UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

META PLATFORMS, INC., SONY ELECTONICS INC., and
LG ELECTRONICS INC.,

Petitioners,

v.

JAWBONE INNOVATIONS, LLC,

Patent Owner.

Patent No. 8,321,213
Filing Date: October 26, 2009
Issue Date: November 27, 2012

Inventor: Nicolas Petit, Gregory Burnett, and Zhinian Jing
Title: ACOUSTIC VOICE ACTIVITY DETECTION (AVAD)
FOR ELECTRONIC SYSTEMS

DECLARATION OF AKBAR M. SAYEED, Ph.D.

Case No. IPR2023-01084

TABLE OF CONTENTS

Page(s)

I. INTRODUCTION 2

A. Background and Qualifications 2

II. COMPENSATION 6

III. MATERIALS CONSIDERED 7

IV. LEGAL PRINCIPLES 7

V. LEVEL OF SKILL IN THE ART 8

VI. THE CLAIMED INVENTION OF THE ’213 PATENT 9

VII. CLAIM CONSTRUCTION 13

VIII. THE ALLEGED PRIOR ART 20

A. Avendano ’880 (U.S. Patent No. 8,194,880) 20

B. Visser (U.S. Patent No. 7,464,029) 23

C. Avendano/Visser Combination 27

IX. PETITIONER’S COMBINATIONS DO NOT RENDER ANY CLAIM OBVIOUS 29

A. GROUND 1: The Combination of Avendano and Visser Does Not Render Obvious Claims 14-17, 35-42 29

1. Petitioner’s Combination Does Not Disclose or Render Obvious “the processing component forming a filter that describes a relationship for speech between the first physical microphone and the second physical microphone” as Required by All Challenged Claims 29

i. Avendano's Delay Node Does Not Render Obvious The Claimed Filter 30

ii. A POSITA Would Not Have Been Motivated to Make the Avendano/Visser Combination 31

2. The Avendano/Visser combination does not render claim 17 obvious because they do not disclose or render obvious "applying a delay to the first intermediate signal" 33

B. GROUND 2: Avendano, Visser, and Bisgaard Do Not Render Obvious Claims 18-26, 33, and 34 36

1. The Avendano/Visser/Bisgaard Combination does not remedy Ground 1's deficiencies as to Claim 17 36

2. The Avendano/Visser/Bisgaard Combination does not disclose or render obvious claim 19. 37

C. GROUND 3: Avendano, Visser, Bisgaard, and Hou Do Not Render Obvious Claims 27-29 40

D. GROUND 4: Avendano, Visser, Bisgaard, Hou, and "Frequency Art" Do Not Render Obvious Claims 30-32 40

X. CONCLUSION 41

I, Akbar M. Sayeed, declare as follows:

1. I have been asked by counsel for Patent Owner, Jawbone Innovations, LLC ("Jawbone" or "Patent Owner"), to review U.S. Patent No. 8,321,213 (the "'213 Patent") entitled ACOUSTIC VOICE ACTIVITY DETECTION (AVAD) FOR ELECTRONIC SYSTEMS, and to provide my technical review, analysis, insights, and opinions regarding the '213 Patent in view of the prior art cited by Petitioner Meta Platforms, Inc. ("Meta" or "Petitioner"). I submit this declaration in support of Patent Owner's Response in this IPR proceeding. I have personal knowledge of the matters stated herein and would be competent to testify to them if required.

2. I submitted a previous declaration in connection with Jawbone's Preliminary Response. While this declaration takes into account issues raised by the Board and by Meta in its Preliminary Reply, I maintain the opinions in my previous declaration.

3. I have been retained on behalf of Jawbone for the above-captioned *inter partes* review proceeding. I understand that the '213 Patent is currently assigned to Jawbone.

4. I am over 18 years of age. I have personal knowledge of the facts stated in this Declaration and could testify competently if asked to do so.

I. INTRODUCTION

A. Background and Qualifications

5. I have reviewed and am familiar with the specification of the ’213 Patent. I understand the ’213 Patent has been provided as Exhibit 1001. I will cite to the specification using the following format: ’213 Patent at col.:line.

6. My CV was previously submitted as Ex. 2003.

7. I received my B.S. degree in Electrical and Computer Engineering (ECE) from the University of Wisconsin-Madison in 1991. I received my M.S. and Ph.D. degrees in ECE from the University of Illinois at Urbana-Champaign in 1993 and 1996, respectively. In my undergraduate and graduate studies, I look a variety of courses generally focusing on the areas of signal processing, communication theory, information theory, electromagnetics and antennas, and statistical techniques in signal processing and communications. My Ph.D. dissertation was entitled “Statistical Time-Frequency Analysis” in which I proposed a new framework for statistical signal processing using time-frequency representations, mathematical tools that extend the powerful theory of Fourier transforms and Fourier analysis to time-varying signals and systems.

8. I currently work as an Independent Researcher, Engineer and Technical Consultant, and worked as professor of Electrical and Computer Engineering at the University of Wisconsin-Madison from 1997-2021, where I directed the *Wireless*

Communications and Sensing Laboratory until my retirement on August 1, 2021 to pursue a career as an independent researcher and consultant.

9. My current work as an independent researcher and consultant spans STEM (science, technology, engineering & mathematics) fields through the lens of information science and technology. My STEM skills and experience cover a broad range of areas, including:

- Wireless technologies for sensing, processing and communication of information (5G/6G/XG), internet of things, and emerging technologies for untapped spectrum (e.g., millimeter-wave, THz).
- Sensing and acquisition of data in new modalities through technological innovation.
- Machine learning and statistical techniques for extracting useful information from data.
- Quantum information science and technology (computing, sensing and communication) spanning foundational concepts, algorithms, and platforms.
- Conception, design and development of new information technologies through basic theory, computational modeling, and hybrid software-hardware prototyping and experimentation.

10. I have written/co-written 200+ papers in leading journals and conferences. I am an inventor/co-inventor of 10 patents.

11. After receiving my Ph.D. in 1996, I spent a year at Rice University as a postdoctoral research fellow where I expanded the scope of my research to the growing field of wireless communications that underpins much of mobile cellular technology. In particular, the propagation channels encountered in mobile cellular communication are inherently time-varying in nature and that provided a natural connection to my Ph.D. research. After spending a year at Rice University, I accepted a position as an Assistant Professor in the ECE Department at the University of Wisconsin-Madison in 1997. I was granted tenure and promoted to Associate Professor in 2003, and was promoted to Professor in 2008.

12. I founded and directed the Wireless Communication and Sensing Laboratory at UW-Madison and engaged in research, teaching and innovation in the basic science, engineering, and technology of sensing, processing, and communication of information. In particular, I led a team in the conception, design, development and successful demonstration of a state-of-the-art prototype for multi-beam MIMO (multiple input multiple output) communication at 28 GHz based on the concept of a lens array that I had pioneered.

13. I co-founded and co-led the NSF-sponsored Research Coordination Network (RCN) on millimeter-wave wireless (2016-2019) for bringing together

researchers from academia, industry, and government agencies to address critical technical challenges in the area.

14. From 2017-2019 I served as a Program director in the Electrical, Communications, and Cyber Systems (ECCS) Division of the Directorate of Engineering at the National Science Foundation. At the NSF, I managed existing programs and initiated new research programs involving sensing, processing, and communication of information in all possible physical modalities, architectures, and technologies, including quantum information science and engineering. I actively contributed to three NSF-wide working groups: Electromagnetic Spectrum Management (ESM) working group, and the working groups of two of the 10 NSF Big Ideas: i) Quantum Leap and ii) Harnessing the Data Revolution. In particular, I co-led the team that developed the NSF solicitation for Quantum Leap Challenge Institutes (Feb 2019) in response to the National Quantum Initiative (NQI) Act signed in December 2018.

15. I have a demonstrated history (25 yrs.) of leading an interdisciplinary research and development group at the forefront of foundational and technological innovations in wireless communication and sensing. Physics-based accurate and computationally tractable system modeling has been a key underlying theme in my work, spanning communication & sensor networks, multipath propagation over

highly dynamic environments, wideband MIMO systems, mm-wave and THz wireless, and prototype development (hardware & software).

16. My technical background, experience and skills are directly relevant to the subject matter in this case. The modeling, analysis and processing of information bearing signals and waves, including transmission and reception through antennas and antenna arrays, and propagation over channels, has been an underlying theme throughout my career. In particular, I have extensive experience in the conception, modeling, development and analysis of techniques and algorithms for processing and filtering of signals from an array of sensors or antennas, including beamforming, interference and noise suppression, and adaptive filtering methods. While most of my work has focused on electromagnetic signals, the underlying physics and mathematics are very similar to acoustic signals.

II. COMPENSATION

17. My compensation for time worked on this proceeding is not dependent on any issues related to the ’213 Patent, the outcome of this proceeding, or the substance of my opinions. My compensation for time worked on this proceeding is at my customary rate of $200 per hour. I have no financial interest in, or affiliation with, the Patent Owner or any of the real parties in interest.

III. MATERIALS CONSIDERED

18. In providing my technical review, analysis, insights, and opinions, I have considered the ’213 Patent and its prosecution history.

19. I have also considered the Petition filed by the Petitioner in this proceeding and the relevant exhibits relied on by Petitioner, including the expert declaration submitted by Dr. Thomas Kenny (“Ex. 1003”). I have further considered and reviewed the transcript of Dr. Kenny’s deposition (“Ex. 2004”).

20. I have also considered my own experience and knowledge, as discussed above and described more fully in my CV, in areas including signal processing, electrical engineering, and other related technologies.

IV. LEGAL PRINCIPLES

21. I understand that a patent claim is unpatentable as “obvious” if the subject matter of the claim as a whole would have been obvious to a person of ordinary skill in the art (POSITA) as of the time of the invention at issue.

22. I understand that the use of “the person of ordinary skill” rubric is to prevent one from improperly, in the present day, using hindsight to decide whether a claim is obvious.

23. I understand that the following factors must be evaluated to determine whether the claimed subject matter is obvious: (1) the scope and content of the prior art; (2) the difference or differences, if any, between the scope of the patent claim

and the scope of the prior art; and (3) the level of ordinary skill in the art at the time of the invention.

24. I understand that certain secondary considerations, such as commercial success, skepticism of experts, surprise, and copying, may provide evidence of non-obviousness. I further understand that such considerations are often the most probative and determinative of obviousness or non-obviousness.

25. I understand that I must construe a claim in accordance with the ordinary and customary meaning of the language of such claim as understood by one of ordinary skill in the art and the prosecution history pertaining to the patent.

V. LEVEL OF SKILL IN THE ART

26. I understand that I should perform my analysis from the viewpoint of a person of ordinary skill in the art. I understand that this hypothetical person of ordinary skill in the art is considered to have the normal skills of a person in a certain technical field. I understand that factors that may be considered in determining the level of ordinary skill in the art include: (1) the education level of the inventor; (2) the types of problems encountered in the art; (3) the prior art solutions to those problems; (4) rapidity with which innovations are made; (5) the sophistication of the technology; and (6) the education level of active workers in the field.

27. In my opinion, a person of ordinary skill in the art would have a bachelor's degree in computer engineering, computer science, electrical

engineering, mechanical engineering, or a similar field, and approximately three years of industry or academic experience in a field related to acoustics, speech recognition, speech detection, or signal processing. Work experience can substitute for formal education and additional formal education can substitute for work experience.

28. My opinions would not change if the Board adopts Dr. Kenny's level of skill in the art.

VI. THE CLAIMED INVENTION OF THE '213 PATENT

29. I have reviewed the '213 Patent, entitled ACOUSTIC VOICE ACTIVITY DETECTION (AVAD) FOR ELECTRONIC SYSTEMS (Ex. 1001). It issued on November 27, 2012, and was filed on October 26, 2009.

30. The '213 Patent discloses and claims systems and devices for acoustic voice activity detection in microphone arrays. The claimed inventions "use microphones to generate virtual directional microphones which have very similar noise responses and very dissimilar speech responses." Furthermore, the "ratio of the energies of the virtual microphones is then calculated over a given window size and the ratio can then be used with a variety of methods to generate a VAD signal. The virtual microphones can be constructed using either an adaptive or a fixed filter". Ex. 1001 Abstract.

31. The Patent explains that the claimed acoustic voice activity detection is based on Jawbone's PSAD (Pathfinder Speech Activity Detection) and NAVSAD (Non-Acoustic Sensor Voiced Speech Activity Detection) systems. '213 Patent, 4:27-30, 10:12-52; 11:1-29. The specification describes that one of the key advances of the technology is using the "difference in frequency content between two microphones to calculate a relationship between the signals of the two microphones." *Id.*, 11:20-23. The prior art, in contrast, "attempt[s] to use the time/phase difference of each microphone to remove the noise outside of an 'area of sensitivity.'" *Id.*, 11:23-26.

32. Figures 4 and 3 of the '213 Patent provide illustrative embodiments for creating the virtual microphone signals, V_1 and V_2, respectively, from two omnidirectional physical microphone signals (O_1 and O_2). Each embodiment includes: i) a calibration filter (α), ii) a filter that describes a relationship for speech (β), and iii) a delay filter (z).

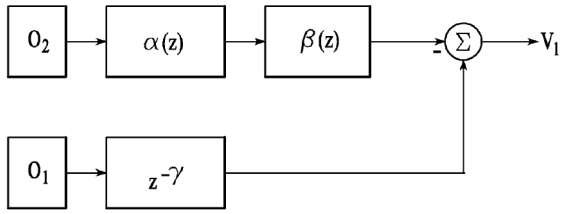


FIG.4

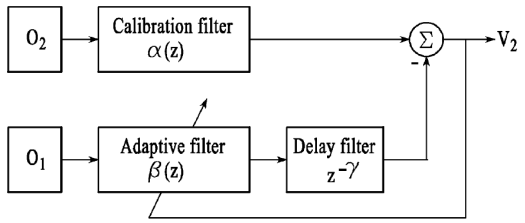


FIG.3

'213 Patent, Figs. 3, 4. '213 Patent, 5:31-35 ("α(z) is a calibration filter used to compensate O_2's response so that it is the same as O_1, β(z) is a filter that describes the relationship between O_1 and calibrated O_2 for speech, and γ is a fixed delay that depends on the size of the array.").

33. The '213 Patent explains that "It is assumed that O_1 is closer to the speech source (the user's mouth)" *Id*. 4:64-65. In the above Figures α(z) and β(z) represent filters, where β(z) can be fixed or adaptive *Id*. 6:17-18.

34. The relationship of V1 and V2 can be described in the digital frequency domain (z variable) as:

$$V_1(z) = -\beta(z)\alpha(z)O_2(z) + O_1(z)z^{-\gamma}$$

$$V_2(z) = \alpha(z)O_2(z) - \beta(z)O_1(z)z^{-\gamma}$$

Id. 5:23-35. In the above equations (and Fig. 3 and Fig. 4 from Ex 1001), "α(z) is a calibration filter used to compensate O_2's response so that it is the same as O_1, β(z) is a filter that describes the relationship between O_1 and calibrated O_2 for speech, and Y is a fixed delay that depends on the size of the array" *Id*. 5:29-33.

35. Ideally the microphones O_1 and O_2 are assumed to be calibrated so that "their response to a source located the same distance away was identical for both amplitude and phase." Ex 1001, 28:65-67. However, this is not always feasible in practice and the calibration filter α(z) can be determined using adaptive algorithms, for example, so that in the presence of speech only signals (with no or low levels of noise) the microphone signals O_1 and O_2 satisfy the relationship $O_1(z) = \alpha(z)\, O_2(z)$. Ex. 1001 28:64-29:31.

36. After the calibration filter α(z) is determined, the filter β(z) can be determined, either from the geometry of microphones relative to the target speech source, or, in the claimed embodiments, by using an adaptive filter algorithm to satisfy the relationship:

$$\tilde{\beta}(z) = \frac{a(z)O_2(z)}{z^{-v}O_1(z)}$$

Ex. 1001 5:53-6:11.

37. The Voice Activity Detection (VAD) is based on computing the ratio R:

$$R = \frac{\|V_1(z)\|}{\|V_2(z)\|} = \sqrt{\frac{\left(-\tilde{\beta}(z)\alpha(z)O_2(z) + O_1(z)z^{-v}\right)^2}{\left(\alpha(z)O_2(z) - \tilde{\beta}(z)O_1(z)z^{-v}\right)^2}}$$

where "double bar indicates norm and again any size window may be used. If $\tilde{\beta}(z)$ has been accurately calculated, the ratio for speech should be relatively high (e.g., greater than approximately 2) and the ratio for noise should be relatively low (e.g., less than approximately 1.1)." Ex. 1001 6:20-32. If the ratio R is above a threshold, speech is present, otherwise speech is absent. Ex 1001, Fig. 5 (510). "Many techniques known to those skilled in the art (e.g., smoothing, etc.) can be used to make R more amenable to use in generating a VAD and the embodiments herein are not so limited." Ex. 1001, 6:38-41.

VII. CLAIM CONSTRUCTION

38. I understand that claim terms are to be given their plain and ordinary meaning as a POSITA would have understood them in light of the specification and prosecution history.

39. In my previous declaration, I opined that the term "filter that describes a relationship for speech between the first physical microphone and the second physical microphone" would have been understood by a POSITA at the time of the invention as "A filter that alters the magnitude of a signal based on the relative responses to speech of the first and second physical microphones."

40. While that statement remains true, I provide here a revised construction that focuses on the key dispute between Meta and Jawbone: whether a POSITA would understand the "filter that describes a relationship for speech between the first

physical microphone and the second physical microphone" (hereinafter, "relationship filter") to encompass a delay filter, such as the delay filter described in the '213 patent. In my opinion, a POSITA would understand that the relationship filter must do more than merely delay a signal.

41. Accordingly, in my opinion the term "filter that describes a relationship for speech between the first physical microphone and the second physical microphone" would have been understood by a POSITA at the time of the invention as "A filter that alters a signal based on the non-temporal components of relative responses to speech of the first and second physical microphones." This construction is consistent with the plain and ordinary meaning of the term as used in the specification.

42. The specification explains that one of the invention's major advances over the prior art is to use "the difference in frequency content" not "time/phase difference" to calculate the relationship for speech between the microphones:

> The NAVSAD and PSAD systems include an array algorithm for speech detection that uses the difference in frequency content between two microphones to calculate a *relationship between the signals of the two microphones*. This is in contrast to conventional arrays that attempt to use the ***time/phase difference*** of each microphone to remove the noise outside of an "area of sensitivity". *The methods described herein provide a significant*

advantage, as they do not require a specific orientation of the array with respect to the signal.

'213 Patent, 11:20-29. As explained above, the claims are intended to cover the PSAD and NAVSAD systems. '213 Patent, 4:27-30, 10:12-52; 11:1-29.

43. A POSITA reading this statement would understand that a key advance and advantage of the patent is to move away from a delay-based method, which is the "time/phase difference" disparaged above. As the patent explains, there are several advantages in moving away from delay-based processing.

> Further, the systems described herein are sensitive to noise of every type and every orientation, unlike conventional arrays that depend on specific noise orientations. Consequently, the frequency-based arrays presented herein are unique as they depend only on the relative orientation of the two microphones themselves with no dependence on the orientation of the noise and signal with respect to the microphones. This results in a robust signal processing system with respect to the type of noise, microphones, and orientation between the noise/signal source and the microphones.

11:30-39.

44. For example, "[t]he systems described herein are sensitive to noise of every type and every orientation, unlike conventional arrays that depend on specific noise orientations." '213 Patent, 11:30-39. "This results in a robust signal processing system with respect to the type of noise, microphones, and orientation between the noise/signal source and the microphones." *Id.*

45. To achieve its goal of improving on time/phase processing, the specification describes three different filters, a calibration filter (α(z)), a delay filter ($z^{-\gamma}$), and the filter that describes a relationship for speech (β(z)). '213 Patent 5:20-51, Fig. 3, Fig. 4. The delay γ is "a fixed delay that depends on the size of the array." '213 Patent, 5:32-33.

46. Figure 3 illustrates the relationship for V2 (and Fig. 4 illustrates it for V1):

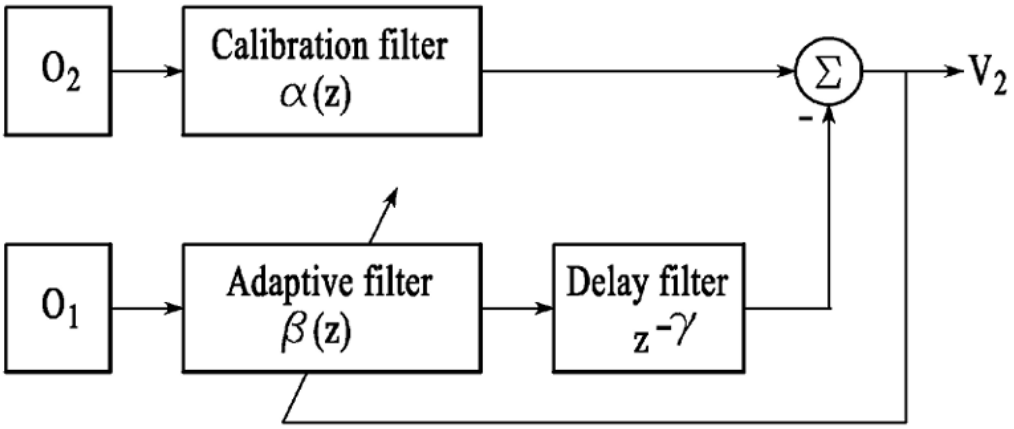


FIG.3

47. The specification further specifically identifies β(z) as the filter describing the relationship for speech. '213 Patent, 5:31-32 ("β(z) is a filter that describes the relationship between O_1 and calibrated O_2 for speech"). The specification provides two methods to calculate β(z): 1) as a function of the distance

between each microphone and the user's mouth ('213 Patent, 5:52-59); and 2) adaptively (*id.*, 6:1-16).

$$\beta(z) = \frac{d_1}{d_2} = \frac{d_1}{d_1 + d} \qquad \tilde{\beta}(z) = \frac{a(z)O_2(z)}{z^{-\nu}O_1(z)}$$

48. In either case, the β(z) compensates for the difference in response to speech of the two microphones, not just the temporal difference. By increasing or decreasing the magnitude of the filtered signal, for example, as a function of frequency content, the filter adjusts the signal to account for the relative strengths of the speech signal received over the range of frequencies of interest—i.e., the closer microphone is attenuated to the same speech level received at the farther microphone and this attenuation may change as a function of frequency content.

49. Indeed, the separate delay, which is defined in terms of distance and the speed of sound, compensates for the temporal difference based on the distance between the microphones. '213 Patent, 5:35-45.

$$\gamma = \frac{d}{c}$$

50. The constant d is the distance between the two microphones. '213 Patent, 5:43-44; 4:66-67 ("If the microphones and the user's mouth are all on a line,

then $d=2d_0$, the distance between the microphones."). The constant c is "the speed of sound in air." *Id.*, 5:43-44.

51. α is similarly contrasted with the relationship filter. '213 Patent 5:20-51, 6:1-16, Fig. 3, Fig. 4.

52. Every embodiment disclosed in the specification includes a β filter. The patent spends significant time and space on explaining how to calculate and utilize the β filter. '213 Patent, 5:53-6:41, 7:8-8:38 (explaining how to implement adaptation in the β filter), 8:39-10:8 (describing experimental results for β filters). For example, when calculating the ratio R (which is used to actually determine if speech is present), the patent emphasizes that β's accuracy is critical: "it is important to use the adapted filter β(z) in $V_2(z)$ for best performance." *Id.*, 6:37-39. Similarly, the specification further emphasizes: "The accuracy of the adaptation to the β(z) of the system is a factor in determining the effectiveness of the AVAD. A more accurate adaptation to the actual β(z) of the system leads to lower energy of the speech response in V_2, and a higher ratio R." *Id.*, 7:8-18. Every embodiment of the claimed system includes a β filter.

53. In contrast, the specification spends relatively little space on the α and delay filters. In the voice activity detection context, the discussion is centered on how to define β so as to overcome poor calibration. *Id.*, 5:20-34, 5:61-63, 7:59-64. The remaining discussion is spent discussing the α and delay filters with respect to

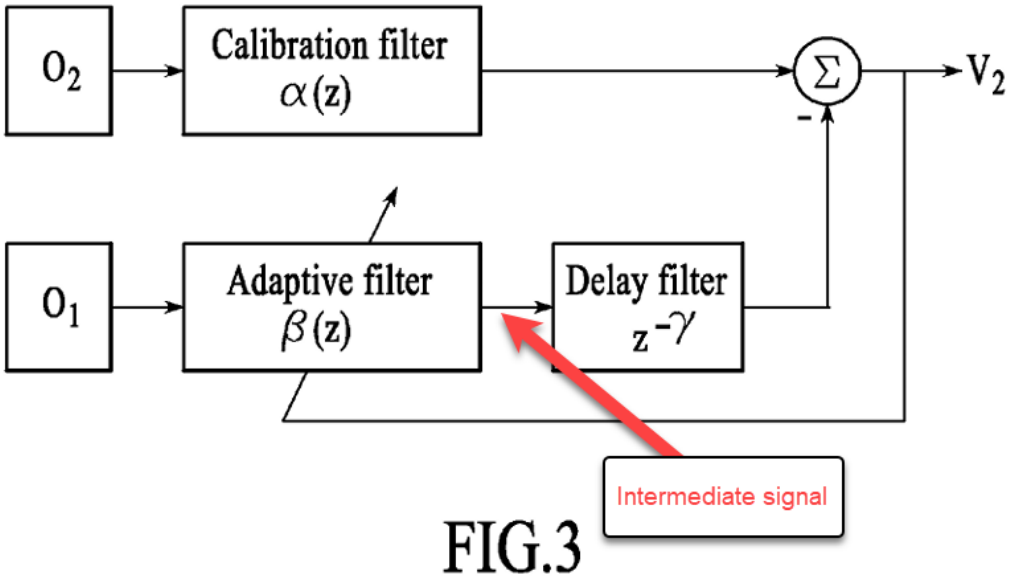
noise removal, which is not claimed in this patent. *Id.*, 23:9-53, 26:62-64, 28:64-30:24.

54. A POSITA would therefore understand that the key component of the invention—the relationship filter—was not the delay or calibration filters. Instead, it is the β filter that the specification discusses in detail.

55. I understand that Meta argued that adopting Jawbone's construction would exclude embodiments (specifically the α and delay filters) that were described in the specification. I disagree.

56. Each embodiment of the invention (that is, the claimed device) includes a β filter. The α and delay filters are always ***in addition*** to the β filter. The specification nowhere describes the relationship filter as being either of the α and delay filters. Moreover, nothing in Jawbone's construction would preclude including those filters in addition to the relationship filter.

57. Further, dependent claims add the α and delay filters as additional elements. For example, Claims 4 and 17 apply each apply a delay to the intermediate signal. Similarly, dependent claims 3 and 15 add limitations that cover a calibration applied to the second signal. These are the exact embodiments disclosed in the specification and shown in Figure 3.



'213 Patent, Fig. 3 (annotations added).

58. In my opinion, therefore, a POSITA, reading the claims in light of the specification, would therefore understand that the "filter that describes a relationship for speech between the first physical microphone and the second physical microphone" refers to "a filter that alters a signal based on the non-temporal components of relative responses to speech of the first and second physical microphones."

VIII. THE ALLEGED PRIOR ART

A. Avendano '880 (U.S. Patent No. 8,194,880)

59. I have reviewed U.S. Patent No. 8,194,880 ("Avendano") (Ex. 1005). Avendano was filed on January 29, 2007 and issued on June 5, 2012.

60. Avendano describes "systems and methods for utilizing inter-microphone level differences (ILD) to attenuate noise and enhance speech". Avendano, Abstract. In exemplary embodiments, "primary and secondary acoustic signals are received by omni-directional microphones and converted into primary and secondary electric signals". *Id.* Further, "a differential microphone array module processes the electric signals to determine a cardioid primary signal and a cardioid secondary signal." *Id.* The cardioid signals are "filtered through a frequency analysis module" and "Energy levels of the signals are then computed, and the results are processed by an ILD module using a non-linear combination to obtain the ILD." In exemplary embodiments, "the non-linear combination comprises dividing the energy level associated with the primary microphone by the energy level associated with the secondary microphone. The ILD is utilized by a noise reduction system to enhance the speech of the primary acoustic signal." *Id.*

61. Fig. 4a in Avendano represents an illustrative embodiment utilized in the Petition and the Declaration of Dr. Kenny.

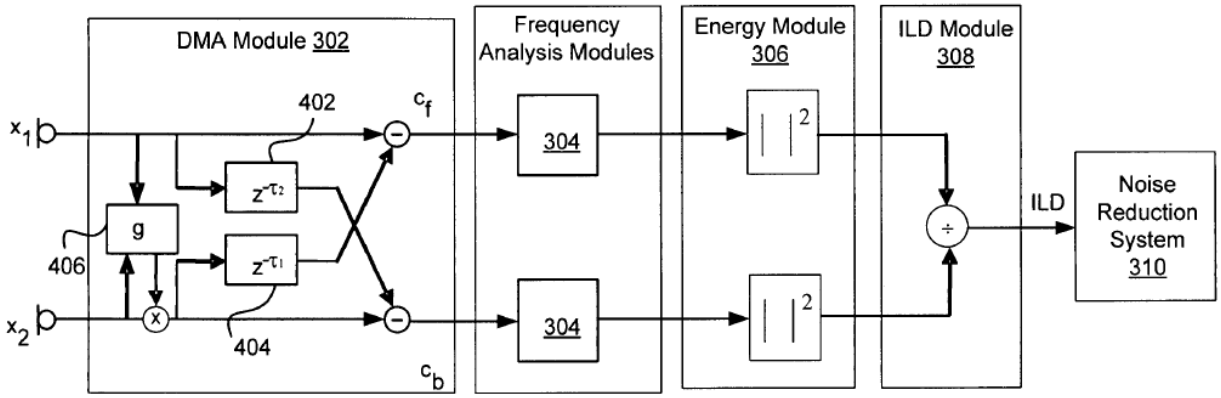


FIG. 4a

Id., Fig. 4a; Pet., 10; Ex. 1003, ¶ 60.

62. Avendano uses two "delay nodes" 402 and 404 to delay the primary and secondary acoustic signals. Avendano, 5:15-24. The gain module 406 computes a gain factor to equalize the signal levels. *Id.*, 5:36-39. The two subtraction units then create cardioid primary (C_f) and cardioid secondary (C_b) signals. *Id.*, 5:25-35.

63. The C_f signal is the cardioid signal in the forward direction. Ex. 2004, 33:24-34:6. The C_b is the cardioid backward signal. *Id.*, 34:7-11. As a result of this processing, the C_f and C_b signals are strongly directional, with the C_f having a strong response in the forward direction and a null in the backward direction, and the C_b signal having a strong response in the backward direction and a null in the forward direction. *Id.*, 35:11-36:11. Avendano, Fig. 6.

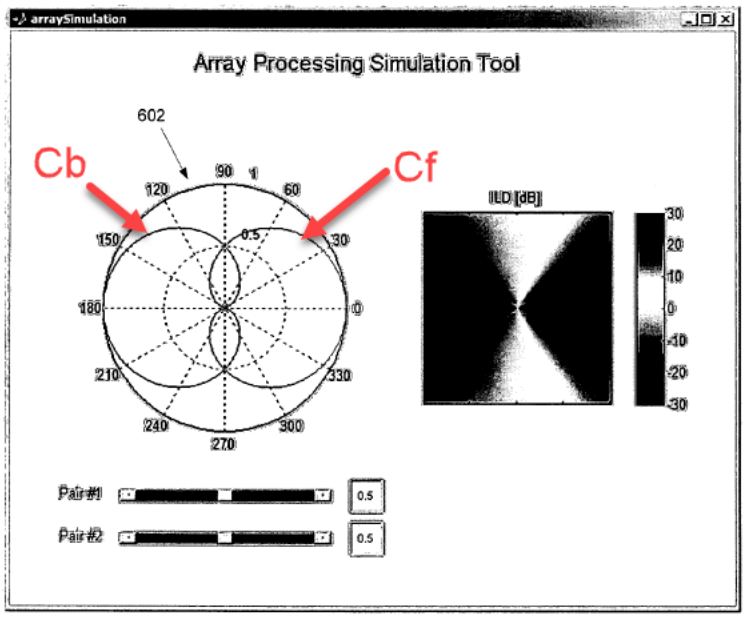


FIG. 6

Avendano, Fig. 6 (annotations added).

64. Following the DMA Module, the signals are processed through a frequency analysis module 304. *Id.*, 5:40-48. The frequency analysis module may filter the signals. *Id.* The system then calculates the power estimates associated with each signal, and then uses those estimates to generate the inter-microphone level differences. *Id.*, 6:7-20.

B. Visser (U.S. Patent No. 7,464,029)

65. I have reviewed U.S. Patent No. 7,464,029 ("Visser") (Ex. 1006). Visser was filed on July 22, 2005 and issued on December 9, 2008.

66. Visser discusses "a signal separation process [that] is associated with a voice activity detector." Visser, Abstract. Notably, "[t]he voice activity detector is a

two-channel detector, which enables a particularly robust and accurate detection of voice activity." *Id.* Visser uses Independent Component Analysis (ICA) to separate speech signals from noise sources. Visser, 3:3-20. Visser's method is also described as blind source separation. *Id.* Fig. 5 from Visser shown below is an illustration of the application of the ICA or blind source separation.

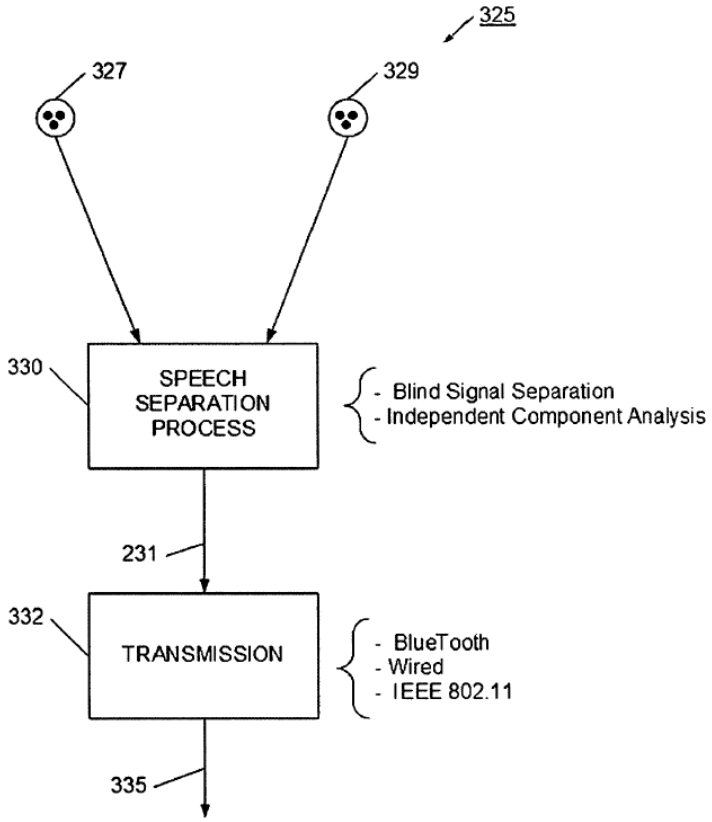


FIG. 5

67. While Petitioner mentions Visser's voice activity detector (VAD), it does not rely on the VAD or argue that it would be part of its combination. Pet., 16.

68. Visser explains (in ref to Fig. 1) that its VAD "receives two input signals 105, with one of the signals defined to hold a stronger speech signal." Visser,

8:31-35. When the VAD detects speech, it "generates a control signal 107." Visser, 8:36-38. As Visser explains, the VAD "enables a more stable and effective speech separation process." *Id.*, 8:45-47.

69. The Petition states that Visser's VAD takes as input "processed signals" that include "(i) a 'speech channel' that 'contains predominantly desired signals,' and (ii) a 'noise channel' that 'contains predominantly noise signals' (e.g., generated by 'ICA or BSS processing')." Pet., 16. Dr. Kenny states that "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.' " Ex. 1003, ¶ 74.

70. Both the Petition and Dr. Kenny leave out a critical portion of Visser. Visser explains that the input signals to the VAD are received directly from the microphones. Visser, Figs. 2, 3; 11:27-12:1. The VAD then makes a comparison based on those signals as to whether speech is occurring. *Id.*, Fig. 3.

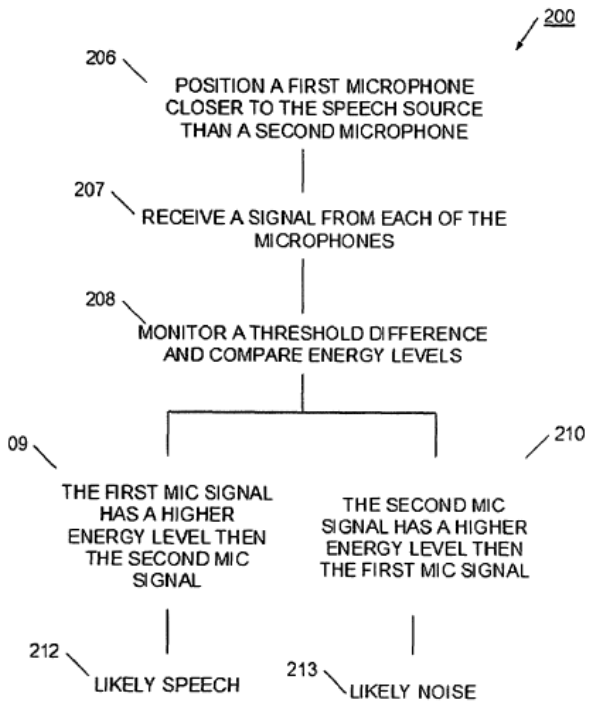


FIG. 3

Visser, Fig. 3.

71. After processing, the signal is then compared again, to form a two channel VAD that generates a control signal. Visser, 8:48-9:5, Fig. 4.

72. The control signal is used to control multiple aspects of the remainder of the Visser system. Visser, 9:6-10:51. The control signal controls whether the speech separation process is active (Visser, 8:38-43), when the filters may adapt (Visser, 8:40-43, 9:7-21, 9:29-32), and other aspects of Visser's post processing.

73. In the relevant embodiment, Visser uses crossfilters w_{21} and w_{12} to separate noise and speech. The filters are adaptive filters. Visser, 17:29-18:29.

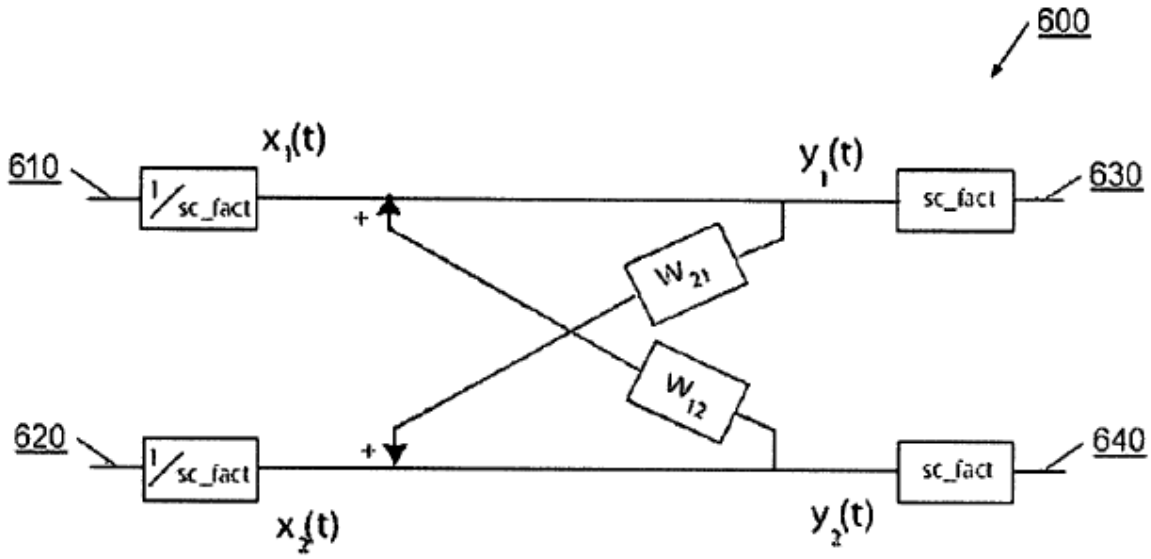


FIG. 10

74. In Figure 10, input signals X_1 and X_2 are received from the input channels 610 and 620. Visser, 17:29-50. The filters w_{21} and w_{12} are applied to each signal to produce the channels of separated signals. *Id.*

C. Avendano/Visser Combination

75. Petitioner asserts that a POSITA would "incorporate Visser's 'ICA process' into Avendano's system to further enhance the differences in speech content in Avendano's two processed signals, such that speech is detected with a greater degree of accuracy, 'even in a changing acoustic environment.'" Pet., 17.

76. Petitioner then specifies that the only portions of Visser to be incorporated into Avendano are "Visser's 'cross filters w21 and w12.'" Pet. 19-20. Petitioner illustrates its combination with a modified Fig. 4a from Avendano:

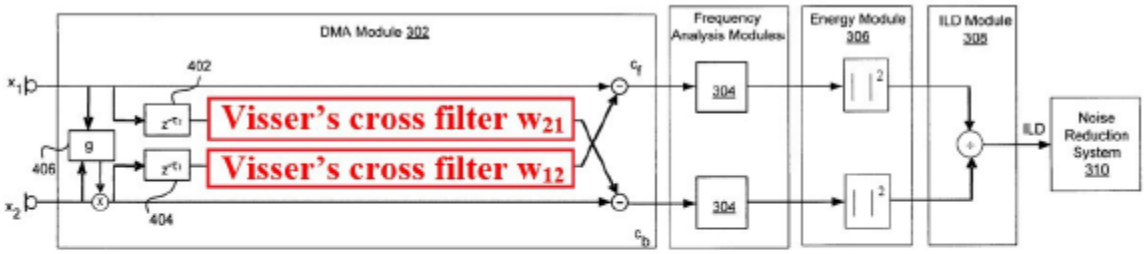


FIG. 4a

Ex. 1005, FIG. 4a (modified)

Pet., 20.

77. Dr. Kenny confirmed that, other than the W_{21} and W_{12} filters and possible connections between them, no other portion of Visser would be incorporated into the combination. Ex. 2004, 61:14-63:6.

78. In Petitioner's combination, microphones are arranged as they are in Avendano. Ex. 2004, 64:15-24. Thus, the signals are received from two microphones: one that is nearer to a speaker and one that is further away. Avendano, 3:27-35; Ex. 2004, 37:8-19.

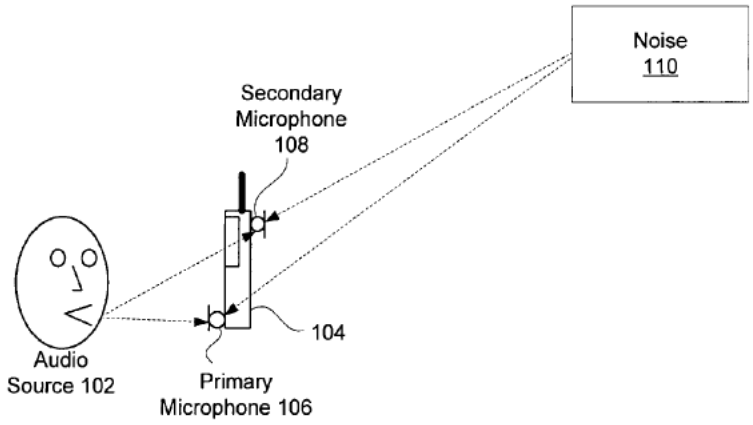


FIG. 1a

IX. PETITIONER'S COMBINATIONS DO NOT RENDER ANY CLAIM OBVIOUS

A. GROUND 1: The Combination of Avendano and Visser Does Not Render Obvious Claims 14-17, 35-42

79. In my opinion, the combination of Avendano and Visser does not render claims 14-17, 35-42 of the '213 Patent obvious.

1. Petitioner's Combination Does Not Disclose or Render Obvious "the processing component forming a filter that describes a relationship for speech between the first physical microphone and the second physical microphone" as Required by All Challenged Claims

80. All Challenged Claims require "forming a filter that describes a relationship for speech between the first physical microphone and the second physical microphone." I understand that Petitioner argues that this limitation is

disclosed or rendered obvious by Avendano alone, or, alternatively, the combination of Avendano and Visser. I disagree.

i. Avendano's Delay Node Does Not Render Obvious The Claimed Filter

81. The Petition first argues that Avendano discloses this limitation via its "delay node 402." Pet. at 25-27. As I explain above, under the proper construction, the claim requires a filter that alters a signal based on the non-temporal components of relative responses to speech of the first and second physical microphones. Avendano's delay node, even if implemented as an all-pass filter, does not meet this requirement.

82. The only relationship for speech that Petitioner identifies is "a temporal relationship for speech between the first and second signals." Pet., 26. Petitioner describes this relationship in strikingly similar terms to the '213 Patent's delay filter:

> In Avendano, "due to a 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 "difference in times of arrival of the signal from a speech source to the microphones."

Pet., 26.

83. Simple mathematics shows that the temporal relationship between the microphones is defined as the distance between them divided by the speed of sound in air. That is exactly the relationship handled by the separate delay filter in the '213 Patent. '213 Patent, 5:35-51.

$$\gamma = \frac{d}{c}$$

84. As I explain above, the specification specifically distinguishes the inventive speech detection process from those based on delays.

85. Unlike the claimed β(z) filter, the delay node does not change the the signal passing through it based on non-temporal components.[1] It simply inserts a delay to account for the distance between the microphones.

86. Accordingly, in my opinion, Avendano does not disclose or render obvious this limitation.

ii. A POSITA Would Not Have Been Motivated to Make the Avendano/Visser Combination

87. In addition to its theory regarding Avendano alone, Petitioner also argues that the Avendano/Visser Combination would render this limitation obvious. I disagree.

88. In my opinion, a POSITA would not make the combination because it would not provide a benefit or function properly.

[1] The Petition does not argue that Avendano's gain is the relationship filter. I agree that the gain is not the relationship filter because it is simply a gain applied to match signal strengths, and does not describe a relationship for speech.

89. As discussed above, Visser uses its VAD to determine when to enter a learning stage where filters may adapt. Visser, 9:6-32 ("In this way, the speech separation process only applies its adaptive learning processes when desired speech is likely occurring."), 9:29-33 ("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.").

90. As Visser explains, the adaptation is controlled based on the input signals received from the microphones. Visser, 9:25-29 ("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.").

91. Petitioner's combination does not include Visser's VAD. Pet., 17-21. Ex. 2004, 61:14-63:6. Instead, Visser's cross filters (including possible connections between them) are the only components of Visser included in the combination. *Id.* Dr. Kenny specifically confirmed that the combination does not include Visser's VAD. *Id.*

92. However, this leads to the case where the adaptive W_{21} and W_{12} filters will continuously and inappropriately adapt, regardless of whether speech is occurring. Visser, 9:29-33.

93. This inappropriate adaptation will result in, for example, the W_{21} filter adapting to remove noise, rather than speech, from the first signal. The resulting filtered signal, when subtracted from the second signal, will result in the C_b cardioid signal including more speech content (from the X_1 microphone) and less noise content than it otherwise would.

94. Because the C_b signal includes more content from the X_1 microphone, it will no longer be as strongly directional as is shown in Avendano Fig. 6. This will lead to signal distortion, while will make the ILD comparison less reliable. Ultimately, introducing the Visser filters results in a ***less accurate*** VAD than Avendano alone.

95. A POSITA would understand the above from reading Avendano and Visser. A POSITA would not be motivated to introduce additional circuitry when the result is a less effective system.

96. Accordingly, in my opinion, a POSITA would not make Petitioner's proposed combination.

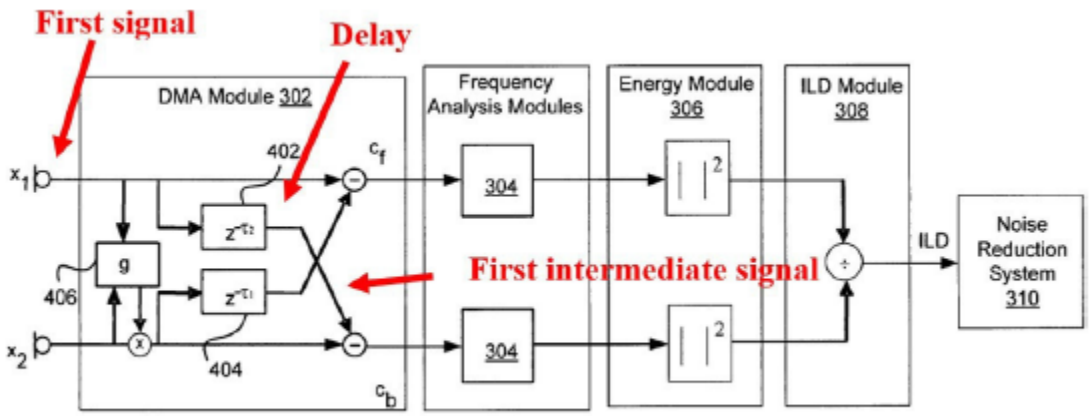
2. The Avendano/Visser combination does not render claim 17 obvious because they do not disclose or render obvious "applying a delay to the first intermediate signal"

97. Claim 17 depends from claim 15 and adds the limitation "the device of claim 15, comprising applying a delay to the first intermediate signal."

98. Petitioner argues that this limitation is disclosed by Avendano's delay node 402 or the Visser cross filters. Pet., 36-37. I disagree.

99. As recited in claim 14, the first intermediate signal is generated by "applying the filter to the first signal to generate a first intermediate signal." The filter is the relationship filter discussed above. The plain language of the claim states that the intermediate signal exists only ***after applying the filter***.

100. Avendano does not disclose this limitation because, under Petitioner's theory, the delay node ***is*** the "filter." Thus, the intermediate signal is the signal output from the delay filter. Petitioner's own annotated figure demonstrates the contradiction in Petitioner's theory.



Pet., 36.

101. As Petitioner admits, "Avendano's first intermediate signal is an output of 'delay node 402,' which is generated by applying 'delay node 402' (*i.e.,* a filter)

to the first signal." Pet., 36. In other words, the delay is applied to the "first signal" not the first intermediate signal.

102. Avendano therefore does not disclose this limitation. Petitioner does not suggest that it would be obvious to introduce a further delay after delay node 402. Pet., 35-36.

103. Petitioner's Avendano/Visser combination has the same flaw, as shown in Petitioner's own figure:

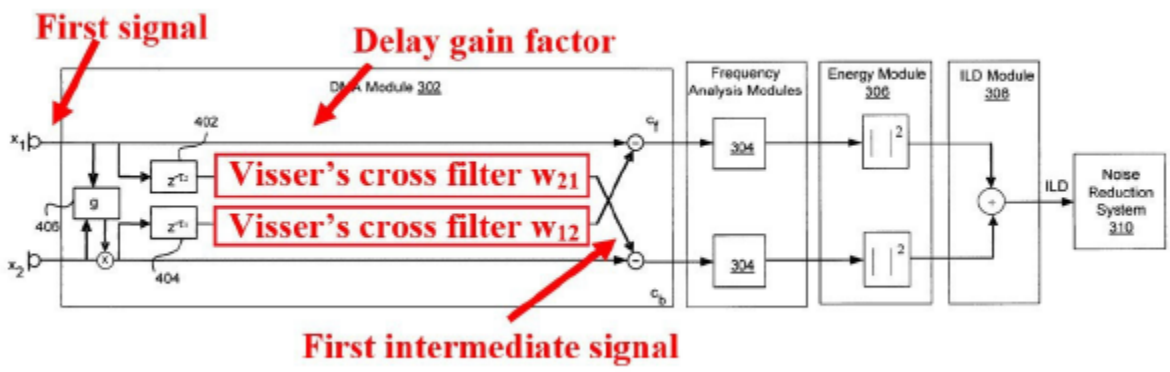


FIG. 4a

Pet., 37.

104. Just as with Avendano alone, Petitioner has identified the first intermediate signal as the output from the Visser cross filter. But that signal does not exist until after the cross filter. Pet., 37. *See also* Pet., 32 ("According to the modification, the ***first intermediate signal is an output of 'cross filter w_{21},'*** which is generated by applying 'cross filter w_{21}' to a first signal x_1." (emphasis added)).

105. Avendano thus does not disclose this limitation. Petitioner does not suggest that it would be obvious to introduce a further delay after the cross filter.

106. In my opinion, therefore, the combination of Avendano and Visser does not disclose clam 17's limitation and does not render claim 17 obvious.

B. GROUND 2: Avendano, Visser, and Bisgaard Do Not Render Obvious Claims 18-26, 33, and 34

107. Claims 18-26 and 34 are dependent on claim 17 and its parent claim 14. The Petition relies on its Ground 1 arguments for the limitations of claims 14 and 17.

108. The Petition does not argue that the addition of Bisgaard remedies its Ground 1 deficiencies as to claim 14.

1. The Avendano/Visser/Bisgaard Combination does not remedy Ground 1's deficiencies as to Claim 17

109. The Petition also does not argue that the addition of Bisgaard remedies its Ground 1 deficiencies as to claim 17. In particular, Petitioner does not use Bisgaard to add a delay to the intermediate signal, but only to set the amount of delay introduced by either Avendano's delay node 402 or Visser's cross filter ***before the intermediate signal***. Pet., 44:

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 (*e.g.*, to account for the delay in the arrival of sound between the two microphones). Ex. 1010, [0042]; Ex. 1011, 5:22-24; Ex. 1003, ¶234.

Pet., 44. I understand that Dr. Kenny confirmed this at his deposition. Ex. 2004, 75:14-78:2.

110. Therefore, introducing Bisgaard does not remedy the lack of a delay applied to the intermediate signal.

2. The Avendano/Visser/Bisgaard Combination does not disclose or render obvious claim 19.

111. Claim 19 depends from claim 18, which ultimately depends from claim 14.

112. Claim 19 adds the limitation: "The device of claim 18, wherein the forming of the first virtual microphone comprises applying the filter to the second signal."

113. Petitioner argues that its combination meets this limitation through Visser's cross filter W_{12} as incorporated into Avendano. Pet., 47. I disagree.

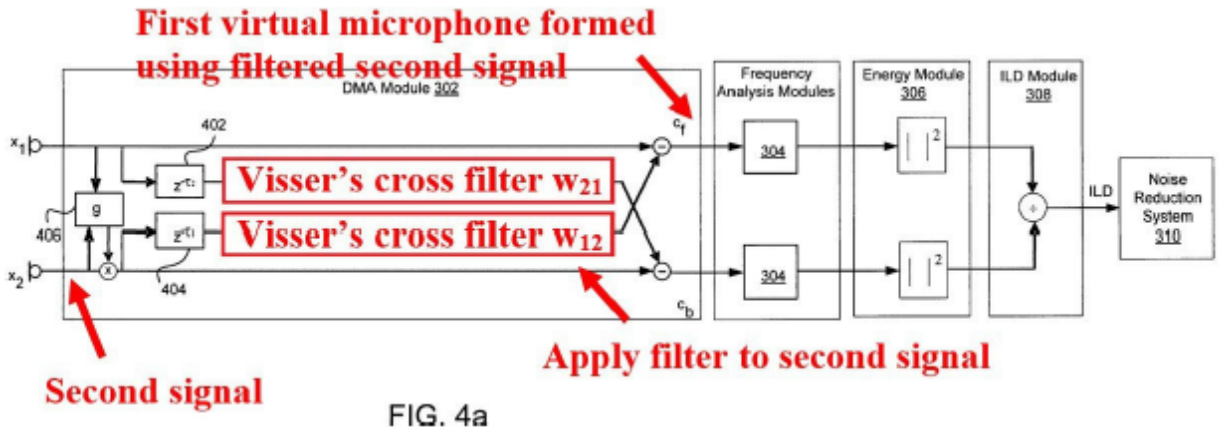


FIG. 4a

Pet., 47.

114. The claim requires "applying ***the filter*** to the second signal." As I discuss above and as recited in claim 14, "the filter" is the "filter that describes a relationship for speech between the first physical microphone and the second physical microphone."

115. As I further discuss with respect to claim 17, the filter is applied to the first signal to generate the first intermediate signal. Claim 14 states this explicitly: "the processing component forming a second virtual microphone by applying the filter to the first signal to generate a first intermediate signal."

116. Claim 19's additional limitation therefore requires that the same filter be applied to both the first and second signals.

117. Under Petitioner's mapping, the W_{21} filter is applied to the first signal. Pet., 32 ("According to the modification, the first intermediate signal is an output of

'cross filter w_{21},' which is generated by ***applying 'cross filter w_{21}' to a first signal x_1.***" (emphasis added)).

118. However, the separate W_{12} filter is applied to the second signal. Pet., 47 ("In the combination, Visser's cross filter w_{12} would be applied to Avendano's second signal x_2.").

119. As a POSITA would recognize, the W_{12} and W_{21} filters are individual filters with separate purposes. Petitioner states that the W_{21} filter acts "to further ***decrease*** speech content in the second virtual microphone Cb." Pet., 20. In contrast, the W_{12} filter acts "to further ***increase*** speech content in the first virtual microphone C_f." *Id.* A POSITA would understand that the two filters (even if interconnected as IIR filters) are separate filters and would have different coefficients. *See* Visser, 17:28-67.

120. Petitioner makes no arguments that it would be obvious to apply the same filter to both signals.

121. Therefore, the Avendano/Visser/Bisgaard combination does not render claim 19 obvious.

122. In sum, no claims are rendered obvious by the Avendano/Visser/Bisgaard combination.

C. GROUND 3: Avendano, Visser, Bisgaard, and Hou Do Not Render Obvious Claims 27-29

123. Claims 27-29 are dependent, through various parent claims, on claims 19, 17, and 14. The Petition relies on its Grounds 1-2 arguments for the limitations of those claims.

124. As shown above, however, Avendano and Visser fail to disclose or render obvious claims 14 and 17, and Avendano/Visser/Bisgaard fail to render obvious claim 19.

125. The Petition does not argue that the addition of Bisgaard or Hou remedies those deficiencies.

126. Therefore, in my opinion, no claims are rendered obvious under Ground 3.

D. GROUND 4: Avendano, Visser, Bisgaard, Hou, and "Frequency Art" Do Not Render Obvious Claims 30-32

127. Claims 30-32 are dependent, through various parent claims, on claims 19, 17, and 14. The Petition relies on its Grounds 1-3 arguments for the limitations of those claims.

128. As shown above, however, Avendano and Visser fail to disclose or render obvious claims 14 and 17, and Avendano/Visser/Bisgaard fail to render obvious claim 19.

129. The Petition does not argue that the addition of Bisgaard, Hou, or the "Frequency Art" remedies those deficiencies.

130. Therefore, in my opinion, no claims are rendered obvious under Ground 4.

X. CONCLUSION

131. In signing this declaration, I recognize that the declaration will be filed as evidence in a contested case before the Patent Trial and Appeal Board of the United States Patent and Trademark Office. I also recognize that I may be subject to cross-examination in the case and that cross-examination will take place within the United States. If cross-examination is required of me, I will appear for cross-examination within the United States during the time allotted for cross-examination.

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

Executed this 2nd day of May 2024

Respectfully submitted,

Akbar M Sayeed

_____