the step of filtering another microphone signal, in order to generate a second processed signal representative of noise (*e.g.,* by combining the filtered

signal with the other microphone signal). APPLE-1005, 5:15-35, FIG. 4a; APPLE-1006, 17:36-44, FIG. 10.

117. For example, Avendano's system is configured to filter a microphone signal $x_1$ using a "delay node 402," and generate a second processed signal representative of noise (*e.g.,* a "cardioid secondary signal ($C_b$)") by combining the filtered signal with another microphone signal $x_2$. APPLE-1005, 5:15-35, FIG. 4a.

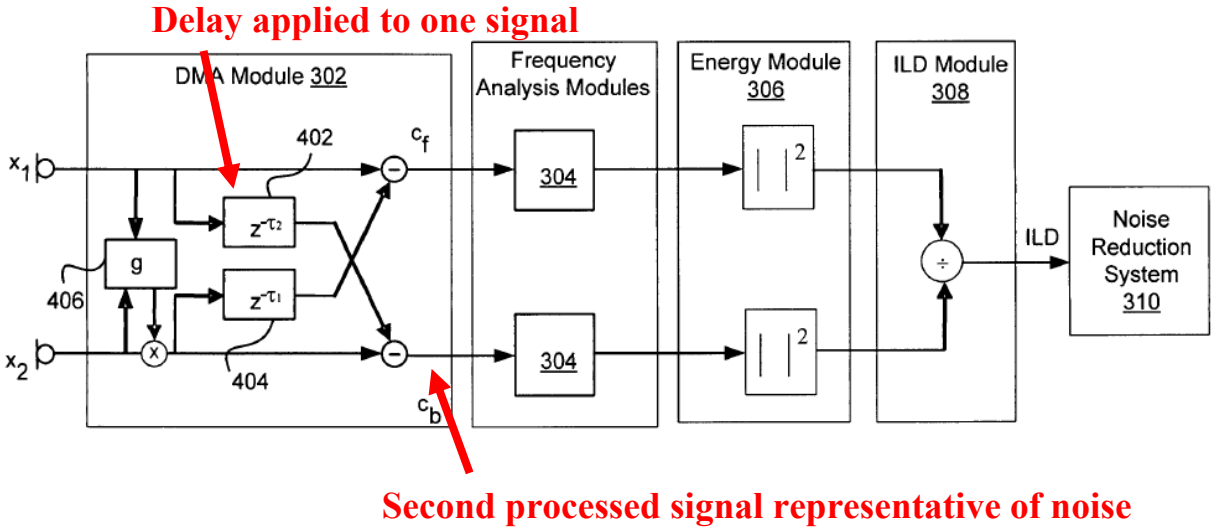
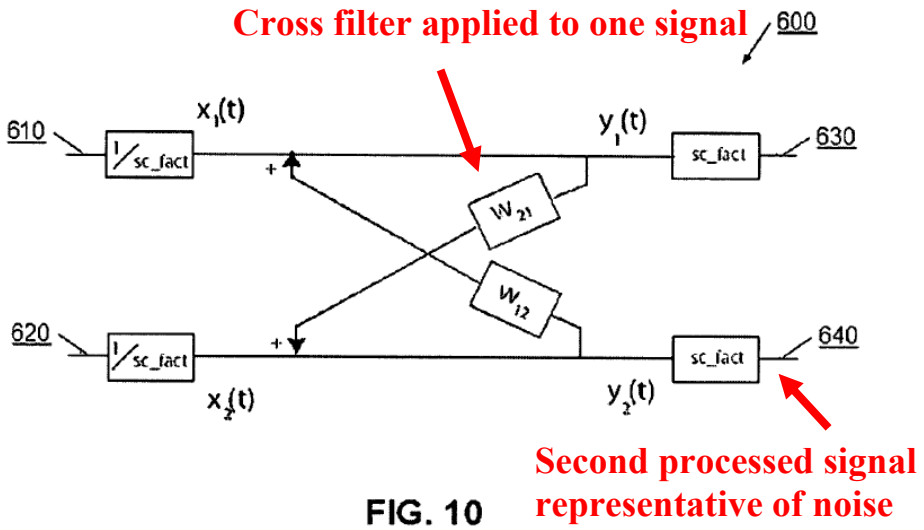



FIG. 4a

**APPLE-1005, FIG. 4a (annotated)**

118. Similarly, a POSITA would have recognized that, like Avendano's "delay node 402," Visser's "cross filter $w_{21}$" is configured to filter one microphone signal (*e.g.*, a "channel 610") in order to generate a second processed signal representative of noise (*e.g.,* a "noise channel, such as the "channel 640") by

combining the filtered signal with another microphone signal (*e.g.,* a "channel 620"). APPLE-1006, 17:36-44, FIG. 10.




**APPLE-1006, FIG. 10 (annotated)**

119. Based on this disclosure, a POSITA would have found it obvious to incorporate Visser's "cross filters $W_{21}$ and $W_{12}$" into Avendano's system to apply additional filtering to Avendano's signals. Such a modification would merely involve using a known technique (*e.g.,* user of Visser's "cross filters") to improve a similar device in a similar way (*e.g.,* to further enhance the separation of speech in Avendano's system). Further, because of the important problem of suppressing environmental noise in microphone signals, a POSITA would have been motivated to achieve the filtering benefits of adding Visser's filters to Avendano's system to provide enhanced separation of speech and noise channels and improved voice activity detection.

120. As an example of applying additional filtering to Avendano's signals, a POSITA would have applied Visser's cross filter $w_{12}$ to Avendano's second signal $x_2$ to further increase speech content in the first virtual microphone $C_f$. Further, a POSITA would have applied Visser's cross filter $w_{21}$ to Avendano's first signal $x_1$ to further decrease speech content in the second virtual microphone $C_b$. An example of this modification is shown below:

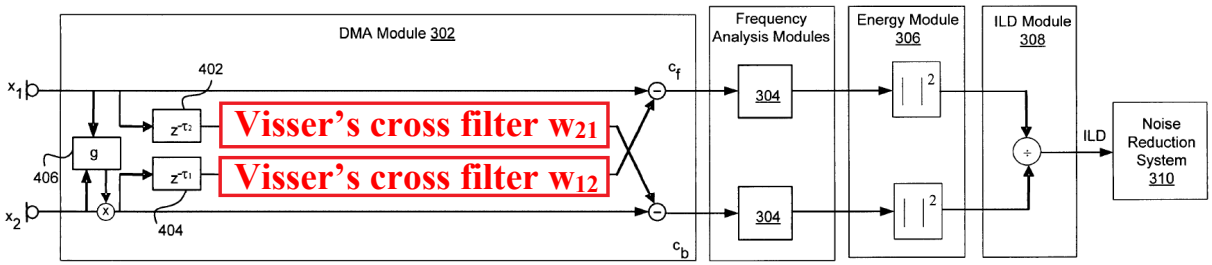



FIG. 4a

**APPLE-1005, FIG. 4a (modified)**

121. Further, a POSITA would have had a reasonable expectation of success in incorporating Visser's "cross filters" into Avendano's system.

122. First, Avendano's "DMA Module 302" and Visser's "cross filters" are both configured to receive similar types of input signals, namely signals from two physical microphones. APPLE-1005, 5:15-6:34, FIG. 4a; APPLE-1006, 17:36-44, FIG. 10.

123. Second, Avendano's "DMA Module 302" and Visser's "cross filters" are both configured to output signals for a similar purpose, namely to produce two processed signals in which one processed signal has a greater degree of speech than the other. *Id.*

124. Given these similarities, a POSITA would understand that Visser's "cross filters" could be incorporated into Avendano's system with only limited modifications. Further, a POSITA would understand that this combination would be compatible with other operations in Avendano's system, as Visser's "cross filters" are configured to receive similar input signals and produce output signals for a similar purpose as Avendano's "DMA Module 302."

125. For these reasons, the combination of Avendano and Visser would have been well within the grasp of a POSITA and a POSITA would have had a reasonable expectation of success in making the combination.

### B. Claim 1

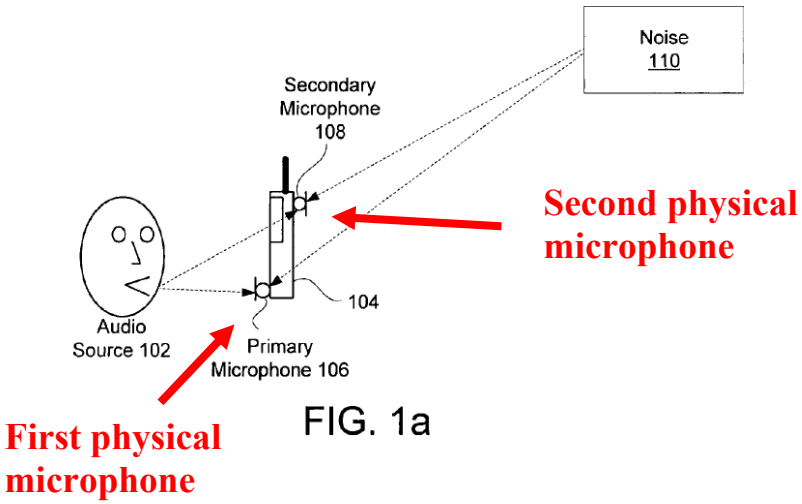
*[1pre]: A method comprising:*

126. To the extent that the preamble of claim 1 is limiting, Avendano discloses a method. APPLE-1005, FIGS. 1a-7, 3:12-7:9, 9:43-10:25 ("an exemplary method for utilizing ILD of omni-direction microphones for noise suppression and speech enhancement.").

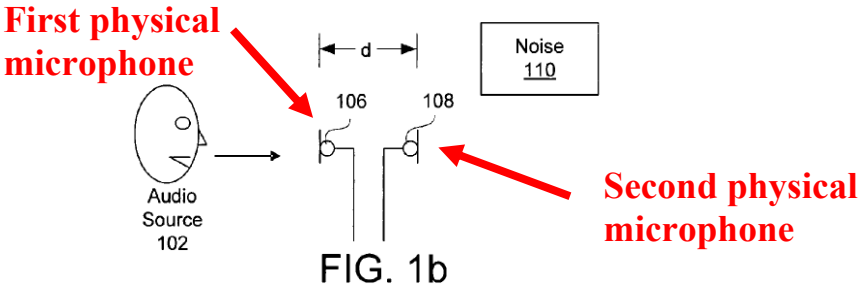
*[1a]: forming a first virtual microphone by combining a first signal of a first physical microphone and a second signal of a second physical microphone*

127. Avendano discloses forming a first virtual microphone by combining a first signal of a first physical microphone and a second signal of a second physical microphone.

128. In Avendano, the first physical microphone is a "primary microphone 106" and the second physical microphone is a "secondary microphone 108." APPLE-1005, 3:27-55, FIGS. 1a-1b.




**APPLE-1005, FIG. 1a (annotated)**



**APPLE-1005, FIG. 1b (annotated)**

129. Further, in Avendano, the first signal is a "primary acoustic signal ($X_1$)" that is output by the "primary microphone 106," and the second signal is a "secondary acoustic signal ($X_2$)" that is output by the "secondary microphone 108." *Id.* 4:20-27, FIGS. 4a-4b.

130. Further, the first virtual microphone is a "cardioid primary signal ($C_f$)," which is formed as a combination of the "primary acoustic signal" (*i.e.,* the first signal) and the "secondary acoustic signal" (*i.e.,* the second signal). *Id.* 4:27-6:10, FIGS. 4a-4b.

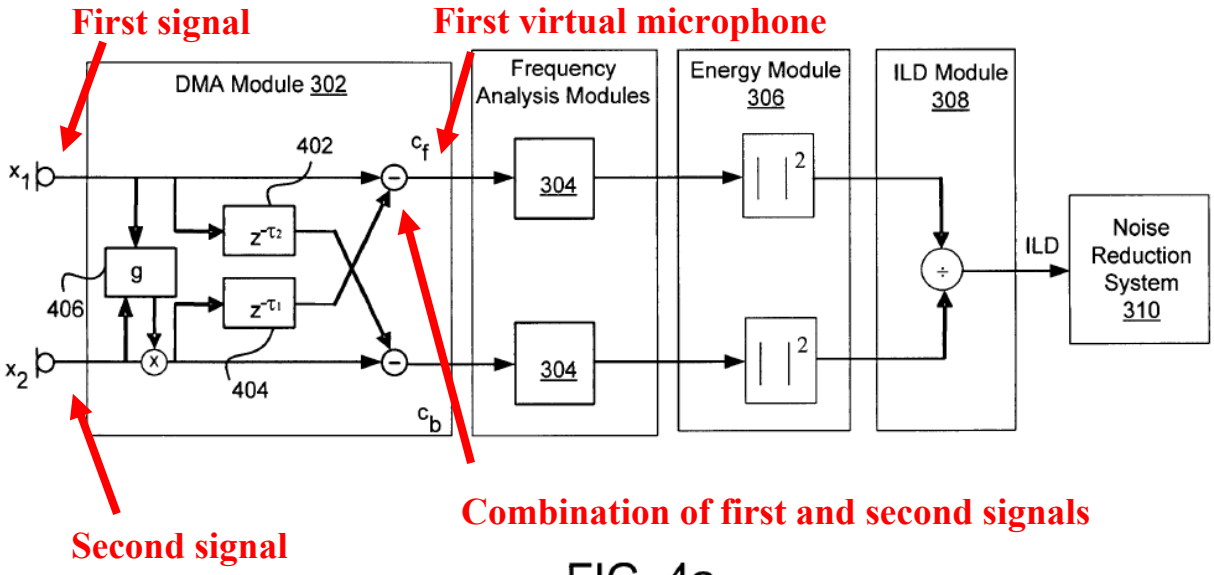



FIG. 4a

**APPLE-1005, FIG. 4a (annotated)**

131. As described in Avendano, "The DMA Module 302 may determine sounds and signals in a front and back cardioid region about the audio device 104 by delaying and subtracting the acoustic signals captured by the microphones 106 and 108… For a two-microphone system, the DMA module 302 can create two different directional patterns about the audio device." *Id.*, 4:27-43.

132. In particular, Avendano discloses an equation for generating the "cardioid primary signal ($C_f$)" from the combination of the "primary acoustic signal" (*i.e.,* the first signal) and the "secondary acoustic signal" (*i.e.,* the second signal):

In exemplary embodiments, a cardioid primary signal ($C_f$) is mathematically determined in the frequency domain (Z transform) as

$$C_f = X_1 - z^{-\tau 1} g X_2$$

**First virtual microphone** **Second signal** **First signal**

**APPLE-1005, 5:25-30 (annotated)**

*[1b]: forming a filter that describes a relationship for speech between the first physical microphone and the second physical microphone, and*

Avendano

133. Avendano describes forming a filter that describes a relationship for speech between the first physical microphone and the second physical microphone.

134. In Avendano, the filter is a "delay node 402." Avendano's "delay node 402" describes a relationship for speech between the first physical microphone and the second physical microphone. APPLE-1005, 4:28-34, 5:19-35, FIGS. 4a-4b.

135. Avendano's system is configured such that a first physical microphone is positioned closer to a source of speech (*e.g.,* an "audio source 102"), and a second physical microphone is positioned farther from the source of speech. *See* Section VI(B), claim element [1a]; APPLE-1005, 3:45-49, FIGS. 1a-1b. As Avendano explains, "due to [this] space difference between the microphones," and because the speed of sound propagation in air is widely known to be approximately 330 m/s, there will be a small, but finite "difference in times of arrival of the signal from a speech source to the microphones." *Id.*, 1:33-36. Thus, the "relationship for speech" between Avendano's two physical microphones is, at least in part, that the arrival of speech at one physical microphone (*i.e.,* the physical microphone positioned at the rear of the audio device) will be slightly delayed in time relative to the arrival of speech at the other physical microphone (*i.e.,* the physical microphone positioned in the front of the audio device).

136. Avendano's "delay node 402" is configured to delay the first signal, which would otherwise include speech content before the second signal. APPLE-1005, 5:15-35. Accordingly, Avendano's "delay node 402" describes, at least in part, a *temporal* relationship for speech between the first signal and the second signal.

137. Further, Avendano describes that a delay node can be implemented using one or more filters. APPLE-1005, 8-38-51. For example, as explained by Avendano, "[t]o implement a fractional delay, allpass filters 416 and 418 ... are applied to the signals." *Id.*

138. This also is consistent with the specification of the '611 patent, which describes that a time delay can be implemented using a "delay filter." APPLE-1001, FIG. 3. For example, as shown in FIG. 3, of the '611, a delay is implemented using a "Delay filter $z^{-\gamma}$":

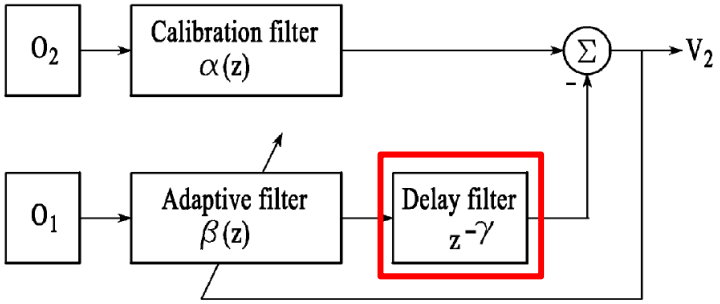



FIG.3

**APPLE-1001, FIG. 3 (annotated)**

Avendano and Visser

139. To the extent that Avendano alone does not describe forming a filter that describes a relationship for speech between the first physical microphone and the second physical microphone, it would have been obvious to modify Avendano's system in view of Visser to include such a feature.

140. As I explain in Section VII(A) (incorporated herein), it would have been obvious to incorporate Visser's "ICA or BSS processing function" into Avendano's system. According to this modification, Avendano's system would include Visser's cross filters $w_{12}$ and $w_{21}$. Each of Visser's cross filters $w_{12}$ and $w_{21}$ describes a relationship for speech between a first physical microphone and a second physical microphone.

141. Also, as I explain in Section VII(A), according to one example modification, Visser's cross filter $w_{12}$ would have been applied to Avendano's second signal $x_2$ to further increase speech content in the first virtual microphone $C_f$. Further, Visser's cross filter $w_{21}$ would have been applied to Avendano's first signal $x_1$ to further decrease speech content in the second virtual microphone $C_b$).

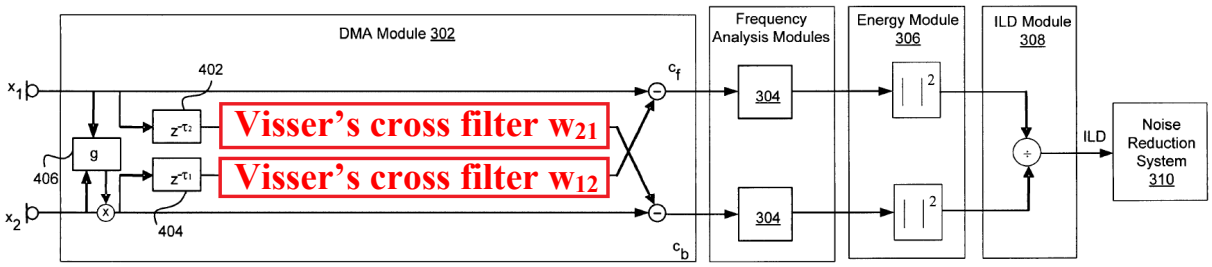



FIG. 4a

**APPLE-1005, FIG. 4a (modified)**

142. Further, Visser discloses that different physical microphones may have different responses to speech, depending on the location of the audio source

relative to each physical microphone. APPLE-1006, 15:43-16:2. As Visser, explains, "[s]ince each of the microphones is separate from every other microphone, each microphone will generate a somewhat different composite signal. For example, the relative content of noise and speech may vary, as well as the timing and delay for each sound source." *Id.*

143. Visser teaches that a system can leverage the different responses of the physical microphones to speech in order to distinguish speech from noise. *Id.*, 8:16-18, 16:3-28, 17:29-44, FIG. 10. For example, Visser discloses generating cross filters $w_{12}$ and $w_{21}$ using independent component analysis (ICA) to form a "speech channel 630" (which contains predominately speech) and a "noise channel 640" (which contains predominately noise). *Id.* As Visser explains, "[the] ICA process filters the received composite signals using cross filters, which are preferably infinitive impulse response filters with nonlinear bounded functions." *Id.,* 16:8-11. Further, Visser's filters can be "adapt[ed] during operation," such that the filters can better separate speech from noise based on the signals received from the physical microphones. *Id.*, 21:17-18:44, FIG. 12. In particular, the "speech separation process ... may be adaptive and learn according to the specific acoustic environment," and to "adapt to particular microphone placement, the acoustic environment, or a particular user's speech." *Id.*, 9:8-12.

144. Accordingly, each of Visser's cross filters $w_{12}$ and $w_{21}$ describes a relationship for speech between at least a first physical microphone and a second physical microphone in order to distinguish speech from noise.

*[1c]: forming a second virtual microphone by applying the filter to the first signal to generate a first intermediate signal, and summing the first intermediate signal and the second signal*

<u>Avendano</u>

145. Avendano discloses forming a second virtual microphone by applying the filter to the first signal to generate a first intermediate signal, and summing the first intermediate signal and the second signal.

146. In Avendano, the first intermediate signal is an output of the "delay node 402." APPLE-1005, 5:15-35, FIGS. 4a-4b.

147. The output of the "delay node 402" is generated by applying the "delay node 402" (*i.e.,* a filter) to the first signal. *Id.*

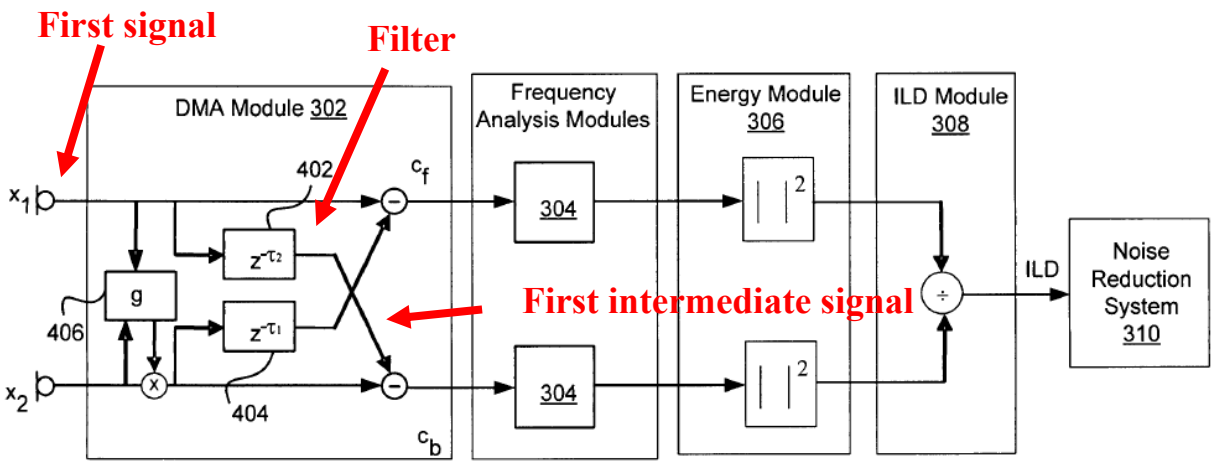



FIG. 4a

**APPLE-1005, FIG. 4a (annotated)**

148. Further, in Avendano, a second virtual microphone is a "cardioid secondary signal ($C_b$)." *Id.* The "cardioid secondary signal ($C_b$)" is formed by subtracting the first intermediate signal from the second signal. *Id.*

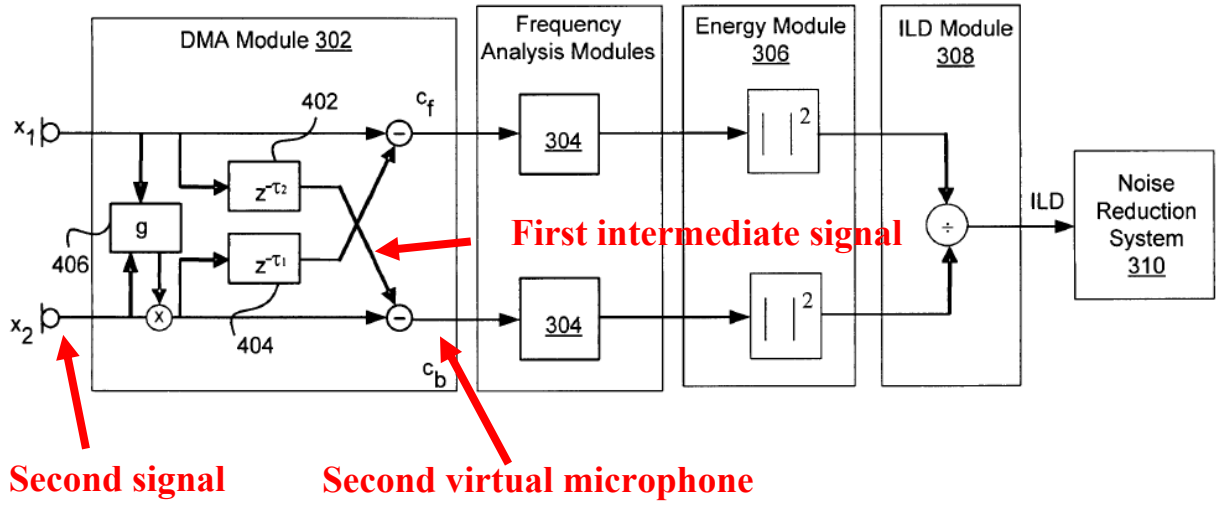



FIG. 4a

**APPLE-1005, FIG. 4a (annotated)**

149. As Avendano explains, a first signal can be subtracted from a second signal by summing the second signal and an inverse of the first signal (*e.g.*, a negative value of the first signal). *Id.,* FIG. 4b. For example, as shown in Avendano's FIG. 4b, the operation of subtracting the first intermediate signal from the second signal can be performed by summing (i) the second signal $x_2$, and (ii) the inverse of the first intermediate signal. *Id.* This operation is represented using a summation node and a negative sign by one of the inputs to the summation node. *Id.*

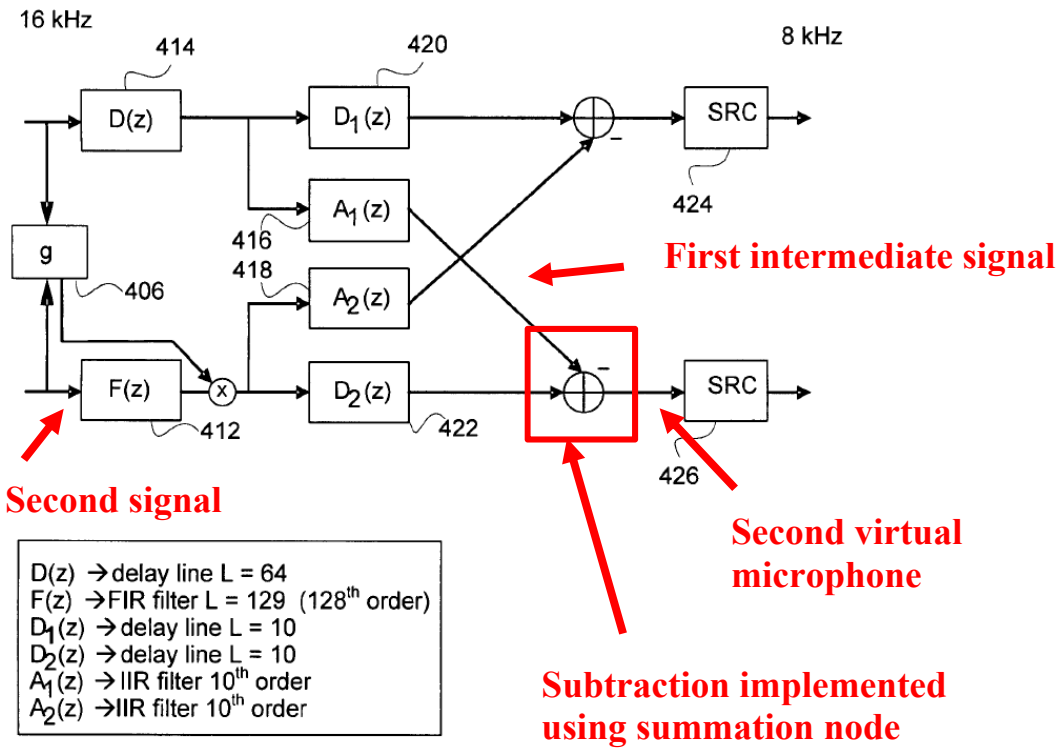



FIG. 4b

**APPLE-1005, FIG. 4b (annotated)**

150. This is consistent with the specification of the '611 patent, which describes that "processing paths" can be "summed to form virtual microphones," and that "varying the magnitude and sign of the delays and gains of the processing paths leads to a wide variety of virtual microphones." APPLE-1001, 21:26-52 (emphasis added), FIGS. 25-26.

151. Further, the '611 patent indicates that a summation can include summing a first value with the inverse of a second value to form a second virtual microphone $V_2$. APPLE-1005, FIG. 3. For example, as shown in FIG. 3, the '611 describes

that a summation (signified by "Σ") can include summing a first value with the inverse of a second value (signified using a negative sign by one of the inputs to "Σ") to form a second virtual microphone $V_2$. *Id.* This summation is the mathematical equivalent of subtracting the second value from the first value.

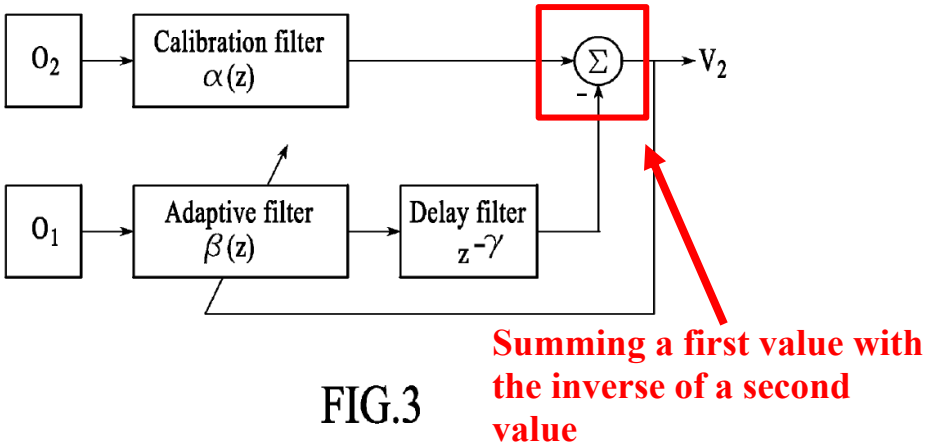



FIG.3

**APPLE-1001, FIG. 3 (annotated)**

152. Accordingly, as in the '611 patent, Avendano discloses forming a second virtual microphone by applying the filter to the first signal to generate a first intermediate signal, and summing the first intermediate signal and the second signal.

Avendano and Visser

153. To the extent that Avendano alone does not describe forming a second virtual microphone by applying the filter to the first signal to generate a first intermediate signal, and summing the first intermediate signal and the second

signal, it would have been obvious to modify Avendano's system in view of Visser to include such a feature.

154. As I explain in Section VII(A) (incorporated herein), it would have been obvious to incorporate Visser's "ICA or BSS processing function" into Avendano's system to further enhance the separation between speech and noise.

155. First, Visser and Avendano each describe determining the presence of speech using similar principles.

156. For example, Visser discloses generating two processed signals, including (i) a "speech channel" that "contains predominantly desired signals," and (ii) a "noise channel" that "contains predominantly noise signals." APPLE-1006, 17:36-44. This is analogous to Avendano's technique, in which two processed signals are generated, including (i) a "cardioid primary signal ($C_f$)" and (ii) a "cardioid secondary signal ($C_b$)," where the "cardioid primary signal ($C_f$)" is expected to have a greater response to speech than the "cardioid secondary signal ($C_b$)." APPLE-1005, 4:47-52, 5:15-35.

157. Further, like Avendano, Visser's processed signals can be generated by applying a delay to a signal received from one of the physical microphones, and combining the delayed signal with a signal received from the other physical microphone. APPLE-1005, 5:15-35; APPLE-1006, 17:39-62, FIG. 10.

158. Second, Visser additionally describes that the separation between the "speech channel" and the "noise channel" can be further enhanced by generating and adapting the filters according to an "ICA process." *Id.,* 9:20-62, 17:29-25:9, FIGS. 10-13. In particular, Visser discloses filtering each of two microphones using "cross filters," prior to combining each of the filtered microphone signals with the other microphone signal. *Id.,* 17:35-44, FIG. 10. As taught by Visser, such a process would enable a system to adapt to "the specific acoustic environment," including "[a] particular microphone placement, the acoustic environment, or a particular user's speech," in order "to provide a sufficient level of separation, even in a changing acoustic environment." *Id.,* 9:8-12, 16:65-17:4.

159. A POSITA would have found it obvious to implement Visser's cross-filters in Avendano's system to further enhance the separation between speech and noise in Avendano's "cardioid primary signal ($C_f$)" and "cardioid secondary signal ($C_b$)." For instance, a POSITA would have found it obvious to use Visser's cross-filters to filter each of Avendano's two microphone signals, prior to combining each of the filtered microphone signals with the other microphone signal.

160. As discussed in Section VII(A), according to one example modification, Visser's cross filter $w_{12}$ would have been applied to Avendano's second signal $x_2$ to further increase speech content in the first virtual microphone $C_f$. Further,

Visser's cross filter $w_{21}$ would have been applied to Avendano's first signal $x_1$ to further decrease speech content in the second virtual microphone $C_b$.

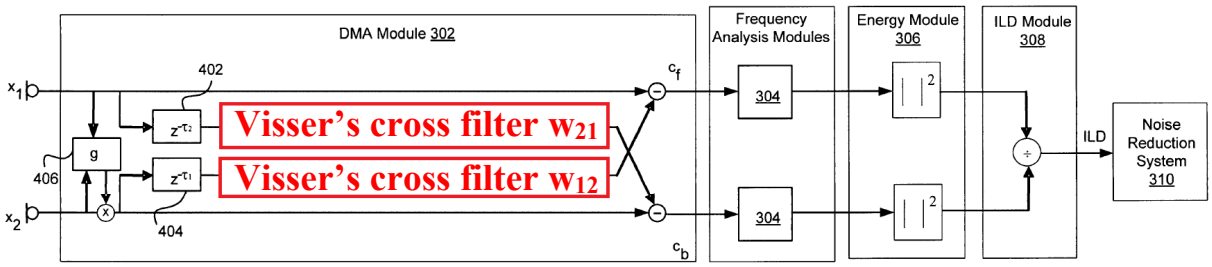



FIG. 4a

**APPLE-1005, FIG. 4a (modified)**

161. According to the modification, the first intermediate signal is an output of the cross filter $w_{21}$. This first intermediate signal is generated by applying the cross filter $w_{21}$ to a first signal $x_1$.

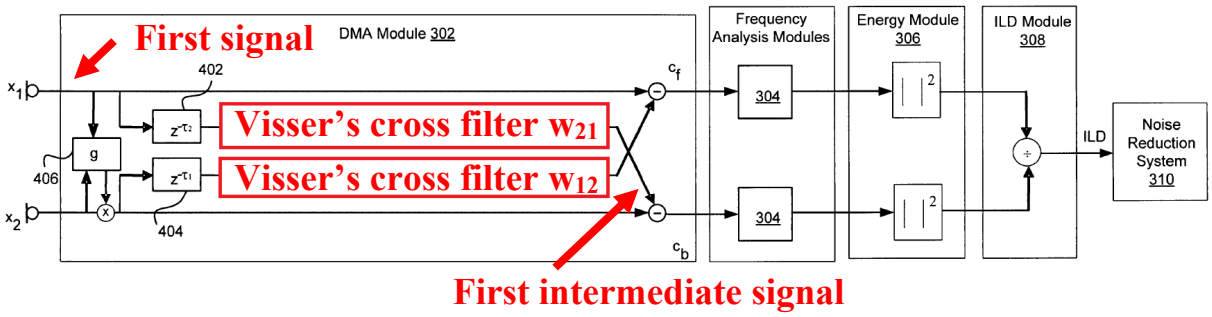
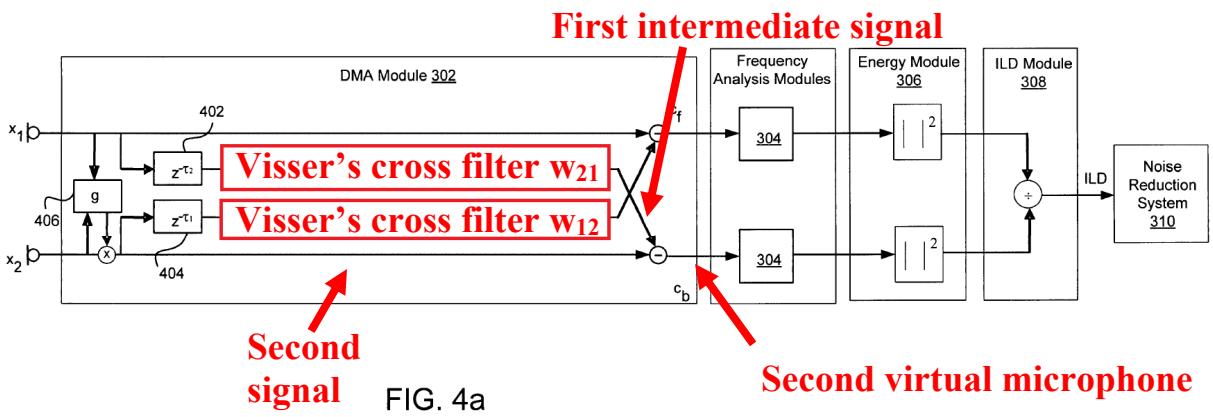



FIG. 4a

**APPLE-1005, FIG. 4a (modified)**

162. Further, according to the modification, a second virtual microphone is a "cardioid secondary signal ($C_b$)." This second virtual microphone is formed by

subtracting the first intermediate signal from a second signal $x_2$. APPLE-1005, FIG. 4a.




**APPLE-1005, FIG. 4a (modified)**

*[1d]: generating an energy ratio of energies of the first virtual microphone and the second virtual microphone*

163. Avendano discloses generating an energy ratio of energies of the first virtual microphone and the second virtual microphone.

164. In Avendano, the energy of the first virtual microphone is $E_f$. The energy $E_f$ is calculated by integrating the square of the signal from the first virtual microphone over time. APPLE-1005, 5:49-61.

$$E_f(t, \omega) = \int_{frame} |C_f(t', \omega)|^2 \, dt',$$

**Energy of the first virtual microphone** **First virtual microphone**

**APPLE-1005, 5:61 (annotated)**

165. The energy of the second virtual microphone is $E_b$. The energy $E_b$ is calculated by integrating the square of the signal from the second virtual microphone over time. *Id.*, 5:49-6:5.

$$E_b(t, \omega) = \int_{frame} |C_b(t', \omega)|^2 \, dt'.$$

**Energy of the second virtual microphone**

**Second virtual microphone**

**APPLE-1005, 6:5 (annotated)**

166. The energy ratio is "Inter-Level Difference" (ILD). The energy ratio ILD is determined by dividing the energy of the first virtual microphone by the energy of the second virtual microphone. *Id.*, 6:8-34.

$$ILD(t, \omega) = \frac{\int |C_f(t', \omega)|^2 \, dt'}{\int_{frame} |C_b(t', \omega)|^2 \, dt'}.$$

**Energy of the first virtual microphone**

**Energy of the second virtual microphone**

**APPLE-1005, 6:16 (annotated)**

*[1e]: detecting acoustic voice activity of a speaker when the energy ratio is greater than a threshold value*

167. Avendano discloses detecting acoustic voice activity of a speaker when the energy ratio is greater than a threshold value.

168. For example, Avendano describes comparing the energy ratio (ILD) to a "threshold" to determine a value $\lambda_1$. APPLE-1005, 6:35-7:9.

$$\lambda_I(t, \omega) = \begin{cases} \approx 0 & \text{if} \quad ILD(t, \omega) < \text{threshold} \\ \approx 1 & \text{if} \quad ILD(t, \omega) > \text{threshold} \end{cases}$$

**APPLE-1005, 6:61**

169. The value of the energy ratio (ILD) relative to the "threshold" is indicative the presence of absence of speech. *Id.*, 6:58-7:3.

170. For example, if the energy ratio (ILD) is "smaller than a threshold value (e.g., threshold=0.5) above which speech is expected to be, $\lambda_1$ is small." *Id.,* 6:66-7:1. Specifically, when the ratio energy (ILD) is less than the "threshold," $\lambda_1$ is set to one. *Id,* 6:61. This is indicative of an absence of speech. *Id.,* 6:58-7:3.

171. In contrast, if the ratio energy (ILD) "starts to rise (e.g., because speech is present within the large ILD region), $\lambda_1$ increases." *Id.*, 7:2-4. Specifically, when the ratio energy (ILD) is greater than the "threshold," $\lambda_1$ is set to one. *Id.*, 6:61. This is indicative of a presence of speech. *Id.,* 6:58-7:3.

### C. Claim 2

*[2] The method of claim 1, wherein the first virtual microphone and the second virtual microphone are distinct virtual directional microphones.*

172. Avendano (and in the combination with Visser) discloses that the first virtual microphone and the second virtual microphone are distinct virtual directional microphones.

173. As I explain in Section VII(B) (incorporated herein), and specifically in reference to claim elements [1a] and [1c], Avendano discloses a first virtual microphone (*e.g.,* a "cardioid primary signal ($C_f$)) and a second virtual microphone (*e.g.*, a "cardioid secondary signal ($C_b$)).

174. As described in Avendano, "The DMA Module 302 may determine sounds and signals in a front and back cardioid region about the audio device 104 by delaying and subtracting the acoustic signals captured by the microphones 106 and 108… For a two-microphone system, the DMA module 302 can create two different directional patterns about the audio device." APPLE-1005, 4:27-5.

175. In Avendano (and in the combination with Visser), the first virtual microphone and the second virtual microphone are distinct from one another. *Id.*, FIG. 4a, 5:15-35.

### D. Claim 3

*[3] The method of claim 2, wherein the first virtual microphone and the second virtual microphone have approximately similar responses to noise.*

176. Avendano (and in the combination with Visser) discloses that the first virtual microphone and the second virtual microphone have approximately similar responses to noise.

177. As I explain in Section VII(B) (incorporated herein), and specifically in reference to claim elements [1a] and [1c], Avendano discloses generating a first virtual microphone (e.g., "cardioid primary signal ($C_f$)") and a second virtual microphone ("cardioid secondary signal ($C_b$)"). The first virtual microphone has a directional pattern directed to the front of the "audio device 104. APPLE-1005, 4:47-52, 5:25-35, FIG. 6. The second virtual microphone has a directional pattern directed to the back of the "audio device 104." *Id.*

178. As Avendano explains, noise is typically generated in the "background" or "far field," rather than from a location near the audio device. *Id.,* 2:49, FIG. 1a. In addition, Avendano describes that noise can include "any sounds from one or more locations different than the audio source 102, and may include reverberations and echoes." *Id.,* 3:35-41.

179. A POSITA would have recognized that, according to Avendano's configuration, Avendano's virtual microphones would have approximately similar responses to noise.

180. For example, Avendano describes two virtual microphones having respective directional patterns. *Id.*, 4:47-52, 5:25-35, FIG. 6. However, neither of these virtual microphones is directed to a "background" or "far field." *Id.* Instead, each of these virtual microphones is directed to a region in proximity to the audio device, such as immediately in front of the audio device or immediately behind the audio device. *Id.,* 2:49, FIG. 1a. As such, neither virtual microphone would be more sensitive to noise generated in the "background" or "far field" than the other.

181. Further, Avendano explains that an audio device would detect both (i) noise propagating directly from a noise source to the audio device, and (ii) reverberations and echoes of that noise from multiple other locations relative to the audio device. APPLE-1005*,* 3:35-41. Due to the non-directionally specific nature of these reverberations and echoes, neither virtual microphone would be more sensitive to detecting noise (and the reverberations and echoes thereof) than the other.

182. Accordingly, each of Avendano's two virtual microphones would have approximately similar responses to noise.

### E. Claim 4

*[4] The method of claim 3, wherein the first virtual microphone and the second virtual microphone have approximately dissimilar responses to speech.*

183. Avendano (and in the combination with Visser) discloses that the first virtual microphone and the second virtual microphone have approximately dissimilar responses to speech.

184. As I explain in Section VII(B) (incorporated herein), and specifically in reference to claim elements [1a] and [1c], Avendano discloses generating a first virtual microphone ( "cardioid primary signal ($C_f$)") and a second virtual microphone ("cardioid secondary signal ($C_b$)"). The first virtual microphone has a directional pattern directed to the front of the "audio device 104." APPLE-1005, 4:47-52, 5:25-35, FIG. 6. The second virtual microphone has a directional pattern directed to the back of the "audio device 104." *Id.*

185. As Avendano explains, speech is typically generated near the front of the device. *Id.*, 3: 42-55, FIG. 1a.

186. A POSITA would have recognized that, according to Avendano's configuration, Avendano's virtual microphones would have approximately dissimilar responses to speech.

187. For example, the directional pattern of the first virtual microphone is specifically directed to the front of the device (*i.e.,* towards the source of speech).

In contrast, the directional pattern of the first virtual microphone is specifically directed to the rear of the device (*i.e.,* away from the source of speech). Accordingly, Avendano's first virtual microphone would have a greater response to speech than that of the second virtual microphone.

### F. Claim 5

*[5] The method of claim 1, comprising applying a calibration to at least one of the first signal and the second signal.*

188. Avendano (and in the combination with Visser) discloses applying a calibration to at least one of the first signal and the second signal.

189. As I explain in Section VII(B) (incorporated herein), and specifically in reference to claim element [1a], Avendano discloses a first signal (e.g., "first acoustic signal ($X_1$)") and a second signal (*e.g.*, "secondary acoustic signal ($X_2$)").

190. Further, Avendano describes applying a calibration to at least the second signal, specifically "gain factor, g." APPLE-1005, 5:36-39, 9:54-59, FIGS. 4a-4b.

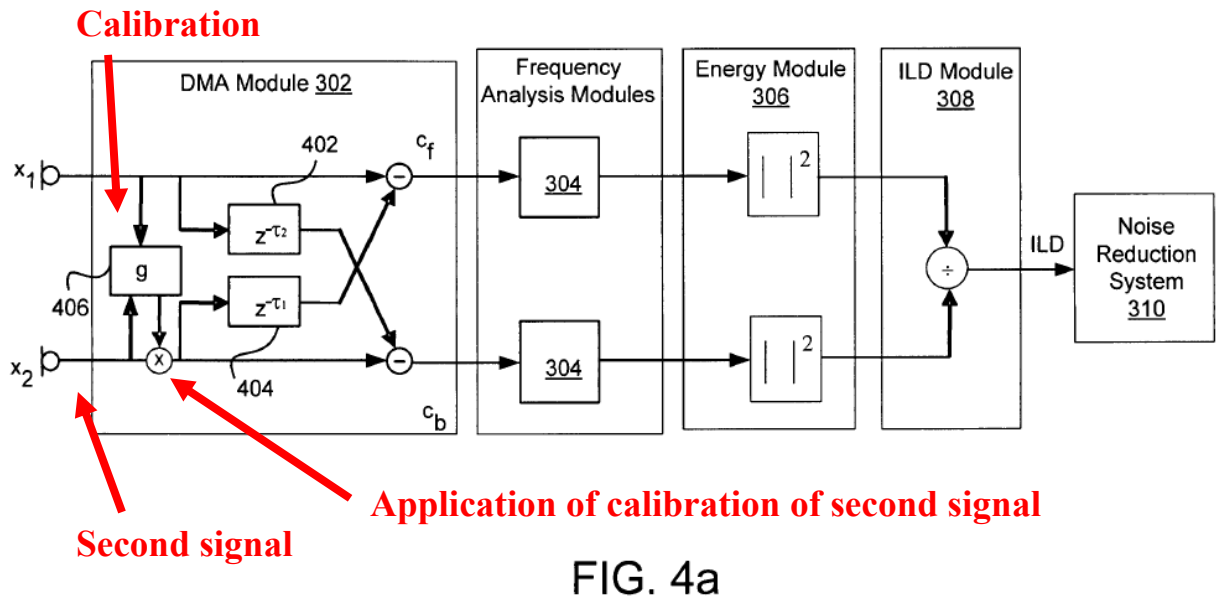



FIG. 4a

**APPLE-1005, FIG. 4a (annotated)**

191. In Avendano's system, "[t]he gain factor, g, is computed by the gain module 406 to equalize the signal levels" of the first and second signals. *Id.*, 5:36-37.

192. That is, the gain factor, g, calibrates the signal levels of the first and second signals. *Id.*

### G. Claim 6

*[6] The method of claim 5, wherein the calibration compensates a second response of the second physical microphone so that the second response is equivalent to a first response of the first physical microphone.*

193. Avendano (and in the combination with Visser) discloses that the calibration compensates a second response of the second physical microphone so that the second response is equivalent to a first response of the first physical microphone.

194. As I explain in Section VII(F) (incorporated herein), Avendano describes applying a calibration (*e.g.,* a "gain factor, g") to at least the second signal.

195. Further, Avendano describes that "[t]he gain factor, g, is computed by the gain module 406 to equalize the signal levels" of the first and second signals. APPLE-1005, 5:36-37. Accordingly, Avendano's calibration that is applied to at least the second signal compensates a second response of the second physical microphone so that the second response is equivalent to a first response of the first physical microphone.

### H. Claim 7

*[7] The method of claim 5, comprising applying a delay to the first intermediate signal.*

196. Avendano (and in the combination with Visser) discloses applying a delay to the first intermediate signal.

197. As I explain in Section VII(B) (incorporated herein), and specifically in reference to claim element [1c], Avendano discloses a first intermediate signal. Specifically, the first intermediate signal is an output of the "delay node 402,"

which is generated by applying the "delay node 402" (*i.e.,* a filter) to the first signal. APPLE-1005, 5:15-35, FIGS. 4a-4b.

198. That is, in Avendano's system, a delay has been applied to the first intermediate signal. *Id.*

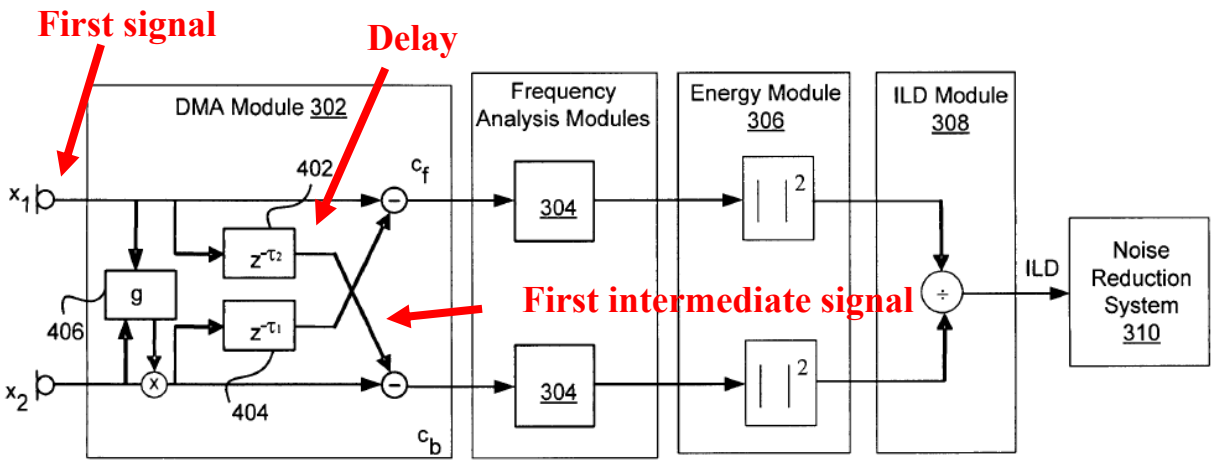



FIG. 4a

**APPLE-1005, FIG. 4a (annotated)**

199. Accordingly, in Avendano, a delay has been applied to the first intermediate signal.

200. Further, as I also explain with reference to claim element [1c], it would have been obvious to incorporate Visser's "ICA or BSS processing function" into Avendano's system. According to one example modification, Visser's cross filter

$w_{12}$ would have been applied to Avendano's second signal $x_2$, and Visser's cross filter $w_{21}$ would have been applied to Avendano's first signal $x_1$.

201. Visser's "cross filters $w_{12}$ and $w_{21}$" can be "gain factors with only one filter coefficient per filter, for example a <u>delay</u> gain factor for the time delay between the output signal and the feedback input signal and an amplitude gain factor for amplifying the input signal." APPLE-1006, 17:58-62 (emphasis added). Accordingly, in the combination of Avendano and Visser, a delay also would have been applied to the first intermediate signal by Visser's cross filter $w_{21}$.

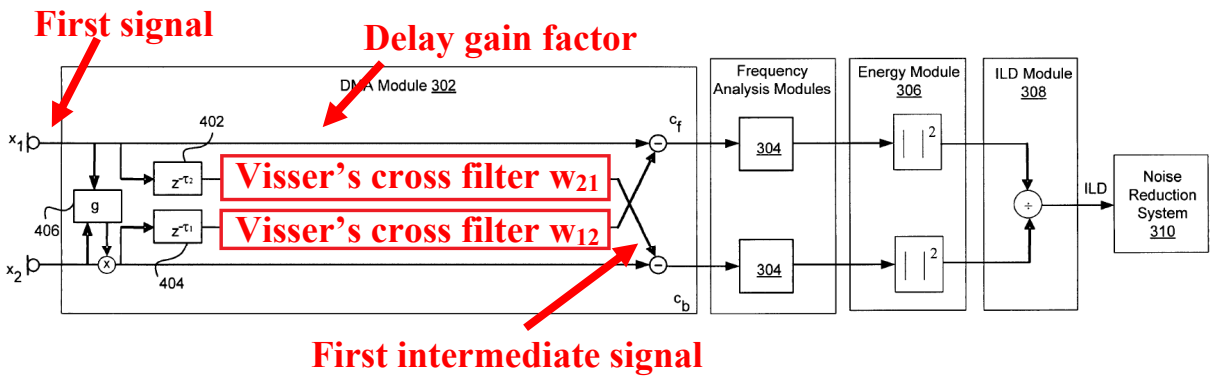



FIG. 4a

**APPLE-1005, FIG. 4a (modified)**

## I. Claim 25

*[25] The method of claim 1, comprising generating a vector of the energy ratio versus time.*

202. Avendano (and in the combination with Visser) discloses generating a vector of the energy ratio versus time.

203. As I explain in Section VII(B) (incorporated herein), and specifically in reference to claim element [1d], Avendano discloses generating an energy ratio (*e.g.,* ILD).

204. As shown in the equation below, Avendano discloses that the energy ratio (ILD) can be generated as a vector over time (t). APPLE-1005, 6:8-34.

**Energy ratio**

$$\boxed{ILD(t,\omega)} = \frac{\int |C_f(t',\omega)|^2\, dt'}{\int_{frame} |C_b(t',\omega)|^2\, dt'}.$$

**APPLE-1005, 6:16 (annotated)**

### J. Claim 26

*[26] The method of claim 1, wherein the first and second physical microphones are omnidirectional microphones.*

205. Avendano (and in the combination with Visser) discloses that the first and second physical microphones are omnidirectional microphones.

206. As I explain in Section VII(B) (incorporated herein), and specifically in reference to claim element [1a], Avendano discloses a first physical microphone (*e.g.,* "primary microphone 106") and a second physical microphone (*e.g.,* "secondary microphone 108").

207. Avendano discloses that each of these physical microphones may be an "omni-directional microphone." APPLE-1005, Abstract, 2:3-19, 3:33-35, 9:39-52.

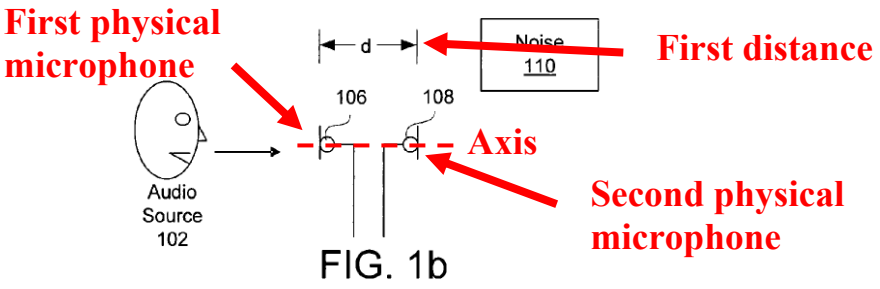
### K. Claim 27

*[27] The method of claim 1, comprising positioning the first physical microphone and the second physical microphone along an axis and separating the first physical microphone and the second physical microphone by a first distance.*

208. Avendano (and in the combination with Visser) discloses positioning the first physical microphone and the second physical microphone along an axis and separating the first physical microphone and the second physical microphone by a first distance.

209. As I explain in Section VII(B) (incorporated herein), and specifically in reference to claim element [1a], Avendano discloses a first physical microphone (*e.g.,* "primary microphone 106") and a second physical microphone (*e.g.,* "secondary microphone 108").

210. As shown in Avendano's FIG. 1b, Avendano's first and second physical microphones are positioned along an axis, and separated by a first distance (d). APPLE-1005, FIG. 1b.




**APPLE-1005, FIG. 1b (annotated)**

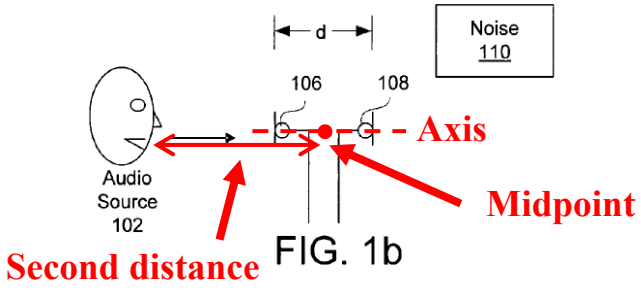
### L. Claim 28

*[28] The method of claim 27, wherein a midpoint of the axis is a second distance from a mouth of the speaker, wherein the mouth is located in a direction defined by an angle relative to the midpoint.*

211. Avendano (and in the combination with Visser) discloses that a midpoint of the axis is a second distance from a mouth of the speaker, wherein the mouth is located in a direction defined by an angle relative to the midpoint.

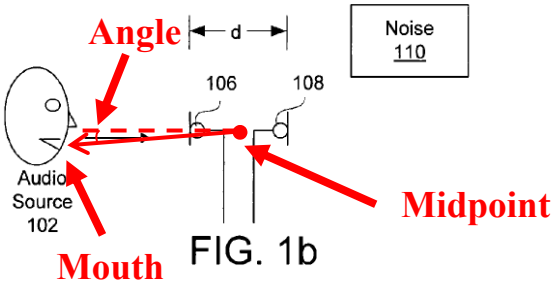
212. As I explain in Section VII(K) (incorporated herein), Avendano discloses an axis along which a first physical microphone and a second physical microphone are positioned. APPLE-1005, FIG. 1b.

213. Further, as I illustrate below, a midpoint of the axis is a second distance away from a mouth of a speaker. *Id.*




**APPLE-1005, FIG. 1b (annotated)**

214. Further, as I illustrate below, the mouth of the speaker is located in a direction defined by an angle relative to the midpoint. *Id.*




**APPLE-1005, FIG. 1b (annotated)**

## VIII. GROUND 2: CLAIMS 8-16, 23, AND 24 ARE RENDERED OBVIOUS BY AVENDANO IN VIEW OF VISSER AND BISGAARD

### A. Combination of Avendano, Visser, and Bisgaard

215. As I explain in Section VII(A) (incorporated herein), a POSITA would have found it obvious to combine Avendano and Visser.

216. Further, as I detail below, a POSITA would have found it obvious to further combine Avendano and Visser with Bisgaard.

217. First, both of Avendano's system and Bisgaard's system are configured to receive signals from two microphones, and process the signals to obtain (i) a cardioid signal directed in a forward direction (*e.g.*, towards a user or other audio source), and (ii) another cardioid signal directed in a backwards direction (*e.g.*, away from the user or other audio source). APPLE-1005, 4:42-5:35, FIGS. 4a-4b, 6; APPLE-1010, [0041]-[0042], FIG. 2; APPLE-1011, 5:18-42, FIG. 2.

218. Further, both of Avendano's system and Bisgaard's system are configured to form each of the cardioid signals by delaying the signal from one microphone, and

combining it with the signal from the other microphone. APPLE-1005, 4:42-5:35, FIGS. 4a-4b, 6; APPLE-1010, [0041]-[0042], FIG. 2; APPLE-1011, 5:18-42, FIG. 2.

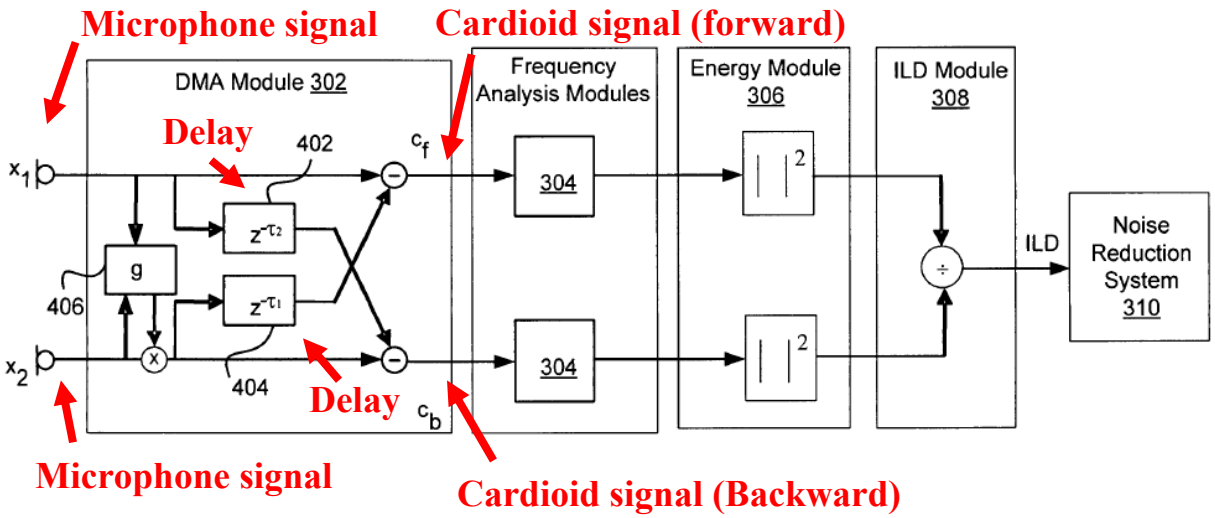



FIG. 4a

**APPLE-1005, FIG. 4a (annotated)**

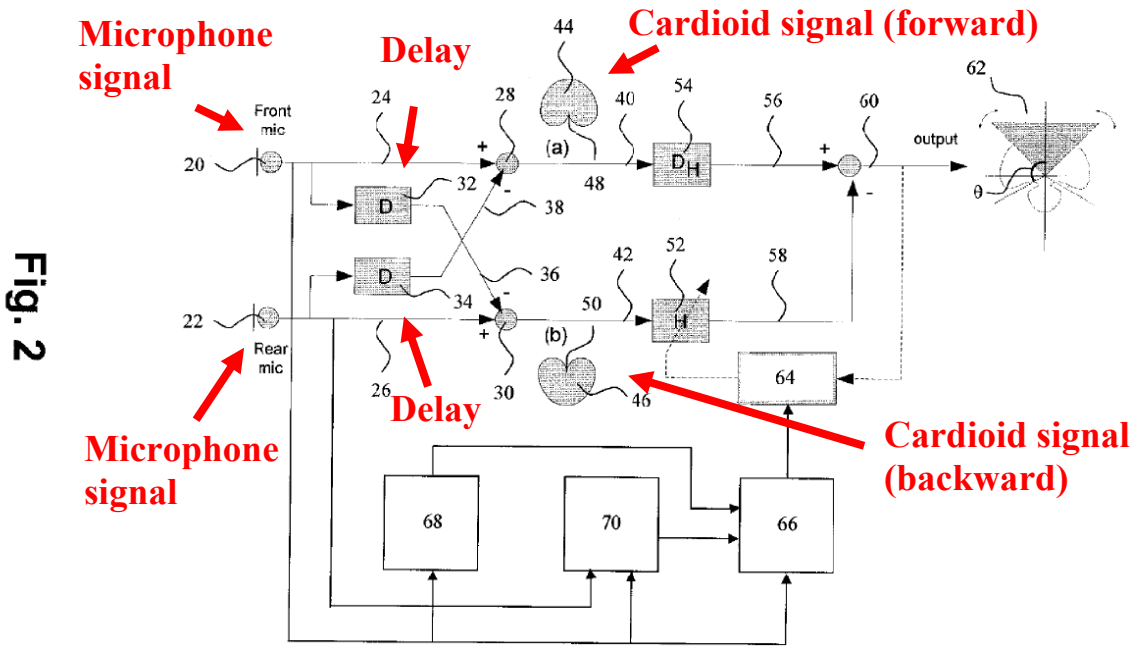



Fig. 2

**APPLE-1010, FIG. 2 (annotated)**

219. A POSITA would have understood that Avendano's "delay nodes" and Bisgaard's "delays" serve the same purpose, namely delaying signals from physical microphones to produce directional cardioid signals (*e.g.,* by combining the delayed microphone signal with another microphone signal). *Id.* Therefore, a POSITA would have looked to Bisgaard for details related to implementing such a delay in Avendano's system.

220. As an example of applying Bisgaard's teaching regarding delays to Avendano's system, a POSITA would have found it obvious to configure each of Avendano's "delay nodes" to "delay[] the digitized sound signal" received by each

of Avendano's physical microphones "by the amount of time used by a sound signal to propagate in the 0° azimuth direction from the front microphone 20 to the rear microphone 22," as taught by Bisgaard, in order to account for "the amount of time used by a sound signal to propagate in the 0° azimuth direction from [a] front microphone ... to [a] rear microphone." APPLE-1010, [0042]; *see also* APPLE-1011, 5:22-24.

221. Further, a POSITA would have had a reasonable expectation of success in incorporating Bisgaard's "delays" in Avendano's system, due to the similarities between Avendano's "delay nodes" and Bisgaard's "delays." For example, as I discussed above, Avendano's "delay nodes" and Bisgaard's "delays" are configured to receive similar input signals and generate similar outputs. APPLE-1005, 4:42-5:35, FIGS. 4a-4b, 6; APPLE-1010, [0041]-[0042], FIG. 2; APPLE-1011, 5:18-42, FIG. 2. Given these similarities, limited modifications would have been necessary to incorporate Bisgaard's "delays" into Avendano's system. Further, other operations in Avendano's system would not have been impacted by the combination, as Bisgaard's "delays" are configured to receive similar input signals and generate similar outputs as Avendano's "delay nodes."

222. For these reasons, the combination of Avendano, Visser, and Bisgaard would have been well within the grasp of a POSITA and a POSITA would have had a reasonable expectation of success in making the combination.

### B. Claim 8

*[8] The method of claim 7, wherein the delay is proportional to a time difference between arrival of the speech at the second physical microphone and arrival of the speech at the first physical microphone.*

223. Avendano (and in the combination with Visser) discloses that the delay is proportional to a time difference between arrival of the speech at the second physical microphone and arrival of the speech at the first physical microphone.

224. In Avendano's system, "due to a space difference between the microphones, the difference in times of arrival of the signals from a speech source to the microphones may be utilized to localize the speech source." APPLE-1005, 1:33-36. Further, Avendano discloses that "embodiments may use a combination of energy level differences and time delays to discriminate speech." *Id.,* 3:56-58. Specifically, as I explained in Section VI(A) (incorporated herein), Avendano system delays a first signal $x_1$ by a "delay node 402," and delays a second signal $x_2$ by a "delay node 404." *Id.*, 15:35, FIG. 4a.

225. A POSITA would have recognized that, in order to account for the "space difference between the microphones," the delays would need to be proportional to a time difference between arrival of the speech at the second physical microphone and arrival of the speech at the first physical microphone. For example, a POSITA would have recognized that, if the "space difference between the microphones"

were to increase, the delay would likewise increase to account for the increase in propagation time. Similarly, a POSITA would have recognized that, if the "space difference between the microphones" were to decrease, the delay would likewise decrease to account for the decrease in propagation time.

226. To the extent that Avendano alone does not describe that the delay is proportional to a time difference between arrival of the speech at the second physical microphone and arrival of the speech at the first physical microphone, it would have been obvious to modify Avendano's system in view of Bisgaard to include such a feature.

227. As I explain in Section VIII(A) (incorporated herein), a POSITA would have found it obvious to further combine Avendano and Visser with Bisgaard by incorporating Bisgaard's "delays" into Avendano's system.

228. Further, Bisgaard's "delays" are used to account for "the amount of time used by a sound signal to propagate in the 0° azimuth direction from [a] front microphone ... to [a] rear microphone"). APPLE-1010, [0042]; *see also* APPLE-1011, 5:22-24.

229. A POSITA would have recognized that, in order to account for "the amount of time used by a sound signal to propagate in the 0° azimuth direction from [a] front microphone ... to [a] rear microphone," the sound signal would be delayed by an amount of time that is proportional to a time difference between arrival of the

speech at one physical microphone and arrival of the speech at another physical microphone.

230. For example, a POSITA would have recognized that, if "the amount of time used by a sound signal to propagate in the 0° azimuth direction from [a] front microphone ... to [a] rear microphone" were to increase, the delay would likewise proportionally increase to account for the increase in propagation time. APPLE-1010, [0042]; *see also* APPLE-1011, 5:22-24. Similarly, a POSITA would have recognized that, if "the amount of time used by a sound signal to propagate in the 0° azimuth direction from [a] front microphone ... to [a] rear microphone" were to decrease, the delay would likewise proportionally decrease to account for the decrease in propagation time. APPLE-1010, [0042]; *see also* APPLE-1011, 5:22-24.

### C. Claim 9

*[9] The method of claim 8, wherein the forming of the first virtual microphone comprises applying the filter to the second signal.*

231. The combination of Avendano and Visser (and in the further combination with Bisgaard) discloses that the forming of the first virtual microphone comprises applying the filter to the second signal.

232. As I explain in Sections VII(A)-(B) (incorporated herein), a POSITA would have found it obvious to combine Avendano and Visser.

233. According to this combination, Avendano's system would include Visser's cross filters $w_{12}$ and $w_{21}$. Sections VII(A)-(B). Specifically, Visser's cross filter $w_{12}$ would be applied to Avendano's second signal $x_2$. *Id.* Further, the filtered second signal would be used to form Avendano's first virtual microphone $C_f$ (*e.g.*, by combining the filtered second signal with Avendano's first signal $x_1$). *Id.*

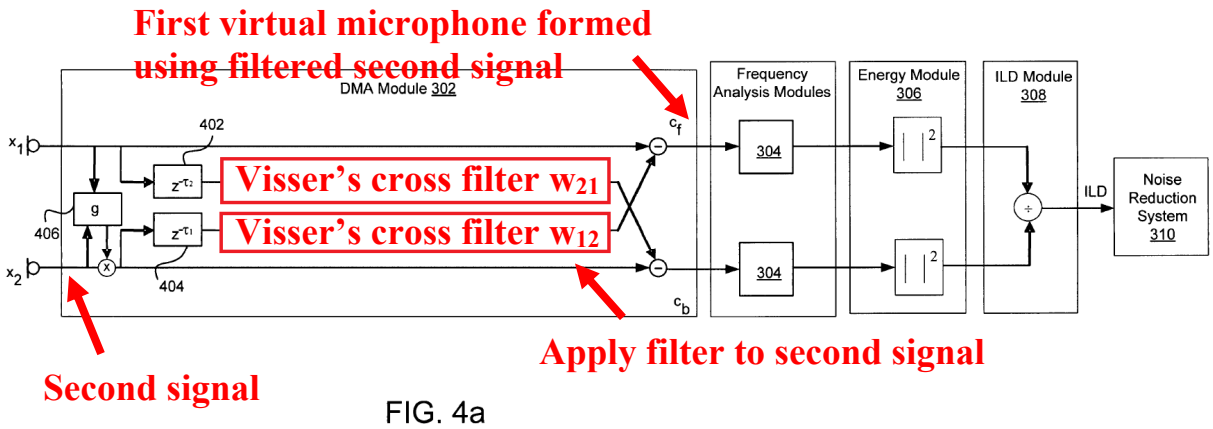



FIG. 4a

**APPLE-1005, FIG. 4a (modified)**

### D. Claim 10

*[10] The method of claim 9, wherein the forming of the first virtual microphone comprises applying the calibration to the second signal.*

234. Avendano (and in the combination with Visser and Bisgaard) discloses that the forming of the first virtual microphone comprises applying the calibration to the second signal.

235. As I explain in Section VII(F) (incorporated herein), Avendano discloses applying a calibration (*e.g.,* a "gain factor, g") to at least the second signal $x_2$.

236. Further, in Avendano's system, the calibrated second signal $x_2$ is used to form the first virtual microphone $C_f$ by combining the calibrated second signal $x_2$ with the first signal $x_1$. APPLE-1005, 4:47-52, 5:25-35.

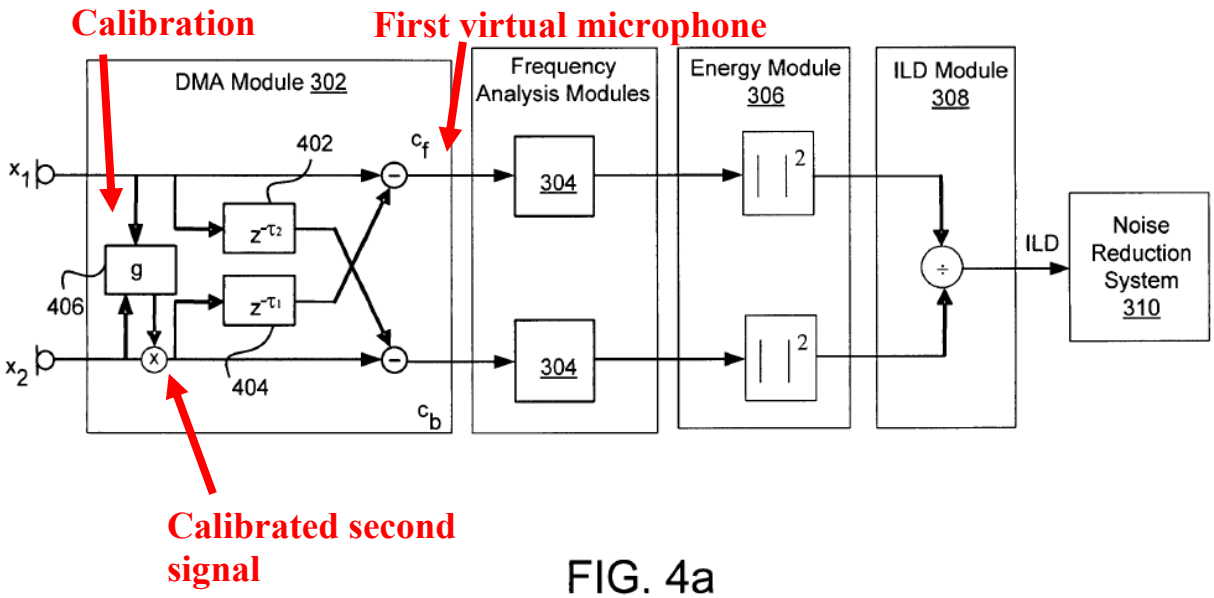



FIG. 4a

**APPLE-1005, FIG. 4a (annotated)**

### E. Claim 11

*[11] The method of claim 10, wherein the forming of the first virtual microphone comprises applying the delay to the first signal.*

237. Avendano (and in the combination with Visser and Bisgaard) discloses that the forming of the first virtual microphone comprises applying the delay to the first signal.

238. Avendano discloses that a "DMA Module 302" is used to form at least the first virtual microphone $C_f$. APPLE-1005, 4:20-5:39, FIGS. 3, 4a. Further, as explained by Avendano, in a "practical implementation of the DMA module 302" shown in FIG. 4b, a delay can be applied to the first signal $x_1$ via a "delay node 414." *Id.,* 8:38-51, FIG. 4b.

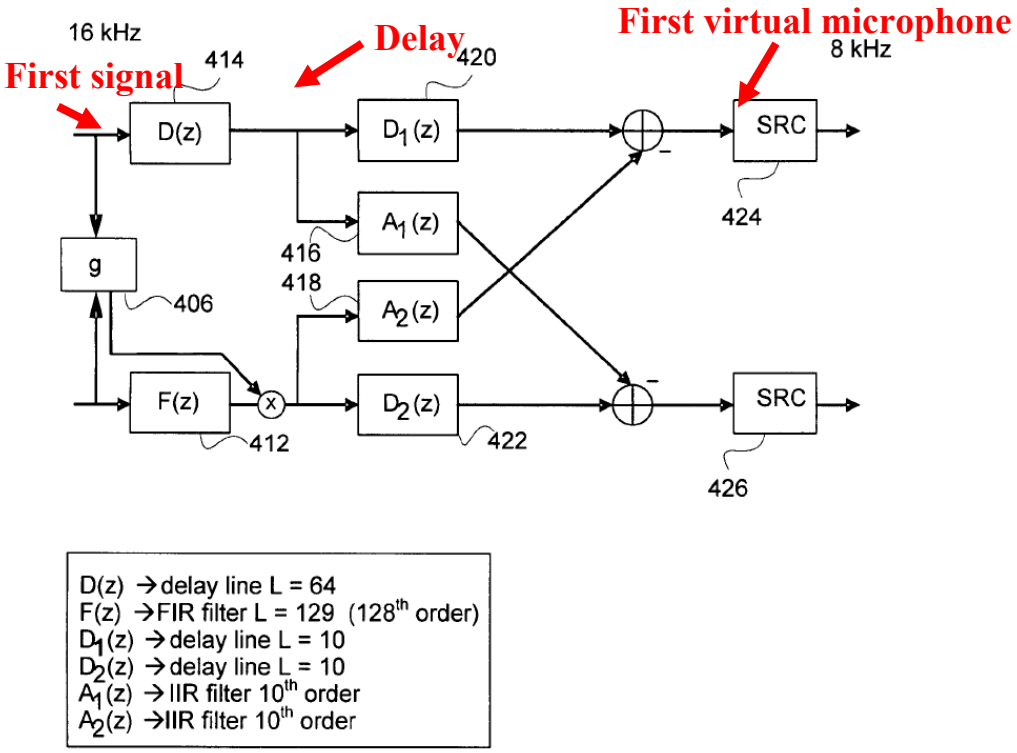



FIG. 4b

**APPLE-1005, FIG. 4b (annotated)**

239. Accordingly, Avendano discloses that forming of the first virtual microphone comprises applying the delay to the first signal.

### F. Claim 12

*[12] The method of claim 11, wherein the forming of the first virtual microphone by the combining comprises subtracting the second signal from the first signal.*

240. Avendano (and in the combination with Visser and Bisgaard) discloses that the forming of the first virtual microphone by the combining comprises subtracting the second signal from the first signal.

241. Avendano discloses that the first virtual microphone $C_f$ is formed by subtracting the second signal $x_2$ from the first signal $x_1$. APPLE-1005, 5:15-35, FIG. 4a.

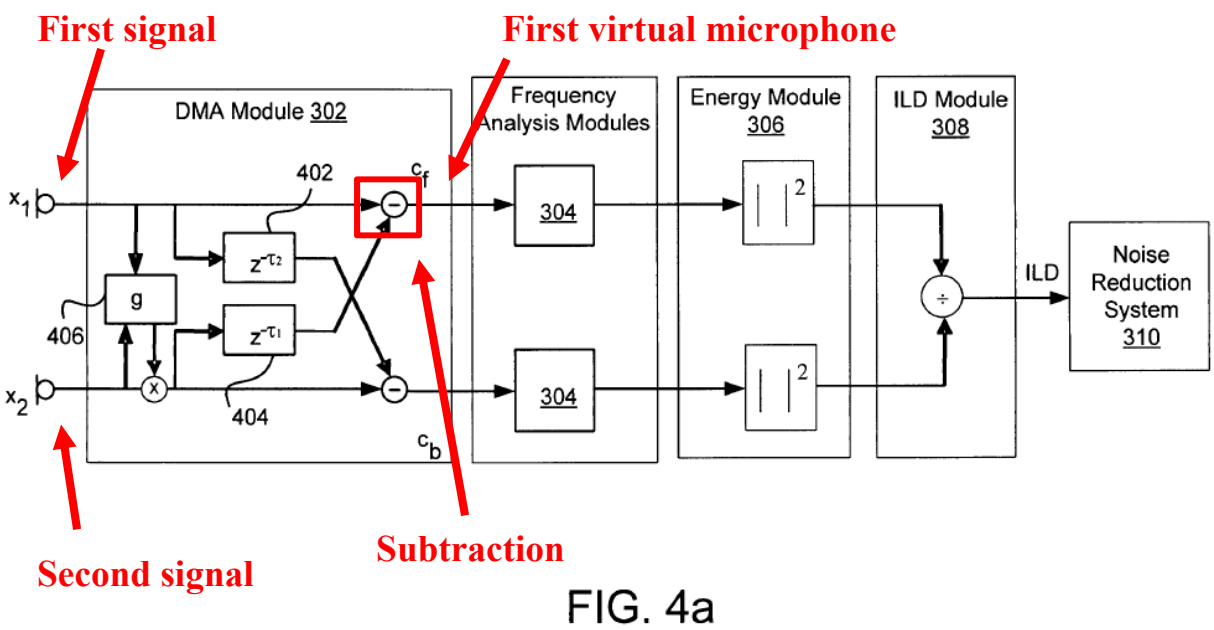



FIG. 4a

**APPLE-1005, FIG. 4a (annotated)**

### G. Claim 13

*[13] The method of claim 12, wherein the filter is an adaptive filter.*

242. The combination of Avendano and Visser (and in the further combination with Bisgaard) discloses the filter is an adaptive filter.

243. As I explain in Sections VII(A)-(B) (incorporated herein), a POSITA would have found it obvious to combine Avendano and Visser. According to this combination, Avendano's system would include Visser's cross filters $w_{12}$ and $w_{21}$. Sections III(A)-(B).

244. As Visser describes, the cross filters $w_{12}$ and $w_{21}$ may be adaptive filters. APPLE-1006, 8:16-18, 16:3-28, 16:57-25:9, FIGS. 5, 7, 10-13. For example, Visser discloses that the cross filters $w_{12}$ and $w_{21}$ are implemented as a part of an "ICA or BSS processing function." *Id.,* 17: 29-32. Further, Visser discloses "adapt[ing]" filters using a "learning stage 752" in an "ICA process." *Id.,* 21:17-58, FIG. 12. Visser also discloses an "ICA process" having filters with "adaptable and adjustable filter coefficient[s]." *Id.*, 16:61-17:4.

### H. Claim 14

*[14] The method of claim 13, comprising adapting the filter to minimize a second virtual microphone output when only speech is being received by the first physical microphone and the second physical microphone.*

245. The combination of Avendano and Visser (and in the further combination with Bisgaard) discloses adapting the filter to minimize a second virtual

microphone output when only speech is being received by the first physical microphone and the second physical microphone.

246. As I explain in Sections VII(A)-(B) (incorporated herein), a POSITA would have found it obvious to combine Avendano and Visser. According to this combination, Avendano's system would include Visser's cross filters $w_{12}$ and $w_{21}$. Sections VII(A)-(B).

247. Further, as I explain in Section VIII(G) (incorporated herein), Visser's cross filters $w_{12}$ and $w_{21}$ may be adaptive filters. In particular, Visser's cross filters $w_{12}$ and $w_{21}$ may be adapted according to an "ICA or BSS processing function." APPLE-1006, 21:17-58. Further, Visser's "ICA process" may include filters with "adaptable and adjustable filter coefficient[s]." *Id.*, 16:61-17:4.

248. Specifically, Visser's cross filters $w_{12}$ and $w_{21}$ are adapted to filter signals received from two physical microphones (e.g., "[i]nput signals $X_1$ and $X_2$") in order to generate two virtual microphones having different speech and noise content. *Id.*, 17:36-44 (emphasis added). The first virtual microphone (also referred to as a "speech channel") "contains predominantly desired signals," namely speech. *Id.* The second virtual microphone (also referred to as a "noise channel") "contains predominantly noise signals." *Id.*

249. According to Visser's configuration, if the two physical microphones were to receive only speech (*i.e.,* and no noise), then the first virtual microphone (*e.g.,*

the "speech channel") would be maximized and the second virtual microphone (*e.g.,* the "noise channel") would be minimized. In contrast, if the two physical microphones were to receive only noise (*i.e.,* and no speech), then the first virtual microphone (*e.g.,* the "speech channel") would be minimized and the second virtual microphone (*e.g.,* the "noise channel") would be maximized.

250. Accordingly, Visser's cross filters $w_{12}$ and $w_{21}$ are adapted such that when the physical microphones only receive speech (*i.e.,* without noise), the second virtual microphone—which would contain "predominantly noise signals—would be minimized.

### I. Claim 15

*[15] The method of claim 13, wherein the adapting comprises applying a least-mean squares process.*

251. The combination of Avendano and Visser (and in the further combination with Bisgaard) discloses that the adapting comprises applying a least-mean squares process.

252. As I explain in Sections VII(A)-(B) (incorporated herein), a POSITA would have found it obvious to combine Avendano and Visser. According to this combination, Avendano's system would include Visser's cross filters $w_{12}$ and $w_{21}$. Sections VII(A)-(B). Further, as discussed in Section IV(H) (incorporated herein), Visser discloses that the cross filters $w_{12}$ and $w_{21}$ may be adaptive filters.

253. Further, Visser teaches that an "adaptive least mean square (NLMS) algorithm" can be used to "build[] a linear filter model." APPLE-1006, 23:13-25.

254. Based on Visser's disclosure, a POSITA would have recognized that adaptive filters (*e.g.,* Visser's cross filters $w_{12}$ and $w_{21}$) can be adapted by applying a least-mean squares process (*e.g.,* Visser's "adaptive least mean square (NLMS) algorithm").

### J. Claim 16

*[16] The method of claim 13, comprising generating coefficients of the filter during a period when only speech is being received by the first physical microphone and the second physical microphone.*

255. The combination of Avendano and Visser (and in the further combination with Bisgaard) discloses generating coefficients of the filter during a period when only speech is being received by the first physical microphone and the second physical microphone.

256. As I explain in Sections VII(A)-(B) (incorporated herein), a POSITA would have found it obvious to combine Avendano and Visser. According to this combination, Avendano's system would include Visser's cross filters $w_{12}$ and $w_{21}$ to produce a first virtual microphone that would "contain[] predominantly desired signals," namely speech, and a second virtual microphone that would "contain[] predominantly noise signals." Sections VII(A)-(B).

257. Further, as I explain in Section VIII(G) (incorporated herein), Visser's cross filters $w_{12}$ and $w_{21}$ may be adaptive filters, such as filters with "adaptable and adjustable filter coefficient[s]."

258. Further, Visser discloses generating coefficients of the filter during a period when only speech is being received by the first physical microphone and the second physical microphone. APPLE-1006, 9:12-4.

259. For example, as described by Visser:

> Generally, the ICA module is not able to perform its main separation function in any time interval when the desired speaker is not speaking, and therefore may be turned off. This "on" and "off" state can be monitored and controlled by the voice activity detection module 185 based on comparing energy content between input channels or desired speaker a priori knowledge such as specific spectral signatures. By turning the ICA off when desired speech is not present, the ICA filters do not inappropriately adapt, thereby enabling adaptation only when such adaptation will be able to achieve a separation improvement. Controlling adaptation of ICA filters allows the ICA process to achieve and maintain good separation quality even after prolonged periods of desired speaker silence and avoid algorithm singularities due to unfruitful separation efforts for addressing situations the ICA stage

> cannot solve. Various ICA algorithms exhibit different degrees of robustness or stability towards isotropic noise but turning off the ICA stage during desired speaker absence, or alternatively noise absence, adds significant robustness to the methodology.

*Id.,* 9:22-41 (emphasis added).

260. Further, as described by Visser:

> To improve the adaptability of the speech separation process, the learning process 188 may be activated responsive to the voice activity control signal 186. In this way, the speech separation process only applies its adaptive learning processes when desired speech is likely occurring. Also, by deactivating the learning processing when only noise is present, or alternatively, absent, processing and battery power may be conserved.

*Id.,* 9:12-19.

261. That is, according to Visser, an adaptive speech separation process can be improved, at least in part, by generating coefficients of the filter during a period when only speech is being received by the first physical microphone and the second physical microphone. *Id.* 9:12-43.

### K. Claim 23

*[23] The method of claim 12, wherein the filter is a static filter.*

262. Avendano (and in the combination with Visser and Bisgaard) discloses that the filter is a static filter.

263. As I explain in Section in Section VII(B) (incorporated herein), and specifically in reference to claim element [1b], Avendano discloses a filter "delay node 402," which describes a relationship for speech between the first physical microphone and the second physical microphone. APPLE-1005, 4:28-34, 5:19-35, FIGS. 4a-4b. Specifically, Avendano's "delay node 402" describes a relationship whereby the arrival of speech at one physical microphone will be delayed in time relative to the arrival of speech at the other physical microphone. *Id.*

264. A POSITA would have recognized that, if the first physical microphone and the second physical microphone do not move relative to one another or to the source of sound, then the delay in time between the arrival of speech at one physical microphone and the arrival of speech at the other physical microphone would not change. Accordingly, a POSITA would have found to obvious to implement the delay as a static filter under these circumstances.

265. To the extent that Avendano alone does not describe the filter is a static filter, the combination of Avendano and Visser (and in the further combination with Bisgaard) would have rendered this feature obvious.

266. As I explain in Section VIII(A) (incorporated herein), a POSITA would have found it obvious to further combine Avendano and Visser with Bisgaard. Further, as I explain in Sections VII(A)-(B) (incorporated herein), according to this combination Avendano's system would include Visser's cross filters $w_{12}$ and $w_{21}$.

267. Visser's cross filters $w_{12}$ and $w_{21}$ may be static filters. For example, Visser's filters can have "filter values ... or taps." APPLE-1006, 22:7-8, FIG. 12. Further, as Visser explains:

> Since the headset or earpiece often has only a limited range of operating conditions, the <u>default values</u> for the taps may be selected to account for the expected operating arrangement. For example, the distance from each microphone to the speaker's mouth is usually held in a small range, and the expected frequency of the speaker's voice is likely to be in a relatively small range. Using these constraints, as well as actual operation values, a set of reasonably accurate tap values may be determined. By carefully selecting default values, the time for the ICA to perform expectable separation is reduced. Explicit constraints on the range of filter taps to constrain the possible solution space should be included. These constraints may be derived from directivity considerations or experimental values obtained through convergence to optimal solutions in previous experiments. <u>It will also be appreciated</u>

that the default values may adapt over time and according to environmental conditions.

*Id.,* 2:8-25 (emphasis added).

268. That is, Visser teaches that, in cases that the device "has only a limited range of operating conditions," "default values" can be selected for the "taps ... to account for the expected operating arrangement." *Id.* Visser discloses that "the default values may adapt over time and according to environment conditions." *Id.* (emphasis added). However, a POSITA would have recognized that such an adaption would not necessarily be required. Indeed, a POSITA would have recognized that if changes to the expected "operating conditions" of the device are limited and "a set of reasonably accurate tap values may be determined, *id.*, there would have been no technical benefit in adapting the "default values." Specifically, a POSITA would have recognized that under these circumstances, the adaption (if any) would likewise be limited and the "reasonably accurate tap values" would remain "reasonably accurate" absent significant changes to the "operating conditions."

269. To the extent that Visser does not expressly describe that the cross filters $w_{12}$ and $w_{21}$ may be static filters, a POSITA would have found it obvious to implement the cross filters $w_{12}$ and $w_{21}$ as static filters.

270. As discussed above, Visser teaches that, in cases that the device "has only a limited range of operating conditions," "default values" can be selected for the "taps ... to account for the expected operating arrangement." APPLE-1006, 2:8-25.

271. Further, Visser discloses that "the separation process could use ... an application specific adaptive filter process using some degree of a priori knowledge about the acoustic environment to accomplish substantially similar signal separation." *Id.,*16:23-27. A POSITA would have recognized that, if the device is not expected to deviate from the "limited range of operating conditions" and/or sufficient "a priori knowledge about the acoustic environment" is known, then there would not have been a need to modify the "taps" of the filter to account for changes in the operating conditions.

272. To the extent that the combination of Avendano and Visser does not describe the filter is a static filter, the combination of Avendano, Visser, and Bisgaard would have rendered this feature obvious.

273. As I explain in Section VII(B) (incorporated herein), and specifically in reference to claim element [1b], Avendano discloses a filter "delay node 402," which describes a relationship for speech between the first physical microphone and the second physical microphone. APPLE-1005, 4:28-34, 5:19-35, FIGS. 4a-4b. Further, as I explain in Sections VIII(A-(B) (incorporated herein), Bisgaard provide specific details regarding how to implement such a delay, such as

configuring the delay to account for "the amount of time used by a sound signal to propagate in the 0° azimuth direction from [a] front microphone ... to [a] rear microphone"). APPLE-1010, [0042]; *see also* APPLE-1011, 5:22-24.

274. A POSITA would have recognized that, if the first microphone and the rear microphone do not move relative to one another or to the source of sound (*i.e.,* the positions of the microphones are static with respect to one another and the source of sound), then "the amount of time used by a sound signal to propagate in the 0° azimuth direction from [a] front microphone ... to [a] rear microphone" would not change. Accordingly, a POSITA would have found to obvious to implement the delay as a static filter under these circumstances.

### L. Claim 24

*[24] The method of claim 23, wherein the forming of the filter comprises: determining a first distance as distance between the first physical microphone and a mouth of the speaker; determining a second distance as distance between the second physical microphone and the mouth; and forming a ratio of the first distance to the second distance.*

275. The combination of Avendano, Visser and Bisgaard discloses that the forming of the filter comprises: determining a first distance as distance between the first physical microphone and a mouth of the speaker; determining a second

distance as distance between the second physical microphone and the mouth; and forming a ratio of the first distance to the second distance.

276. As I explain in Section VIII(A) (incorporated herein), a POSITA would have found it obvious to further combine Avendano and Visser with Bisgaard.

277. Further, as I explain in Section VIII(B) (incorporated herein), a POSITA would have recognized that, in order to account for "the amount of time used by a sound signal to propagate in the 0° azimuth direction from [a] front microphone ... to [a] rear microphone," the sound signal would be delayed by an amount of time that is proportional to a time difference between arrival of the speech at one physical microphone and arrival of the speech at another physical microphone. APPLE-1010, [0042]; *see also* APPLE-1011, 5:22-24.

278. Further, a POSITA would have recognized that, given constant environment conditions, sound would propagate through an environment at a constant speed. Further still, a POSITA would have recognized that an environment's conditions are unlikely to change during the short time period during which sound propagates between two closely positioned microphones, and thus the speed of sound is unlikely to change during this time period. Thus, a POSITA would have recognized that the propagation time of sound from a source of a sound to a destination would be proportional to the distance between the source of the sound and the destination.

279. Accordingly, to determine a time difference between arrival of the speech at one physical microphone and arrival of the speech at another physical microphone, a POSITA would have found it obvious to determine a parameter representing the proportional relationship between the first distance and the second distance. Specifically, a POSITA would have found it obvious to (i) determine a first distance between the first physical microphone and a mouth of the speaker (*i.e.,* the source of the sound), (ii) determine a second distance between the second physical microphone and the mouth, and (iii) form a ratio of the first distance to the second distance, which represents the proportional relationship between the first distance and the second distance.

## IX. GROUND 3: CLAIMS 17-19 ARE RENDERED OBVIOUS BY AVENDANO IN VIEW OF VISSER, BISGAARD, AND HOU

### A. Combination of Avendano, Visser, Bisgaard, and Hou

280. As I explain in Section VIII(A) (incorporated herein), a POSITA would have found it obvious to combine Avendano, Visser, and Bisgaard.

281. Further, as I detail below, a POSITA would have found it obvious to further combine Avendano, Visser, and Bisgaard with Hou.

282. Both Avendano and Hou come from the same field of endeavor of enhancing speech and attenuating noise. APPLE-1005, Abstract, 1:24-26, 3:13-26, 3:42-60; APPLE-1008, Abstract, 2:44-52, 4:40-49.

283. Further, both Avendano and Hou describe enhancing speech and attenuating noise using similar techniques, including equalizing the signal levels between two signals and generating directional signals based on the equalized signals.

284. For example, as I explain in Section VI(A) (incorporated herein), Avendano's "audio device 104" includes at least a "primary microphone 106" and a "secondary microphone 108," each of which may be an "omni-directional microphone." APPLE-1005, 3:27-35. Further, Avendano's "primary microphone 106 is much closer to [an] audio source 102 than the secondary microphone 108." *Id.*, 3:45-49. Due to this configuration, "the intensity level is higher for the primary microphone 106 resulting in a larger energy level during a speech/voice segment." *Id.*

285. Avendano discloses that a "gain factor, g" can be applied to at least a second signal to account for the differences in intensity levels between signals generated by the two microphones. *Id.*, 5:36-39, 9:54-59, FIGS. 4a-4b. In particular, "[t]he gain factor, g, is computed by the gain module 406 to equalize the signal levels" of the first and second signals. *Id.*, 5:36-37. As Avendano explains, this equalization is beneficial, as "systems can suffer loss of performance when the microphone signals have different levels." *Id.* 5:37-39.

286. Further, in Avendano's system, the equalized signal is used to generate directional signals, such as a "cardioid primary signal ($C_f$)" having a directional

pattern for receiving sounds "within a front cardioid region around the audio device 104" and a "cardioid secondary signal ($C_b$)" signal having a directional pattern for receiving sounds "within a back cardioid region around the audio device 104." *Id.,* 41-52, 5:25-35, FIGS. 4a-4b, 6.

287. Hou describes enhancing speech and attenuating noise using similar techniques as those described by Avendano. For example, Hou describes that the "two-microphone directional processing system 500" shown in Figure 5 serves a similar function as Avendano's "gain factor," namely "compensat[ing] (or correct[ing]) for the relative difference in sensitivity between ... mismatched first and second microphones." APPLE-1008, 5:47-50. Similarly, like Avendano, Hou teaches that such a compensation or correction is beneficial, as "[t]he sensitivity of the microphones of the sound pick up system must be matched in order to achieve good directionality." *Id.*, 1:48-50. For example, as Hou explains, "[w]hen the sensitivities of the microphones are not properly matched, then the directionality is substantially degraded and thus the ability to suppress interference coming from a particular direction is poor." *Id.*, 1:50-2:2.

288. Like Avendano, Hou also discloses that the compensated or corrected signal is processed to generate a directional signal (*e.g.,* "an output signal ... hav[ing] robust directionality") to aid in "directional noise suppression." *Id.,* 5:50-56, 9:28-41.

289. A POSITA would have recognized that Avendano teaches the importance of equalizing signal levels between signals generated by two respective microphones, and that Hou discloses a specific "two-microphone directional processing system 500" for performing precisely this function. Accordingly, a POSITA would have been motivated to implement Hou's signal compensation or correction components in Avendano's system to further improve the equalization of the signal output by the microphones. *Id.,* 5:50-56, 9:28-41. Specifically, a POSITA would have been motivated to modify Avendano's system to include Hou's signal compensation or correction components to compensate or correct for the relative difference in sensitivity between mismatched first and second microphones in order to generate a signal "hav[ing] robust directionality" to enhance "directional noise suppression." *Id.*

290. In an example modification, Hou's "first minimum estimate unit 508," "second minimum estimate unit 510," and divide unit 512," would be implemented in Avendano's "gain module 406." Specifically, Hou's "first minimum estimate unit 508" would be configured to receive Avendano's signal $x_2$, and Hou's "second minimum estimate unit 510" would be configured to receive Avendano's signal $x_1$. Further, Hou's "divide unit 512" would be configured to calculate a gain as the quotient of the output of the "first minimum estimate unit 508" and the output of the "second minimum estimate unit 510." Further still, the calculated gain would

be applied to Avendano's signal $x_1$ by the "multiplication unit 514." An example of this modification is shown below:

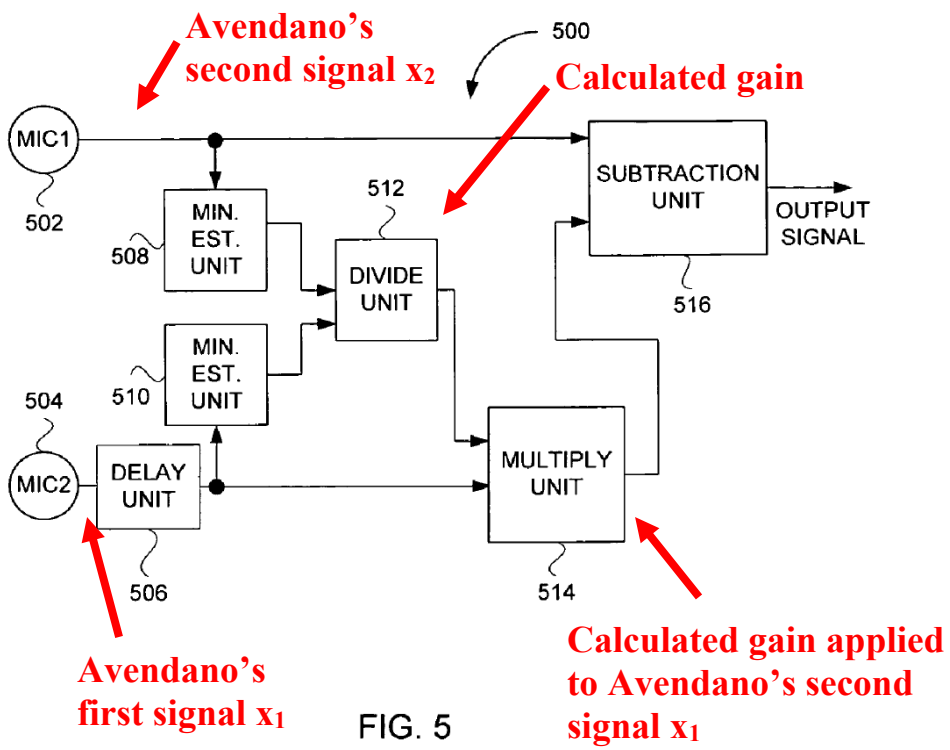



FIG. 5

**APPLE-1008, FIG. 5 (annotated)**

291. Although Hou discloses multiplying one of the signals by the calculated gain, a similar equalization effect also could be achieved by multiplying the other signal by the inverse of the calculated gain. As an illustrative example, if a first signal were greater in magnitude than a second signal, the two signals can be equalized by multiplying the second signal by a particular gain (*e.g.,* to increase the magnitude of the second signal to match that of the first signal). Alternatively, the two signals can be equalized by multiplying the first signal by an inverse of

that gain (*e.g.,* to decrease the magnitude of the first signal to match that of the second signal).

292. This is analogous to Avendano's system in which a gain is calculated based on the signals $x_1$ and $x_2$, and is applied to the signal $x_2$ (*e.g.,* by a multiplexer).

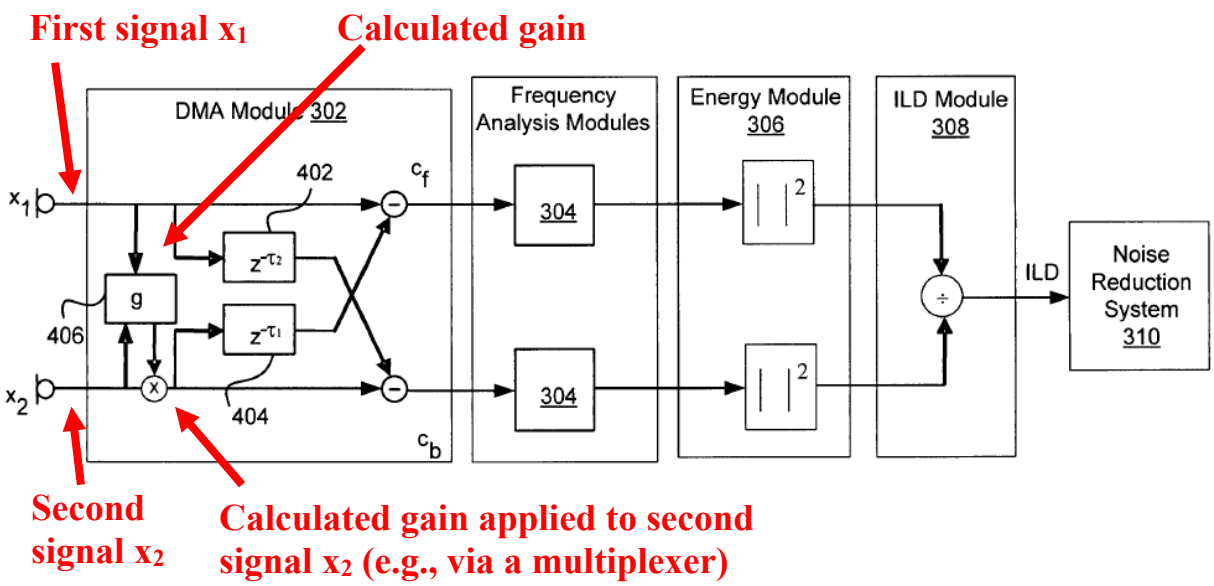



FIG. 4a

**APPLE-1005, FIG. 4a (annotated)**

293. Further, given the similarities between Avendano's "gain factor, g" and Hou's "two-microphone directional processing system 500," a POSITA would have had a reasonable expectation of success in incorporating Hou's "two-microphone directional processing system 500" in Avendano's system. For example, Avendano's "gain factor, g" and Hou's "two-microphone directional processing system 500" are configured to receive similar input signals (*e.g.,* signals

from two microphones), and generate similar outputs (*e.g.,* "equalized" or "compensated" signals). APPLE-1005, 5:36-39, 9:54-59, FIGS. 4a-4b; APPLE-1008, 5:25-50. Given these similarities, limited modifications would have been necessary to incorporate Hou's "two-microphone directional processing system 500" into Avendano's system. Further, other operations in Avendano's system would not have been impacted by the combination, as Hou's "two-microphone directional processing system 500" are configured to receive similar input signals and generate similar outputs as Avendano's "gain factor, g." APPLE-1005, 5:36-39, 9:54-59, FIGS. 4a-4b; APPLE-1008, 5:25-50.

294. For these reasons, the combination of Avendano and Hou would have been well within the grasp of a POSITA and a POSITA would have had a reasonable expectation of success in making the combination.

### B. Claim 17

*[17] The method of claim 13, wherein the forming of the filter comprises: generating a first quantity by applying a calibration to the second signal; generating a second quantity by applying the delay to the first signal; forming the filter as a ratio of the first quantity to the second quantity.*

295. The combination of Avendano and Hou (and in the further combination with Visser and Bisgaard) discloses that the forming of the filter comprises: generating a first quantity by applying a calibration to the second signal; generating a second

quantity by applying the delay to the first signal; forming the filter as a ratio of the first quantity to the second quantity.

296. As I explain in Section IX(A) (incorporated herein), a POSITA would have found it obvious to further combine Avendano, Visser, and Bisgaard with Hou.

297. According to an example modification, Hou's "first minimum estimate unit 508" would be configured to receive Avendano's second signal $x_2$ (*e.g.,* from Avendano's second physical microphone 108), and apply a calibration to the second signal $x_2$. (*e.g.,* estimate a minimum of the second signal) to generate a first quantity. Section V(B); APPLE-1008, 5:38-39, FIG. 5.

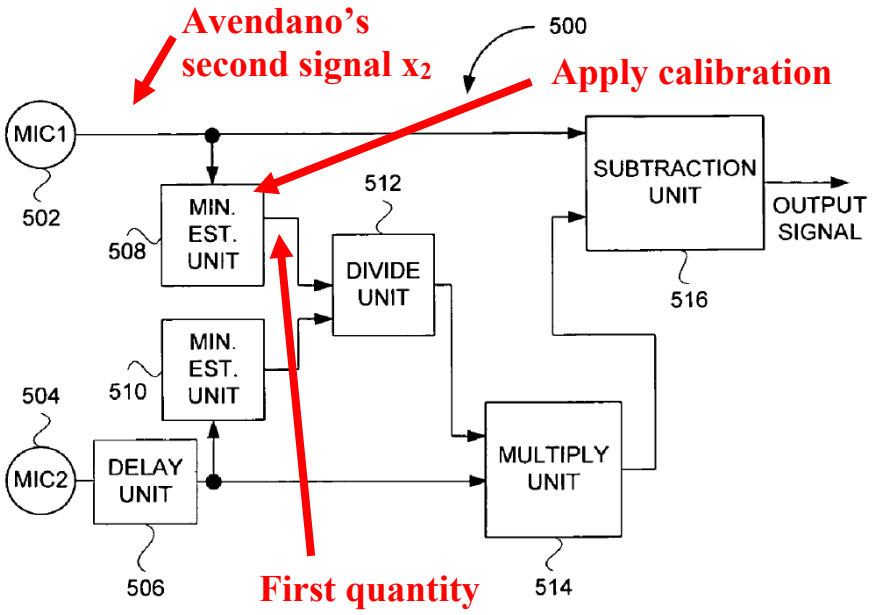
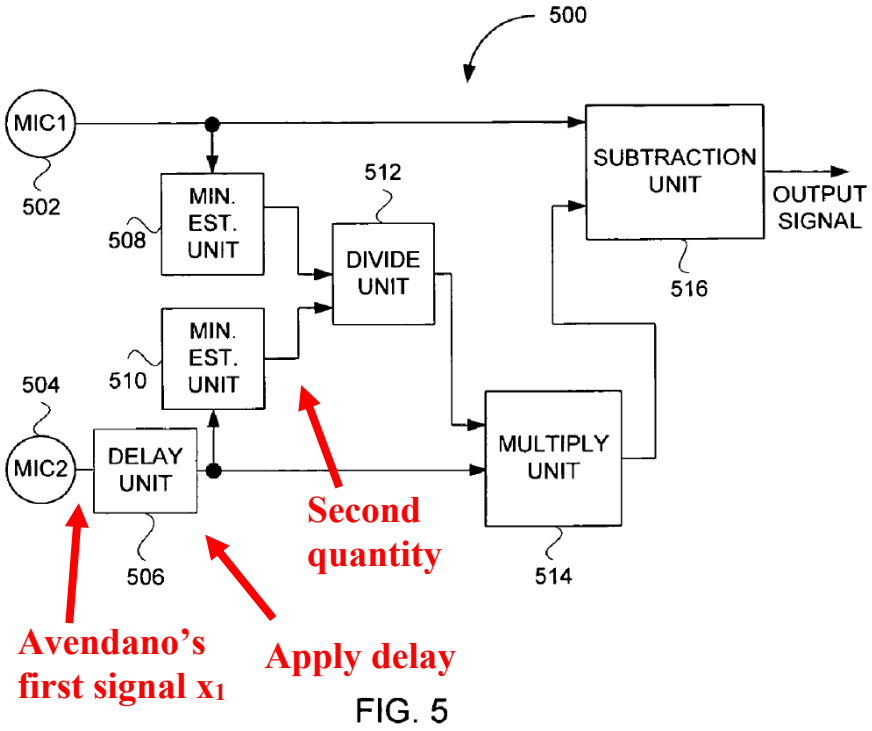



FIG. 5

**APPLE-1006, FIG. 5 (annotated)**

298. Further, Hou's "delay unit 506" would be configured to receive Avendano's first signal $x_1$ (*e.g.,* from Avendano's first physical microphone 106), and apply a delay to that first signal $x_2$ to generate a second quantity. Section IX(A); APPLE-1008, 5:36-38, FIG. 5.




**APPLE-1006, FIG. 5 (annotated)**

299. Further, Hou's "divide unit 512" forms a filter as a ratio of the first quantity to the second quantity. Section IX(A); APPLE-1008, 5:42-45, FIG. 5. Specifically, "[t]he divide unit 512 produces a quotient by dividing the first minimum estimate by the second minimum estimate." APPLE-1008, 5:42-45. This "quotient represents a scaling amount that is sent to a multiplication unit

514." *Id.* That is, the scaling amount represents a filter that would be applied to Avendano's first signal $x_1$ (*e.g.,* using the "multiply unit 514").

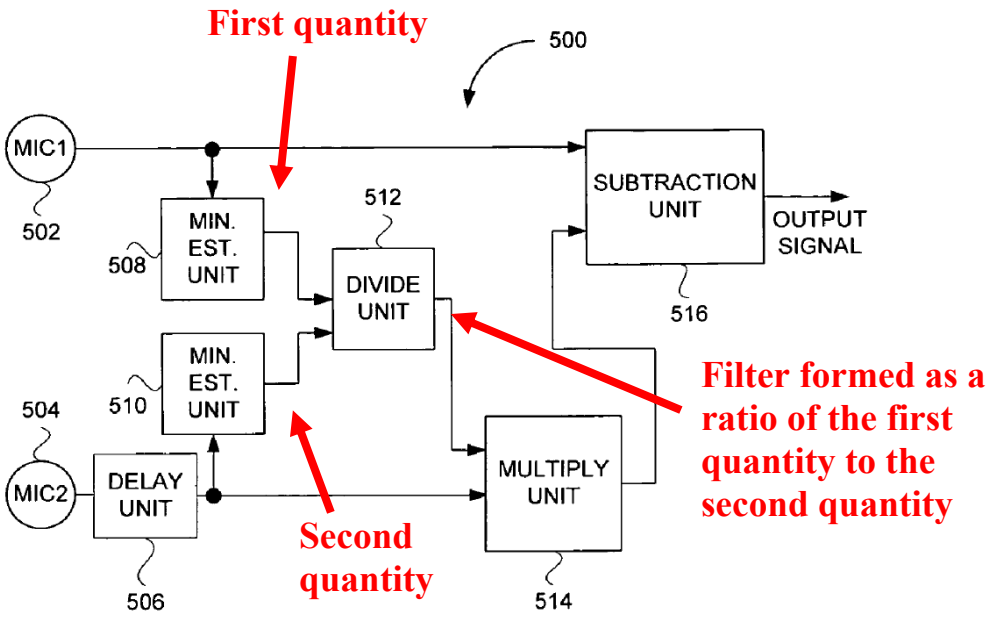



FIG. 5

**APPLE-1006, FIG. 5 (annotated)**

### C. Claim 18

*[18] The method of claim 17, wherein the generating of the energy ratio comprises generating the energy ratio for a frequency band.*

300. Avendano (and in the combination with Visser, Bisgaard, and Hou) discloses that the generating of the energy ratio comprises generating the energy ratio for a frequency band.

301. As described in Section VII(B) (incorporated herein), and specifically in reference to claim element [1d], Avendano discloses generating an energy ratio (*e.g.,* "ILD"). This energy ratio is generated for a frequency band.

302. For example, Avendano's "frequency analysis module 304 takes the cardioid signals and mimics the frequency analysis of the cochlea (i.e., cochlear domain) simulated by a filter bank." APPLE-1005, 4:55-5:2. In particular, Avendano's "frequency analysis module 304" can "separate[] the cardioid signals into frequency bands," and perform "sub-band analysis on the acoustic signal [to] determine[] what individual frequencies are present in the complex acoustic signal during a frame." *Id*.

303. As described by Avendano, "[o]nce the frequencies are determined," the energy ratio (*e.g.,* "ILD") is calculated by the "energy module 306" based on signals that are output by the "frequency analysis module 304." *Id.,* 5:3-10. Specifically, the energy ratio can be calculated "based on bandwidth of the cochlea channel" (*e.g.,* the frequency band(s) for which the "frequency analysis module 304 ... mimics the frequency analysis of the cochlea"). *Id.*

304. Accordingly, Avendano discloses generating the energy ratio for a frequency band.

305. This is also reflected in Avendano's equation for the energy ratio ("ILD"), which indicates that the energy ratio is be generated as a vector for a frequency band (ω):

Energy ratio

$$\boxed{ILD(t, \omega)} = \frac{\int |C_f(t', \omega)|^2 \, dt'}{\int_{frame} |C_b(t', \omega)|^2 \, dt'}.$$

**APPLE-1005, 6:16 (annotated)**

**D. Claim 19**

*[19] The method of claim 17, wherein the generating of the energy ratio comprises generating the energy ratio for a frequency subband.*

306. Avendano (and in the combination with Visser, Bisgaard, and Hou) discloses that the generating of the energy ratio comprises generating the energy ratio for a frequency subband.

307. As described in Section VII(B) (incorporated herein), and specifically in reference to claim element [1d], Avendano discloses generating an energy ratio (*e.g.,* "ILD"). This energy ratio is generated for a frequency subband.

308. For example, Avendano's "frequency analysis module 304 takes the cardioid signals and mimics the frequency analysis of the cochlea (i.e., cochlear domain) simulated by a filter bank." APPLE-1005, 4:55-5:2. In particular, Avendano's "frequency analysis module 304" can "separate[] the cardioid signals into

frequency bands," and perform "sub-band analysis on the acoustic signal [to] determine[] what individual frequencies are present in the complex acoustic signal during a frame." *Id*.

309. As described by Avendano, "[o]nce the frequencies are determined," the energy ratio (*e.g.,* "ILD") is calculated by the "energy module 306" based on signals that are output by the "frequency analysis module 304." *Id.,* 5:3-10. Specifically, the energy ratio can be calculated "based on bandwidth of the cochlea channel" (*e.g.,* the frequency band(s) for which the "frequency analysis module 304 ... mimics the frequency analysis of the cochlea"). *Id.*

310. Accordingly, Avendano discloses generating the energy ratio for a frequency subband.

311. This is also reflected in Avendano's equation for the energy ratio ("ILD"), which indicates that the energy ratio is be generated as a vector for a frequency subband (ω):

**Energy ratio**

$$\boxed{ILD(t, \omega)} = \frac{\int |C_f(t', \omega)|^2 \, dt'}{\int_{frame} |C_b(t', \omega)|^2 \, dt'}.$$

**APPLE-1005, 6:16 (annotated)**

## X. GROUND 4: CLAIMS 20-22 ARE RENDERED OBVIOUS BY AVENDANO IN VIEW OF VISSER, BISGAARD, HOU, AND FREQUENCY ART (BYRNE, BURNETT, AND/OR BERGLUND)

### A. Combination of Avendano, Visser, Bisgaard, Hou, and Frequency Art (Byrne, Burnett, and/or Berglund)

312. As I explained in Section IX(A) (incorporated herein), a POSITA would have found it obvious to combine Avendano, Visser, Bisgaard, and Hou. Further, a POSITA would have found it obvious to further combine Avendano, Visser, Bisgaard, and Hou with one or more of Byrne, Burnett, and Berglund.

313. For example, as I explained in Section VII(A) (incorporated herein), Avendano describes determining the presence or absence of speech based on an energy ratio of energies of the first virtual microphone and the second virtual microphone. Further, as I explained in Sections IX(C)-(D) (incorporated herein), Avendano discloses determining the energy ratio for specific frequency range(s) (*e.g.,* the frequency band(s) or subband(s) for which the "frequency analysis module 304 ... mimics the frequency analysis of the cochlea" and performs "sub-band analysis on the acoustic signal [to] determine[] what individual frequencies are present in the complex acoustic signal during a frame." APPLE-1005, 4:55-5:10.

314. A POSITA would have recognized that the processes described by Avendano (and in combination with Visser, Bisgaard, and Hou) are specifically configured to analyze audible speech in specific frequency range(s). Accordingly,

a POSITA would have found it obvious to perform these processes specifically in frequency range(s) that are known to have significant spectral components of speech (*e.g.,* as taught by Byrne). For example, a POSITA would have found it obvious to perform the processes described by Avendano, Visser, Bisgaard, and Hou in a frequency range that includes at least 500 Hz (*e.g.,* the frequency of sound expected to have the highest intensity for human speech) to more readily detect the presence of English language speech. Example frequency ranges include (i) a frequency range greater than 200 Hz, (ii) a frequency range between 250 Hz and 1250 Hz, and (iii) in a frequency range between 200 Hz and 3000 Hz).

315. A POSITA also would have found it obvious to filter out frequency range(s) that are known to have significant spectral components of noise (*e.g.,* as taught by Berglund). For example, a POSITA would have found it obvious to filter out frequencies corresponding to common noise sources experienced by passengers of "road transportation vehicles" to reduce the likelihood of false positives in a speech detection process. Common noises include "low frequency noise" range (*e.g.,* a range of frequencies less than approximately 250 Hz"). APPLE-1013, FIG. 3.

316. A POSITA also would have recognized that the frequency range may vary, depending on the type of speech that is being detected (*e.g.,* the gender of the

speaker, the spoken language, etc.), the type of noise that is being filtered out, and the expected spectral distributions thereof.

317. Further, a POSITA would have found it obvious to tune the frequency range to accommodate the specific operational requirements of the system.

318. For example, a POSITA would have recognized that the time and/or computational resources required to perform the processes described by Avendano, Visser, and Bisgaard would depend on the range of frequencies that are analyzed. Accordingly, a POSITA would have found it obvious to balance (i) analyzing a larger range of frequencies to conduct a more comprehensive analysis, and (ii) analyzing a smaller range of frequencies to improve speed, accuracy, and/or efficiency.

319. As another example, a POSITA also would have recognized that, as taught by Burnett, analyzing certain frequency ranges may be beneficial in improving the separation of "voiced and unvoiced speech from background acoustic noise." APPLE-1012, [0063]. For example, as taught by Burnett, a system can "group [a speaker's] utterances by their spectral characteristics" by "bandpass[ing] the information from [two microphones] so that it is possible to see which bands in [one of the microphones] are more heavily composed of noise and which are more weighted with speech." *Id.*

320. Specifically, Burnett describes that a system can use bandpass filters to capture signals in certain frequency ranges to identify the presence of unvoiced speech. *Id.,* para. [0064] As an example, a system can use bandpass filters to accept signals in the frequency range of 500 Hz to 4000 Hz to identify the unvoiced speech "k" in "kick." *Id.* As another example, a system can use bandpass filters to accept signals in the frequency range of 1700 Hz to 4000 Hz to identify the unvoiced speech "sh" in "she." *Id.* A POSITA would have recognized that a system can use other bandpass filters to accept signals in other frequency ranges in order to detect other types of unvoiced speech.

321. Similarly, Burnett describes that a system can use bandpass filters to accept signals in other frequency ranges to identify the presence of voiced speech. As an example, a system can use bandpass filters to accept a frequency range of 300 Hz to 2500 Hz to identify the voiced speech "ee." *Id.* As another example, a system can use bandpass filters to accept a frequency range of 900 Hz to 1200 Hz to identify the voiced speech "ah." A POSITA would have recognized that a system can use bandpass filters to accept other frequency ranges in order to detect other types of voiced speech.

322. In view of Burnett, a POSITA would have found to obvious to tune the frequency range, depending on the specific unvoiced speech and/or voiced speech that is to be detected.

323. Accordingly, a POSITA would have found to obvious to tune the frequency range, depending on the specific use case at hand.

### B. Claim 20

*[20] The device of claim 29, wherein the frequency subband includes frequencies in a range from approximately 200 Hz to 3000 Hz.*

324. The combination of Avendano with Byrne, Burnett, and/or Berglund (and in the further combination with Visser, Bisgaard, and Hou) discloses that the frequency subband includes frequencies higher than approximately 200 Hertz (Hz).

325. As I explained in Section X(A) (incorporated herein), a POSITA would have found it obvious to further combine Avendano, Visser, Bisgaard, and Hou with one or more of Byrne, Burnett, and Berglund. In particular, a POSITA would have found it obvious to perform the processes described by Avendano, Visser, Bisgaard, and Hou specifically in frequency range(s) known to have significant spectral components of speech, while filtering out frequency range(s) that are known to have significant spectral components of noise. This would have resulted in performing the processes in a frequency subband that includes frequencies higher than approximately 200 Hz.

326. For example, as taught by Byrne, the spectral components of English language speech are predominantly in a frequency range greater than 200 Hz. APPLE-1009, FIG. 1.

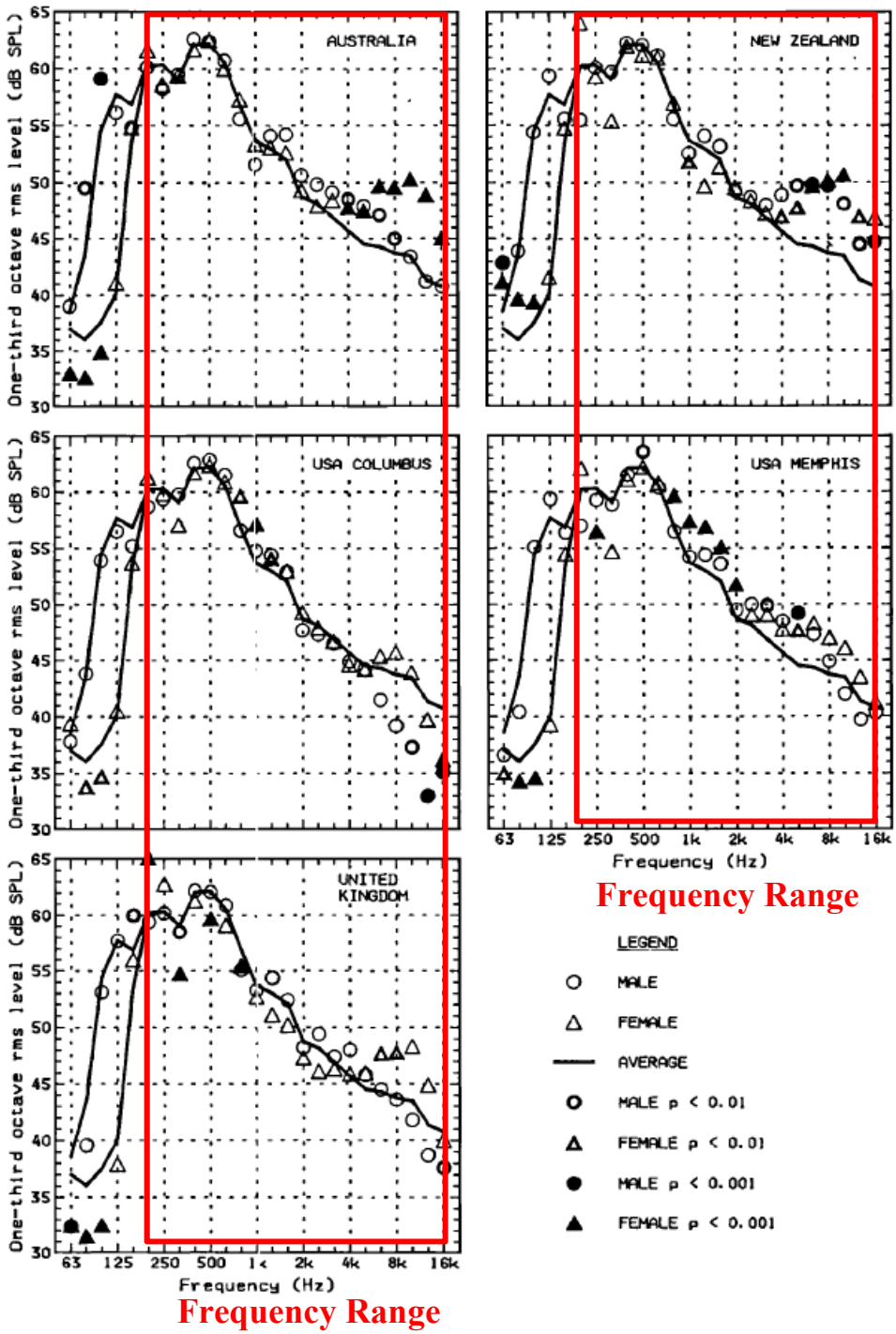



FIG. 1. Male and female long-term average speech spectrum (LTASS) values for five samples of English. Solid line shows LTASS average across 17 speech samples (all samples except Arabic), males and females separately for frequencies below 160 Hz, combined for higher frequencies.

**APPLE-1009, FIG. 1 (annotated)**

327. Further, as taught by Berglund, the spectral components of common noises experienced by passengers of "road transportation vehicles" is highest in the "low frequency noise" range (*e.g.,* a range of frequencies less than approximately 250 Hz"). *Id.,* 2985-2987, FIGS. 1, 3.

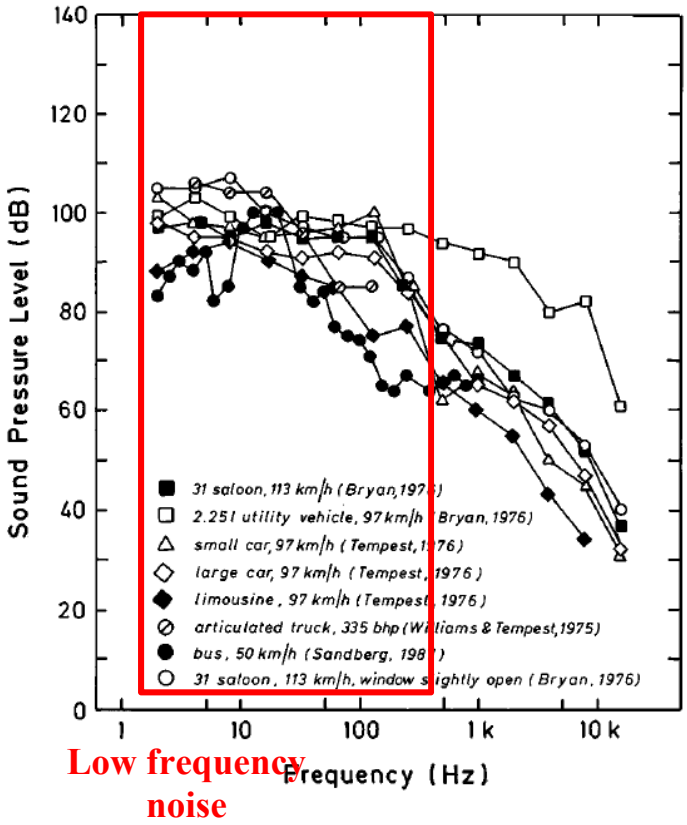



FIG. 3. Passenger noise exposure in road transportation vehicles as a function of frequency.

**APPLE-1013, FIG. 3 (annotated)**

328. The combination of Avendano and Burnett (and in the further combination with Visser, Bisgaard, and Hou) also discloses that the frequency subband includes frequencies higher than approximately 200 Hertz (Hz).

329. As I explained in Section X(A) (incorporated herein), a POSITA would have found it obvious to further combine Avendano, Visser, Bisgaard, and Hou with Burnett. In particular, a POSITA would have found it obvious to perform the processes described by Avendano, Visser, Bisgaard, and Hou specifically in frequency range(s) known to have "voiced and unvoiced speech" in order to improve the separation of "voiced and unvoiced speech from background acoustic noise." APPLE-1012. [0063]. As described by Burnett, these frequency ranges would include frequencies higher than approximately 200 Hz (*e.g.,* 500 Hz to 4000 Hz to identify the unvoiced speech "k" in "kick," 1700 Hz to 4000 Hz to identify the unvoiced speech "sh" in "she," 300 Hz to 2500 Hz to identify the voiced speech "ee," and 900 Hz to 1200 Hz to identify the voiced speech "ah"). APPLE-1012. [0064].

### C. Claim 21

*[21] The method of claim 19, wherein the frequency subband includes frequencies in a range from approximately 250 Hz to 1250 Hz.*

330. The combination of Avendano with Byrne, Burnett, and/or Berglund (and in the further combination with Visser, Bisgaard, and Hou) discloses that the frequency subband includes frequencies in a range from approximately 250 Hz to 1250 Hz.

331. I have been informed by counsel that, when evaluating whether a prior range anticipates a species, courts evaluate whether the claimed species is "critical." I have also been informed by counsel that where there is a range disclosed in the prior art, and the claimed invention falls within that range, the burden of production falls upon the patentee to come forward with evidence of teaching away, unexpected results, or other pertinent evidence of nonobviousness.

332. There is no evidence in the disclosure of the '611 patent that "a range from approximately 250 Hz to 1250 Hz" is critical. For example, in the specification of the '611 patent, the **only** explanation regarding a frequency range from 250 Hz to 1250 Hz reads as follows:

> The ratio R can be calculated for the entire frequency band of interest, or can be calculated in frequency subbands. One effective subband discovered was 250 Hz to 1250 Hz, another was 200 Hz to 3000 Hz, but <u>many others are possible and useful.</u>

APPLE-1001, 6:42-46 (emphasis added).

333. As the '611 patent concedes, a frequency range from 250 Hz and 1250 Hz is but one of "many" frequency ranges that are "possible" and "useful." *Id.* The '611 patent does not explain how using this frequency range would produce any new and unexpected results relative to the prior art. In fact, the '611 patent does not explain how using this frequency range would be more advantageous than using

any other specific frequency ranges or even "the entire frequency band of interest." Further, the '611 patent does not provide any guidance regarding the selection of "possible" and "useful" frequency ranges. Further still, nothing in the prosecution history of the '611 patent asserts any new and unexpected results relative to the prior art that are associated with this frequency range from 250 Hz to 1250 Hz.

334. Further still, the is no evidence in the disclosure or the prosecution history of the '611 patent asserting that the prior art taught away from using a frequency range from 250 Hz to 1250 Hz. Further still, there is no evidence in the disclosure or the prosecution history of the '611 patent regarding any other pertinent secondary considerations associated with a frequency range from 250 Hz to 1250 Hz.

335. Nevertheless, Avendano in combination with Byrne, Burnett, and/or Berglund (and in the further combination with Visser, Bisgaard, and Hou) discloses that the frequency subband includes frequencies in a range from approximately 250 Hz to 1250 Hz.

336. As I explained in Section X(A) (incorporated herein), a POSITA would have found it obvious to further combine Avendano, Visser, Bisgaard, and Hou with Byrne, Burnett, and/or Berglund. In particular, a POSITA would have found it obvious to perform the processes described by Avendano, Visser, Bisgaard, and

Hou specifically in frequency range(s) known to have significant spectral components of speech.

337. For example, as taught by Byrne, English language speech is expected to have significant spectral components in a frequency range between 250 Hz and 1250 Hz, as this frequency range would include the frequencies of sound expected to have the highest intensities for human speech (*e.g.,* frequencies of and around 500 Hertz). APPLE-1009, FIG. 1.

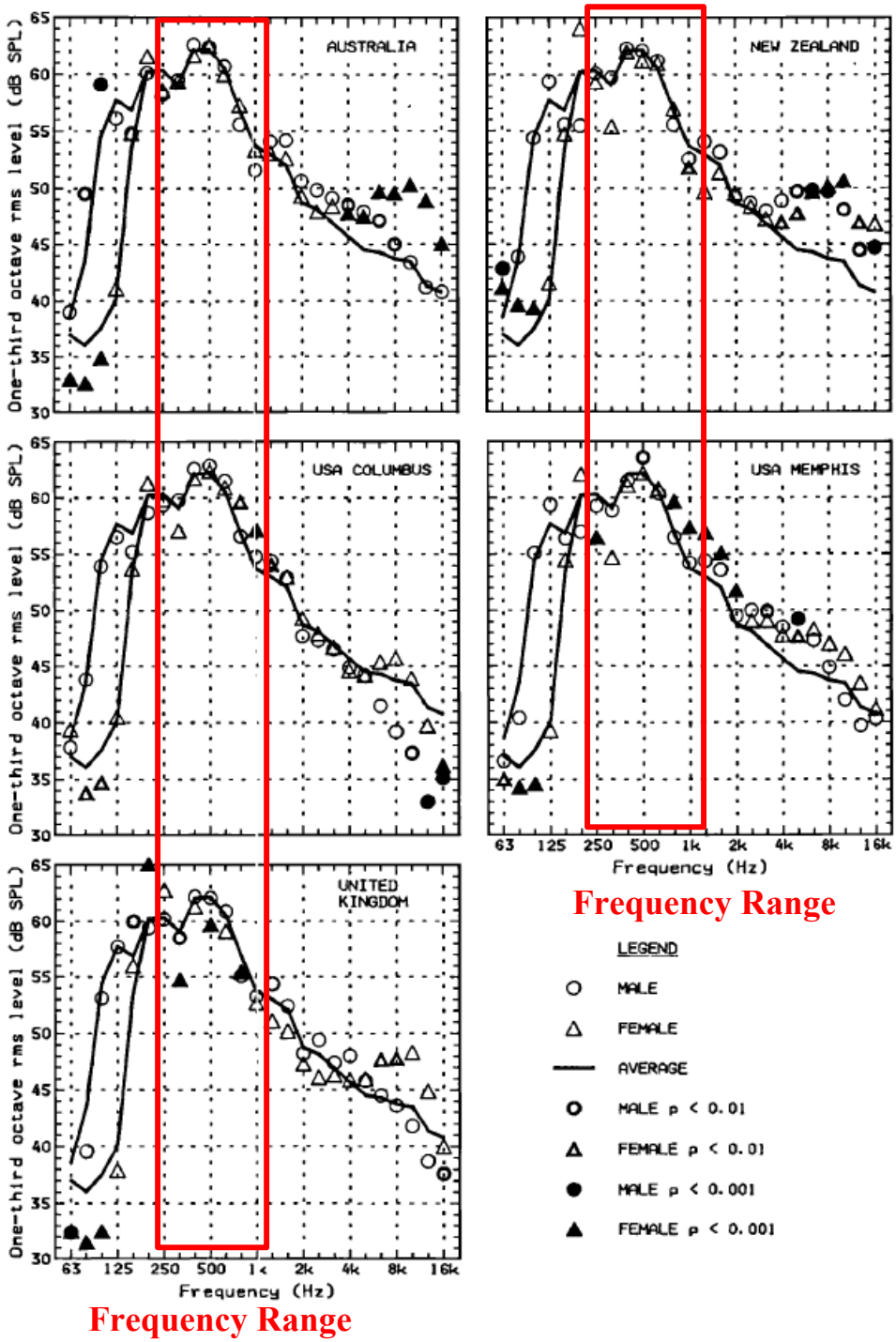



FIG. 1. Male and female long-term average speech spectrum (LTASS) values for five samples of English. Solid line shows LTASS average across 17 speech samples (all samples except Arabic), males and females separately for frequencies below 160 Hz, combined for higher frequencies.

**APPLE-1009, FIG. 1 (annotated)**

338. Accordingly, a POSITA would have found it obvious to perform the processes described by Avendano in a frequency subband that includes at least frequencies in a range from approximately 250 Hz to 1250 Hz to more readily detect the presence of human speech.

339. Further, I have been informed by counsel that, by virtue of the inclusive term "including," claim 21 does not preclude a frequency subband from including frequencies in addition to those between approximately 250 Hz to 1250 Hz. As discussed in Sections X(A)-(B), (incorporated herein), a POSITA would have found it obvious to perform the processes described by Avendano, Visser, Bisgaard, and Hou in a frequency subband that includes at least frequencies higher than approximately 200 Hz to more readily detect the presence of English language speech. Such a frequency subband would include frequencies in a range from approximately 250 Hz to 1250 Hz, among others.

340. Nevertheless, to the extent that claim 21 precludes a frequency subband from including frequencies other than those between approximately 250 Hz to 1250 Hz, a POSITA also would have found it obvious to perform the processes described by Avendano, Visser, Bisgaard, and Hou in such a frequency subband.

341. For example, as I explained in Section X(A) (incorporated herein), a POSITA would have found obvious to filter out frequency range(s) that are known

to have significant spectral components of noise, such as in a "low frequency noise" range less than approximately 250 Hz. *Id.,* 2985-2987, FIGS. 1, 3.

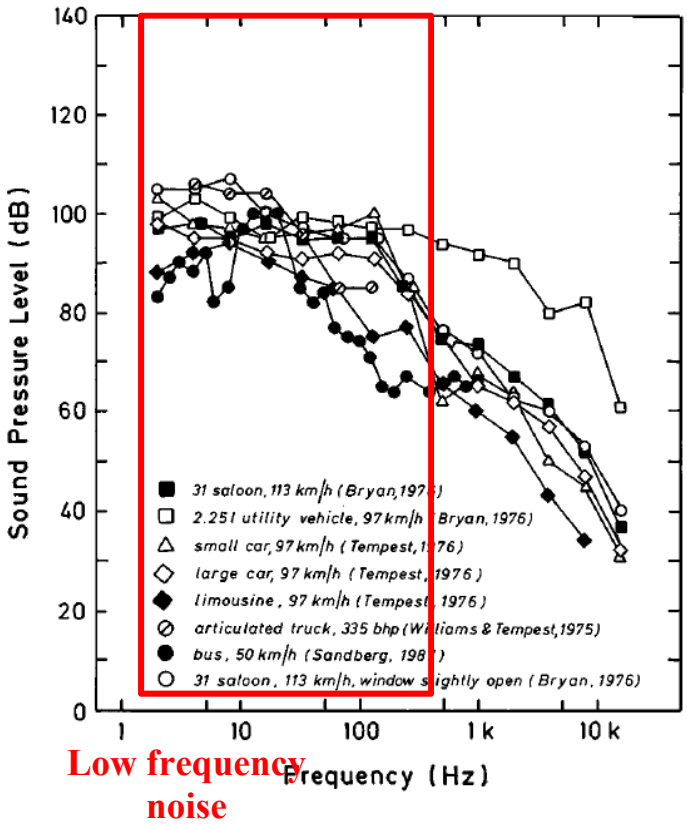



FIG. 3. Passenger noise exposure in road transportation vehicles as a function of frequency.

**APPLE-1013, FIG. 3 (annotated)**

342. Further, as I also explained in Section X(A), a POSITA would have found to obvious to tune the frequency range, depending on the specific use case at hand. In particular, a POSITA would have recognized that the time and/or computational resources required to perform the processes described by Avendano, Visser, Bisgaard, and Hou would depend on the range of frequencies that are analyzed. Accordingly, a POSITA would have recognized that the time and/or computation

resources required to perform these processes can be reduced by limiting analysis to a subset of all frequencies. As such, to improve the speed, accuracy, and/or efficiency in detecting voice activity, a POSITA would have found it obvious to prioritize analysis of frequency range(s) known to have significant spectral components of speech, such as a frequency range between approximately 250 Hz and 1250 Hz. A POSITA also would have found it obvious to use other frequency ranges, depending on the type of speech that is being detected (*e.g.,* the gender of the speaker, the spoken language, etc.), the type of noise that is being filtered out, and the expected spectral distributions thereof.

343. As discussed above, there is no evidence in the disclosure of the '611 patent that "a range from approximately 250 Hz to 1250 Hz" is critical. However, to the extent that the Patent Owner contends that the '611 patent's description of frequency ranges of certain "voiced and unvoiced speech" (*i.e.,* APPLE-1001, 17:18-36) demonstrates the criticality of the claimed range, I respectfully disagree, as the description does not correspond to the claimed range. For example, the '611 patent describes that a system can use bandpass filters to accept signals in various certain frequency ranges to identify the presence of unvoiced speech (*e.g.*, 500 Hz to 4000 Hz to identify the unvoiced speech "k" in "kick," and 1700 Hz to 4000 Hz to identify the unvoiced speech "sh" in "she"), and use bandpass filters to accept other frequency ranges to identify the presence of voiced speech (e.g., 300 Hz to

2500 Hz to identify the voiced speech "ee," and 900 Hz to 1200 Hz to identify the voiced speech "ah"). *Id.,* 17:29-36. However, none of these frequency ranges corresponds to the claimed frequency range of 250 Hz to 1250 Hz. In fact, the claimed frequency range would arbitrarily include the entirety of the frequency range for some types of speech (*e.g.,* 900 Hz to 1200 Hz, corresponding to "ah"), exclude a portion of the frequency range for other types of speech (*e.g.*, 300 Hz to 2500 Hz, corresponding to "ee," and 500 Hz to 4000 Hz, corresponding to "k" in "kick"), and exclude the entirety of the frequency range for yet another type of speech (*e.g.*, 1700 Hz to 4000 Hz, corresponding to "sh" in "she"). The '611 patent does not describe any criticality associated with these arbitrarily selected frequencies. Nor does any criticality exist for any of these arbitrarily selected frequencies.

344. However, even if this description in the '611 patent were to demonstrate the criticality of the claimed range, Burnett nevertheless includes this same description verbatim. APPLE-1012, [0063]-[0064]. As such, the claimed range would still be disclosed by Burnett.

### D. Claim 22

*[22] The method of claim 19, wherein the frequency subband includes frequencies in a range from approximately 200 Hz to 3000 Hz.*

345. The combination of Avendano with Byrne, Burnett, and/or Berglund (and in the further combination with Visser, Bisgaard, and Hou) discloses that the frequency subband includes frequencies in a range from approximately 200 Hz to 3000 Hz.

346. As I have been informed by counsel that, when evaluating whether a prior range anticipates a species, courts evaluate whether the claimed species is "critical." I have also been informed by counsel that where there is a range disclosed in the prior art, and the claimed invention falls within that range, the burden of production falls upon the patentee to come forward with evidence of teaching away, unexpected results, or other pertinent evidence of nonobviousness.

347. There is no evidence in the disclosure of the '611 patent that "a range from approximately 200 Hz to 3000 Hz" is critical. For example, in the specification of the '611 patent, the **only** explanation regarding a frequency range from 200 Hz to 3000 Hz reads as follows:

> The ratio R can be calculated for the entire frequency band of interest, or can be calculated in frequency subbands. One effective subband discovered was 250 Hz to 1250 Hz, another was 200 Hz to 3000 Hz, but <u>many others are possible and useful.</u>

APPLE-1001, 6:42-46 (emphasis added).

348. As the ’611 patent concedes, a frequency range from 200 Hz and 3000 Hz is but one of “many” frequency ranges that are “possible” and “useful.” *Id.* The ’611 patent does not explain how using this frequency range would produce any new and unexpected results relative to the prior art. In fact, the ’611 patent does not explain how using this frequency range would be more advantageous than using any other specific frequency ranges or even “the entire frequency band of interest.” Further, the ’611 patent does not provide any guidance regarding the selection of “possible” and “useful” frequency ranges. Further still, nothing in the prosecution history of the ’611 patent asserts any new and unexpected results relative to the prior art that are associated with this frequency range from 200 Hz to 3000 Hz.

349. Further still, the is no evidence in the disclosure or the prosecution history of the ’611 patent asserting that the prior art taught away from using a frequency range from 200 Hz to 3000 Hz. Further still, there is no evidence in the disclosure or the prosecution history of the ’611 patent regarding any other pertinent secondary considerations associated with a frequency range from 300 Hz to 3000 Hz.

350. Nevertheless, Avendano in combination with Byrne, Burnett, and Berglund (and in the further combination with Visser, Bisgaard, and Hou) discloses that the frequency subband includes frequencies in a range from approximately 200 Hz to 3000.

351. As discussed in Section X(A) (incorporated herein), a POSITA would have found it obvious to further combine Avendano, Visser, Bisgaard, and Hou with Byrne, Burnett, and Berglund. In particular, a POSITA would have found it obvious to perform the processes described by Avendano, Visser, Bisgaard, and Hou specifically in frequency range(s) known to have significant spectral components of speech.

352. For example, as taught by Byrne, human speech is expected to have significant spectral components in a frequency range between 200 Hz and 3000 Hz, as such a frequency range would include the frequencies of sound expected to have the highest intensities for human speech (*e.g.,* frequencies of and around 500 Hertz). APPLE-1009, FIG. 1.

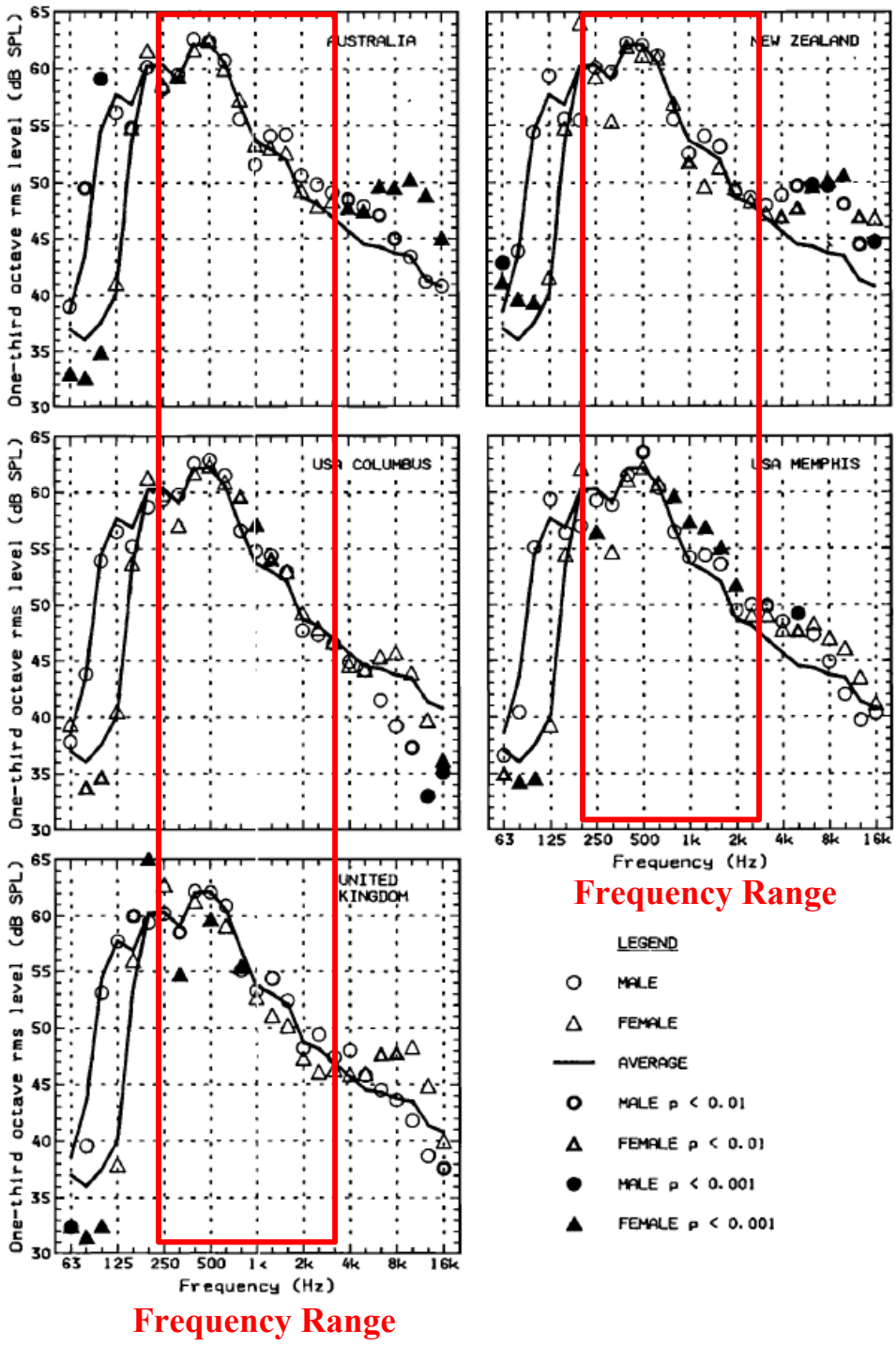



FIG. 1. Male and female long-term average speech spectrum (LTASS) values for five samples of English. Solid line shows LTASS average across 17 speech samples (all samples except Arabic), males and females separately for frequencies below 160 Hz, combined for higher frequencies.

**APPLE-1009, FIG. 1 (annotated)**

353. Accordingly, a POSITA would have found it obvious to perform the processes described by Avendano in a frequency subband that includes at least frequencies in a range from approximately 200 Hz to 3000 Hz to more readily detect the presence of human speech.

354. Further, I have been informed by counsel that, by virtue of the inclusive term "including," claim 22 does not preclude a frequency subband from including frequencies in addition to those between approximately 200 Hz to 3000 Hz. As I explained in Sections X(A)-(B), (incorporated herein), a POSITA would have found it obvious to perform the processes described by Avendano, Visser, Bisgaard, and Hou in a frequency subband that includes at least frequencies higher than approximately 200 Hz to more readily detect the presence of English language speech. Such a frequency subband would include frequencies in a range from approximately 200 Hz to 3000 Hz, among others.

355. Nevertheless, to the extent that claim 22 precludes a frequency subband from including frequencies other than those between approximately 200 Hz to 3000 Hz, a POSITA also would have found it obvious to perform the processes described by Avendano, Visser, and Bisgaard in such a frequency subband.

356. For example, as I explained in Section X(A) (incorporated herein), a POSITA would have found obvious to filter out frequency range(s) that are known

to have significant spectral components of noise, such as in a "low frequency noise" range less than approximately 250 Hz. *Id.,* 2985-2987, FIGS. 1, 3.

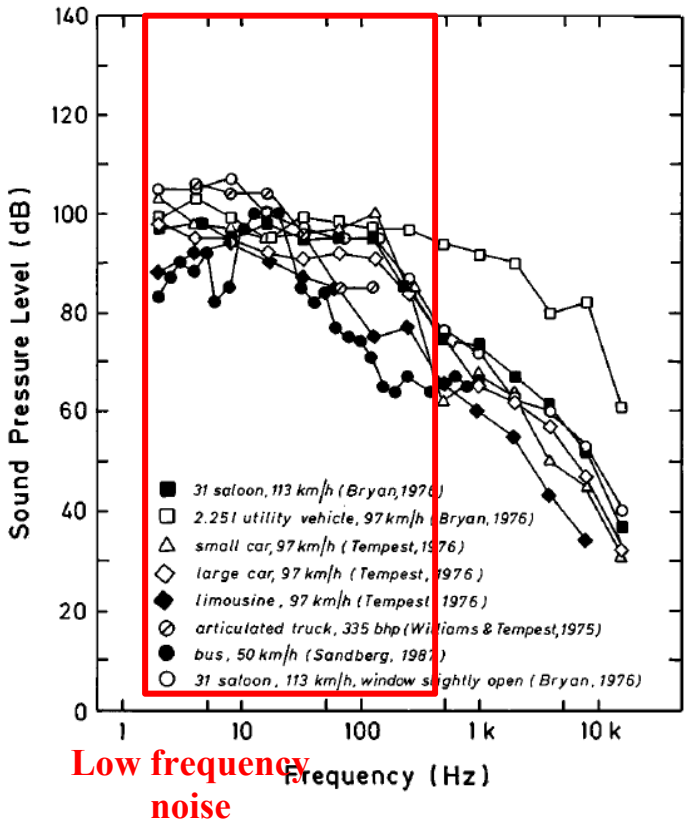



FIG. 3. Passenger noise exposure in road transportation vehicles as a function of frequency.

**APPLE-1013, FIG. 3 (annotated)**

357. Further, as I also explained in Section described in Section X(A), a POSITA would have found to obvious to tune the frequency range, depending on the specific use case at hand. In particular, a POSITA would have recognized that the time and/or computational resources required to perform the processes described by Avendano, Visser, Bisgaard, and Hou would depend on the range of frequencies that are analyzed. Accordingly, a POSITA would have recognized that the time

and/or computation resources required to perform these processes can be reduced by limiting analysis to a subset of all frequencies. As such, to improve the speed, accuracy and/or efficiency in detecting voice activity, a POSITA would have found it obvious to prioritize analysis of frequency range(s) known to have significant spectral components of speech, such as a frequency range between approximately 200 Hz and 3000 Hz. A POSITA also would have found it obvious to use other frequency ranges, depending on the type of speech that is being detected (*e.g.,* the gender of the speaker, the spoken language, etc.), the type of noise that is being filtered out, and the expected spectral distributions thereof.

358. As discussed above, there is no evidence in the disclosure of the '611 patent that "a range from approximately 200 Hz to 3000 Hz" is critical. However, to the extent that the Patent Owner contends that the '611 patent's description of frequency ranges of certain "voiced and unvoiced speech" (*i.e.,* APPLE-1001, 17:18-36) demonstrates the criticality of the claimed range, I respectfully disagree, as the description does not correspond to the claimed range. For example, the '611 patent describes that a system can use bandpass filters to accept certain frequency ranges to identify the presence of unvoiced speech (*e.g.*, 500 Hz to 4000 Hz to identify the unvoiced speech "k" in "kick," and 1700 Hz to 4000 Hz to identify the unvoiced speech "sh" in "she"), and use bandpass filters to accept other frequency ranges to identify the presence of voiced speech (e.g., 300 Hz to 2500 Hz to

identify the voiced speech "ee," and 900 Hz to 1200 Hz to identify the voiced speech "ah"). *Id.,* 17:29-36. However, none of these frequency ranges corresponds to the claimed frequency range of 200 Hz to 3000 Hz. In fact, the claimed frequency range would arbitrarily include the entirety of the frequency range for some types of speech (*e.g.,* 900 Hz to 1200 Hz, corresponding to "ah", and 300 Hz to 2500 Hz, corresponding to "ee"), exclude a portion of the frequency range for other types of speech (*e.g.*, 500 Hz to 4000 Hz, corresponding to "k" in "kick," and 1700 Hz to 4000 Hz, corresponding to "sh" in "she"). The '611 patent does not describe any criticality associated with these arbitrarily selected frequencies. Nor does any criticality exist for any of these arbitrarily selected frequencies.

359. However, even if this description in the '611 patent were to demonstrate the criticality of the claimed range, Burnett nevertheless includes this same description verbatim. APPLE-1012, [0063]-[0064]. As such, the claimed range would still be disclosed by Burnett.

## XI. CONCLUSION

360. I reserve the right to supplement my opinions to address any information obtained, or positions taken, based on any new information introduced throughout this proceeding.

361. I declare under penalty of perjury that the foregoing is true and accurate to the best of my ability.