Filed: July 6, 2023

IPR2023-01177
U.S. Patent No. 8,326,611
Petition for *Inter Partes* Review
Attorney Docket SONY 7.1R-023

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

SONY ELECTRONICS INC.,
Petitioner,

v.

JAWBONE INNOVATIONS, LLC,
Patent Owner.

IPR2023-01177
U.S. Patent No. 8,326,611

PETITION FOR *INTER PARTES* REVIEW OF CLAIMS 1-28 OF U.S. PATENT NO. 8,326,611

TABLE OF CONTENTS

TABLE OF AUTHORITIES v

TABLE OF EXHIBITS vii

LISTING OF CHALLENGED CLAIMS viii

I. REQUIREMENTS FOR IPR 1

A. Grounds for Standing 1

B. Challenge and Relief Requested 1

C. Priority Date 1

1. Dynamic Drinkware Analysis 3

a. A hearing instrument, comprising: 3

b. 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; 3

c. 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; and 3

d. 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. 4

II. BACKGROUND 4

A. The ’611 Patent 4

B. Prosecution History 6

C. Level of Ordinary Skill 6

D. Claim Construction 6

III. GROUND 1: AVENDANO AND VISSER (CLAIMS 1-7, 25-28) 7

A. Avendano Overview 7
B. Visser Overview 11
C. Combination of Avendano and Visser 14
D. Claim 1 21
E. Claim 2 33
F. Claims 3 and 4 34
G. Claims 5 and 6 36
H. Claim 7 37
I. Claim 25 38
J. Claim 26 39
K. Claims 27 and 28 39
IV. GROUND 2: Avendano, Visser, And Bisgaard (Claims 8-16, 23, 24) 40
A. Bisgaard Overview 40
B. Combination of Avendano, Visser, and Bisgaard 41
C. Claim 8 45
D. Claim 9 46
E. Claim 10 47
F. Claim 11 48
G. Claim 12 49
H. Claim 13 50
I. Claim 14 50
J. Claim 15 51
K. Claim 16 52
L. Claim 23 52
M. Claim 24 54

V. GROUND 3: Avendano, Visser, Bisgaard, And Hou (Claims 17- 19) 56

A. Hou Overview 56

B. Combination of Avendano, Visser, Bisgaard, and Hou 58

C. Claim 17 63

D. Claims 18 and 19 66

VI. GROUND 4: AVENDANO, VISSER, BISGAARD, HOU, AND FREQUENCY ART (BYRNE, BURNETT, AND/OR BERGLUND) (CLAIMS 20-22) 68

A. Byrne Overview 68

B. Burnett Overview 70

C. Berglund Overview 70

D. Combination of Avendano, Visser, Bisgaard, Hou, and Frequency Art (Byrne, Burnett, and/or Berglund) 71

E. Claim 20 74

F. Claims 21 and 22 77

VII. INSTITUTION IS APPROPRIATE HERE 87

A. Co-Pending Litigation Does Not Warrant Discretionary Denial Under § 314(A) (*Fintiv*) 87

1. Factor 1: Potential Stay 87

2. Factor 2: Proximity of Trial to FWD 89

3. Factor 3: Investment in Parallel Proceeding 89

4. Factor 4: Overlapping Issues 90

5. Factor 5: Parties in Parallel Proceedings 90

6. Factor 6: Other Circumstances 90

B. The Existence of Previously Filed IPR Petitions Does Not Warrant Discretionary Denial (*General Plastic*) 91

VIII. DISCRETIONARY DENIAL UNDER § 325(d) IS ALSO NOT APPROPRIATE 93

IX. MANDATORY NOTICES 93
A. Real Parties-In-Interest (37 C.F.R. § 42.8(b)(1)) 93
B. Related Matters (37 C.F.R. § 42.8(b)(2)) 93
C. Lead and Backup Counsel (37 C.F.R. § 42.8(b)(3)) 94
D. Service Information (37 C.F.R. § 42.8(b)(4)) 95
E. Payment of Fees (37 C.F.R. § 42.103) 95
X. CONCLUSION 96

TABLE OF AUTHORITIES

Page(s)

CASES

Apple Inc. v. Fintiv, Inc.,
IPR2020-00019, Paper 11 (PTAB Mar. 20, 2020) 87, 91

Cal. Inst. of Tech. v. Broadcom Ltd.,
25 F.4th 976 (Fed. Cir. 2022)..... 90

Celltrion, Inc. v. Genentech, Inc.,
IPR2018-01019, Paper 11 (PTAB Oct. 30, 2018) 91

Central Security Group,
IPR2019-01609, Paper 11 91

ClearValue, Inc. v. Pearl River Polymers, Inc.,
668 F.3d 1340 (Fed. Cir. 2012)..... 77

E.I. DuPont de Nemours & Co. v. Synvina C.V.,
904 F.3d 996 (Fed. Cir. 2018)..... 77

GAF Materials LLC v. Kirsch Rsch. & Dev., LLC,
IPR2021-00192, Paper 14 (PTAB May 25, 2021)..... 89

Global Tel-Link Corp. v. HLFIP Holding, Inc.,
IPR2021-00444, Paper 14 (PTAB July 22, 2021) 89

Google LLC v. Jawbone Innovations LLC,
IPR2022-00630, Paper 10 (PTAB Sept. 13, 2022)..... 90

Huawei Techs. Co. v. WSOU Invs., LLC,
IPR2021-00226, Paper 10 (PTAB June 10, 2021)..... 90

Jawbone Innovations, LLC v. Sony Elecs. Inc.,
No. 2:23-cv-01161-CCC-LDW (D.N.J.)..... 87

Kirsch Rsch. & Dev., LLC v. GAF Materials, LLC,
Civil Action No. 20-13683-JMV, 2021 U.S.
Dist. LEXIS 111667 (D.N.J. June 15, 2021) 88

Mercedes-Benz USA, LLC v. Carucel Invs. L.P.,
IPR2019-01404, Paper 12 (PTAB Jan. 22, 2020) 92

NetNut Ltd. v. Bright Data Ltd.,
IPR2021-00465, Paper 11 (PTAB Aug. 12, 2021) 92

Samsung Elecs. Am. Inc. v. Snik LLC,
IPR2020-01428, Paper 10 (PTAB Mar. 9, 2021) 89

Sand Revolution II, LLC v. Cont'l Intermodal Grp.-Trucking LLC,
IPR2019-01393, Paper 24 (PTAB June 16, 2020) 90

Skechers U.S.A., Inc. v. Nike, Inc.,
IPR2021-00160, Paper 10 (PTAB May 19, 2021) 89

Toshiba Am. Info. Sys., Inc. v. Walletex Microelecs, Ltd.,
IPR2018-01538, Paper 11 (PTAB Mar. 5, 2019) 92

Wellman, Inc. v. Eastman Chem. Co.,
642 F.3d 1355 (Fed. Cir. 2011) 6

STATUTES, RULES & OTHER AUTHORITIES

35 U.S.C. § 325(d) 93

TABLE OF EXHIBITS

Exhibit No.	Description
1001	U.S. Patent No. 8,326,611 to Petit et al. ("the '611 patent")
1002	Excerpts from the Prosecution History of the '611 patent ("the Prosecution History")
1003	Declaration of Dr. Thomas Kenny
1004	Curriculum Vitae of Dr. Thomas Kenny
1005	U.S. Patent No. 8,194,880 B2 ("Avendano")
1006	U.S. Patent No. 7,464,029 B2 ("Visser")
1007	RESERVED
1008	U.S. Patent No. 7,155,019 B2 ("Hou")
1009	Byrne, D. et al., "An international comparison of long-term average speech spectra," 1994 Oct; J. Acoust. Soc. Am.; 96(4): 2108-2120 ("Byrne")
1010	U.S. Publication No. US 2011/0103626 A1 ("Bisgaard")
1011	U.S. Provisional Appln. No. 60/816,244 ("the Bisgaard Provisional")
1012	U.S. Publication No. US 2002/0198705 A1 ("Burnett")
1013	Berglund, B., et al., "Sources and effects of low-frequency noise," 1996 May; J. Acoust. Soc. Am; 99(5): 2985-3002 ("Berglund").
1014	Declaration of June Ann Munford
1015	Declaration of June Ann Munford—Appendix
1016	Curriculum Vitae of June Ann Munford

LISTING OF CHALLENGED CLAIMS

Claim 1	
[1pre]	A method comprising:
[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;
[1b]	forming a filter that describes a relationship for speech between the first physical microphone and the second physical microphone;
[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;
[1d]	generating an energy ratio of energies of the first virtual microphone and the second virtual microphone; and
[1e]	detecting acoustic voice activity of a speaker when the energy ratio is greater than a threshold value.
Claim 2	
[2]	The method of claim 1, wherein the first virtual microphone and the second virtual microphone are distinct virtual directional microphones.
Claim 3	
[3]	The method of claim 2, wherein the first virtual microphone and the second virtual microphone have approximately similar responses to noise.
Claim 4	
[4]	The method of claim 3, wherein the first virtual microphone and the second virtual microphone have approximately dissimilar responses to speech.

Claim 5	
[5]	The method of claim 1, comprising applying a calibration to at least one of the first signal and the second signal.
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.
Claim 7	
[7]	The method of claim 5, comprising applying a delay to the first intermediate signal.
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.
Claim 9	
[9]	The method of claim 8, wherein the forming of the first virtual microphone comprises applying the filter to the second signal.
Claim 10	
[10]	The method of claim 9, wherein the forming of the first virtual microphone comprises applying the calibration to the second signal.
Claim 11	
[11]	The method of claim 10, wherein the forming of the first virtual microphone comprises applying the delay to the first signal.

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.
Claim 13	
[13]	The method of claim 12, wherein the filter is an adaptive filter.
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.
Claim 15	
[15]	The method of claim 13, wherein the adapting comprises applying a least-mean squares process.
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.
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.
Claim 18	
[18]	The method of claim 17, wherein the generating of the energy ratio comprises generating the energy ratio for a frequency band.

Claim 19	
[19]	The method of claim 17, wherein the generating of the energy ratio comprises generating the energy ratio for a frequency subband.
Claim 20	
[20]	The method of claim 19, wherein the frequency subband includes frequencies higher than approximately 200 Hertz (Hz).
Claim 21	
[21]	The method of claim 19, wherein the frequency subband includes frequencies in a range from approximately 250 Hz to 1250 Hz.
Claim 22	
[22]	The method of claim 19, wherein the frequency subband includes frequencies in a range from approximately 200 Hz to 3000 Hz.
Claim 23	
[23]	The method of claim 12, wherein the filter is a static filter.
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.
Claim 25	
[25]	The method of claim 1, comprising generating a vector of the energy ratio versus time.
Claim 26	
[26]	The method of claim 1, wherein the first and second physical microphones are omnidirectional microphones.

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.
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.

Petitioner Sony Electronics Inc. ("Sony" or "Petitioner") petitions for *inter partes* review ("IPR") of claims 1-28 ("Challenged Claims") of U.S. Patent No. 8,326,611 ("the '611 patent").

I. REQUIREMENTS FOR IPR

A. Grounds for Standing

Petitioner certifies that the '611 patent is available for IPR. This petition is being filed within one year of service of the original complaint against Petitioner in the district court litigation. Petitioner is not barred or estopped from requesting this review.

B. Challenge and Relief Requested

Petitioner requests IPR on the following grounds.

Ground	Claims	§ 103 Basis
1	1-7, 25-28	Avendano, Visser
2	8-16, 23, 24	Avendano, Visser, Bisgaard
3	17-19	Avendano, Visser, Bisgaard, Hou
4	20-22	Avendano, Visser, Bisgaard, Hou, and Frequency Art (Byrne, Burnett, and/or Berglund)

C. Priority Date

The '611 patent was filed 10/26/2009 as a continuation-in-part of applications filed 05/25/2007 and 06/13/2008, and claims priority to a provisional application filed 10/24/2008.

The Challenged Claims are not entitled to the 05/25/2007 and 06/13/2008 dates because neither CIP application discloses: "generating an energy ratio of energies of the first virtual microphone and the second virtual microphone" and "detecting acoustic voice activity of a speaker when the energy ratio is greater than a threshold value." Thus, the earliest possible priority date is 10/24/2008 ("Critical Date").

Each reference qualifies as prior art:

Reference	Date	Section
Avendano	01/29/2007 (filed)	§ 102(e)
Visser	07/22/2005 (filed)	§ 102(e)
Bisgaard	06/25/2007 (filed) 06/23/2006 (filed, provisional application)	§ 102(e)
Hou	03/14/2001 (filed)	§ 102(e)
Byrne	October 1994 (published)[1]	§ 102(b)
Burnett	12/26/2002 (published)	§ 102(b)
Berglund	May 1996 (published)[2]	§ 102(b)

Bisgaard qualifies as prior art because its filing date (06/25/2007) and the filing date of its provisional application (06/23/2006) predate the Critical Date.

[1] Ex. 1014 ¶¶6-8.
[2] *Id.* ¶¶9-11.

1. Dynamic Drinkware Analysis

Bisgaard claims priority to U.S. 60/816,244 ("Bisgaard Provisional"). Ex. 1011, Cover. The Bisgaard Provisional is incorporated in its entirety in Bisgaard. Ex. 1010 ¶ [0001]. Bisgaard and the Bisgaard Provisional share a similar specification and similar claims. Ex. 1010 ¶¶ [0002]-[0006], [0008]-[0019], [0022]-[0027], [0033]-[0083], [0088]-[0090], claim 1; Ex. 1011, 1:3-3:18, 3:26-12:18, claim 1. Bisgaard is entitled to the 06/23/2006 filing date because the Bisgaard Provisional includes the relevant prior art disclosure and supports at least one of Bisgaard's claims (claim 1), as shown below.

a. A hearing instrument, comprising:

Ex. 1011, claim 1, 1:3-5, 2:1-9, 2:15, 4:1-4, FIG. 1; Ex. 1003 ¶81.

b. 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;

Ex. 1011, claim 1, 1:13-15, 2:1-9, 4:1-9, 5:1-20, FIGS. 1-2; Ex. 1003 ¶81.

c. 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; and

Ex. 1011, claim 1, 2:1-9, 3:3-13, 5:21-6:6, 11:30-12:18, FIGS. 1-2; Ex. 1003 ¶81.

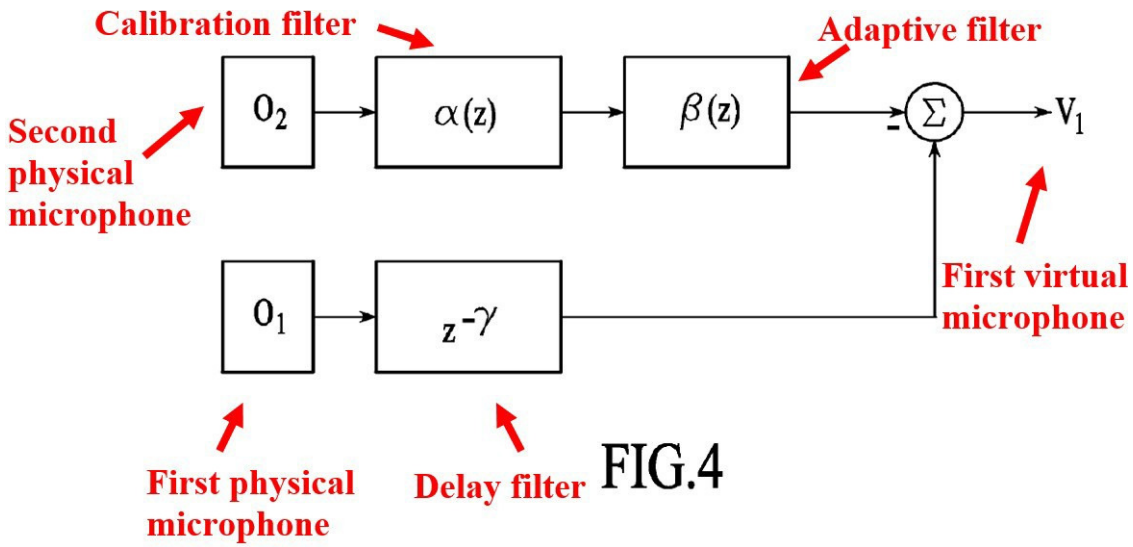
d. wherein the signal processor is further configured to prevent the at least one null direction θ from entering a prohibited range of directions, wherein the prohibited range is a function of a parameter of the electrical sound signals.

Ex. 1011, claim 1, 2:6-9, 3:3-13, 6:18-19; Ex. 1003 ¶81.

II. BACKGROUND

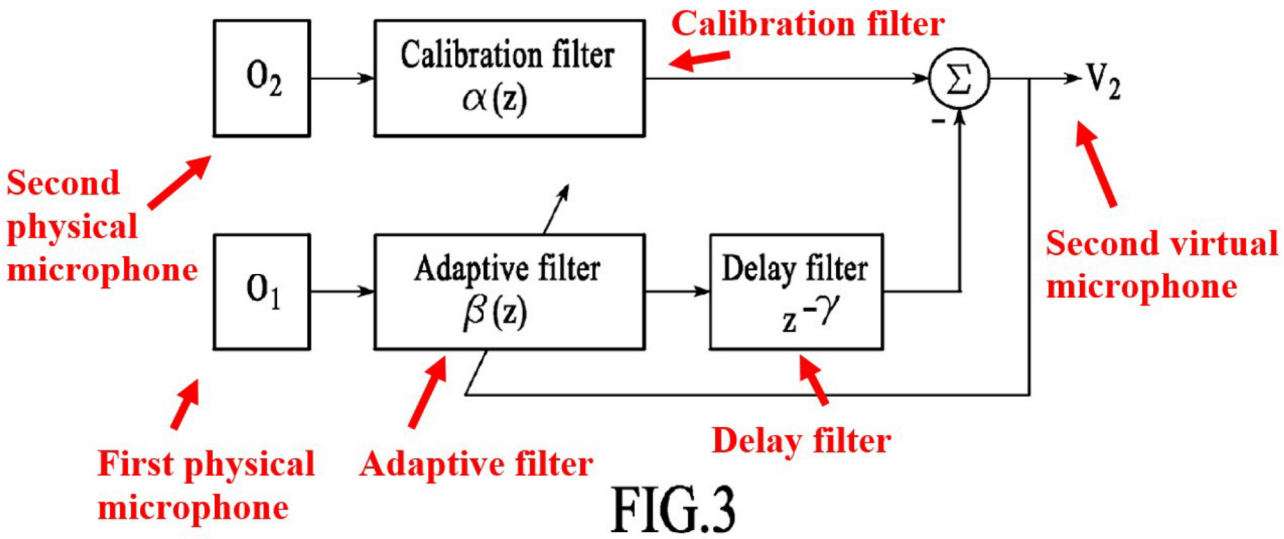
A. The '611 Patent

The '611 patent "relates to noise suppression systems, devices, and methods for use in acoustic applications." Ex. 1001, 1:16-18. A first virtual microphone (V_1) is generated by (i) applying a delay filter ($z^{-\gamma}$) to a signal from a first physical microphone (O_1), (ii) applying a calibration filter ($\alpha(z)$) and an adaptive filter ($\beta(z)$) to a signal from a second physical microphone (O_2), and (iii) combining the filtered signals. *Id.,* 5:20-6:19, FIG. 4. As shown in Figure 4 of the '611 Patent[3]:



[3] Red annotations added throughout.

Further, a second virtual microphone (V_2) is generated by (i) applying an adaptive filter ($\beta(z)$) and a delay filter ($z^{-\gamma}$) to the signal from a first physical microphone (O_1), (ii) applying a calibration filter ($\alpha(z)$) to the signal from a second physical microphone (O_2), and (iii) combining the filtered signals. *Id.* 5:20-6:19, FIG. 3.



Ex. 1001, FIG. 3

The ratio of energies of the first and second virtual microphones is used "to determine when speech is occurring." *Id.* 6:20-10:8, FIGS. 5-11. A ratio that is greater than a threshold value is indicative of acoustic voice activity, whereas a ratio that is less than the threshold value is indicative of an absence of acoustic voice activity. *Id.* 6:47-51, 7:5-7, FIGS. 5-11; Ex. 1003 ¶¶42-49.

B. Prosecution History

The claims were allowed after the filing of a terminal disclaimer over U.S. 12/606,146. Ex. 1002, 195-196, 218-219, 227-233.

C. Level of Ordinary Skill

A person of ordinary skill in the art ("POSITA") would have 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. Ex. 1003 ¶¶22-23. Additional education or industry experience may compensate for a deficit in the other. *Id*.

D. Claim Construction

No formal claim constructions are necessary because "claim terms need only be construed to the extent necessary to resolve the controversy." *Wellman, Inc. v. Eastman Chem. Co.*, 642 F.3d 1355, 1361 (Fed. Cir. 2011).[4]

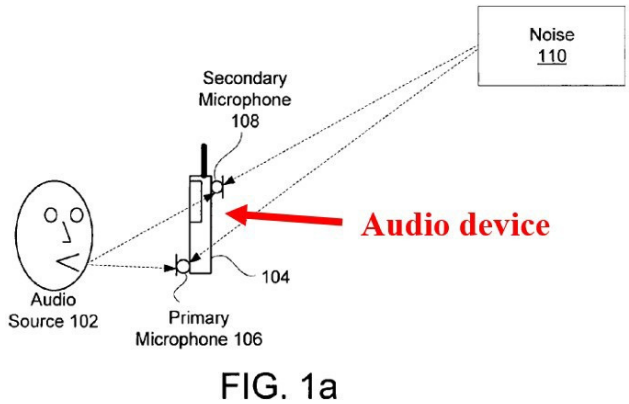
[4] Petitioner is neither conceding that each claim satisfies all statutory requirements, such as §§ 101 and 112, nor waiving any arguments concerning claim scope or grounds that can only be raised in district court.

III. GROUND 1: AVENDANO AND VISSER (CLAIMS 1-7, 25-28)

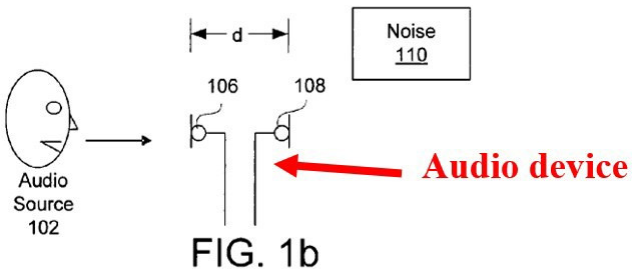
A. Avendano Overview

Avendano determines "inter-microphone level differences (ILD) ... based on energy level differences of a pair of omni-directional microphones," and uses ILD "to attenuate noise and enhance speech." Ex. 1005, 2:5-9.

Avendano discloses "audio device 104" having "primary microphone 106" and "secondary microphone 108," which may be "omni-directional microphone[s]." *Id.* 3:27-35; FIGS. 1a-1b.



Ex. 1005, FIG. 1a



Ex. 1005, FIG. 1b

Avendano's "primary microphone 106 is much closer to [an] audio source 102 than the secondary microphone 108," and thus "the intensity level is higher for the primary microphone 106 resulting in a larger energy level during a speech/voice segment." *Id.* 3:45-55, FIGS. 1a-1b.

Avendano uses this "level difference … to discriminate speech and noise in the time-frequency domain." *Id.* 3:55-57. For example, Avendano receives signals from the two microphones (signals x_1 and x_2), and processes the signals using "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.

Avendano's DMA module 302 generates (i) a first processed signal having a directional pattern for receiving sounds "within a front cardioid region around the audio device 104" (*i.e.,* "cardioid primary signal (C_f)"), and (ii) a second processed signal having a directional pattern for receiving sounds "within a back cardioid region around the audio device 104" (*i.e.,* "cardioid secondary signal (C_b)"). *Id.* 4:47-52, 5:25-35, 9:29-42, FIG. 6 (below).

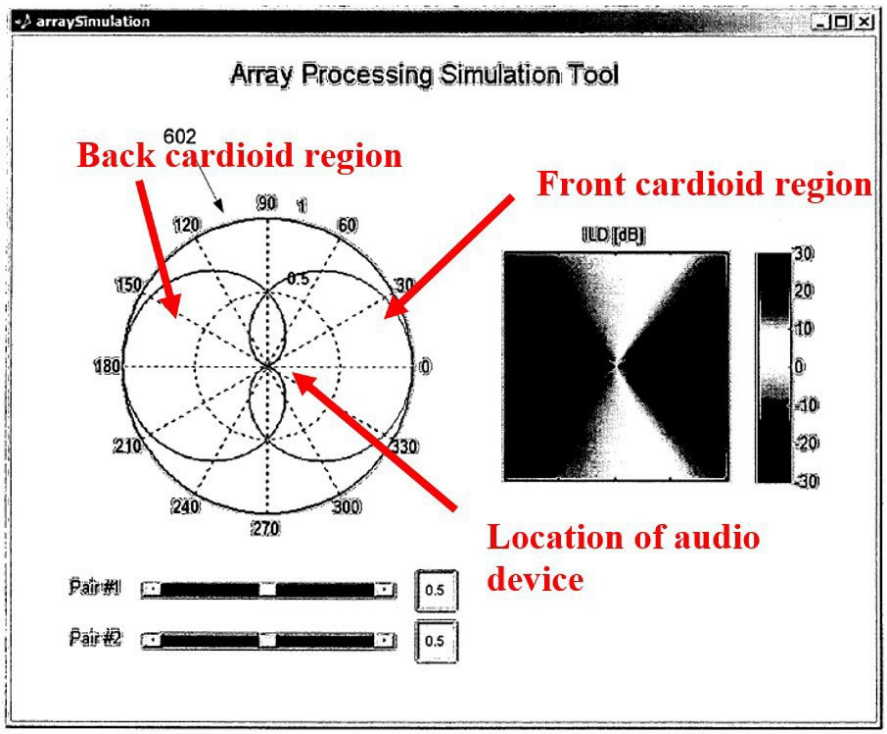


FIG. 6

Ex. 1005, FIG. 6

Avendano's "cardioid primary signal (C_f)" is generated by combining (i) signal x_1 from primary microphone 106, and (ii) signal x_2 from secondary microphone 108 (signal x_2 having been filtered by "delay node 404" and "gain module 406"). *Id.* 5:15-35, FIG. 4a. Avendano's "cardioid secondary signal (C_b)" is generated by combining (i) signal x_2 from secondary microphone 108, and (ii) signal x_1 from primary microphone 106 (signal x_1 having been filtered by "delay node 402"). *Id.* The "delay nodes" are implemented using filters (*e.g.,* "allpass filters"). *Id.* 8:47-51.

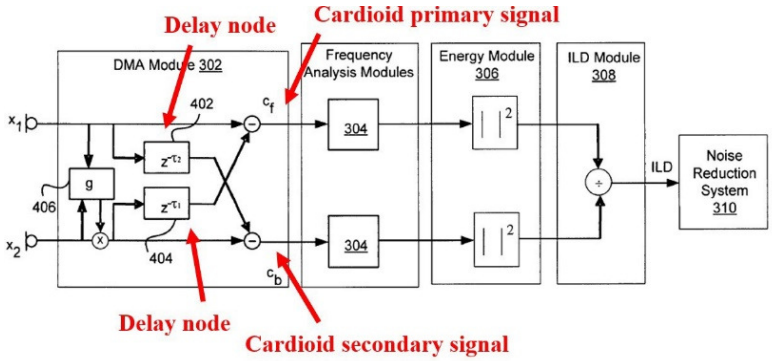


FIG. 4a

Ex. 1005, FIG. 4a

Further, Avendano detects speech based on the ratio between (i) the energy of "cardioid primary signal (C_f)" and (ii) the energy of "cardioid secondary signal (C_b)." *Id.* 5:49-6:34. Specifically, an "energy level" (E_f) associated with "cardioid primary signal (C_f)" is calculated:

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

Ex. 1005, 5:60

Further, an "energy level" (E_b) associated with "cardioid secondary signal(C_b)" is calculated:

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

Ex. 1005, 6:5

The ratio between these two energy levels (ILD) is determined:

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

Energy level of C_f

Energy level of C_b

Ex. 1005, 6:16

Avendano compares the ratio (ILD) to a "threshold" to determine the presence or absence of speech:

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

Ex. 1005, 6:61

If the ratio (ILD) is "smaller than a threshold value (e.g., threshold=0.5) above which speech is expected to be," a value λ_l is set to zero (*e.g.,* indicating an absence of speech). *Id.* 6:58-7:3. However, if the ratio (ILD) "starts to rise (e.g., because speech is present within the large ILD region), λ_l increases" (*e.g.,* is set to one, indicating a presence of speech). *Id.*

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--8:23, 10:18-50, FIGS. 7-8; Ex. 1003 ¶¶51-67.

B. Visser Overview

Visser "improv[es] the quality of a speech signal extracted from noisy acoustic environment" using a "voice activity detector." Ex. 1006, 6:57-60. Visser's "speech separation process 100" separates speech from "sound signals from

microphones ... 102 and 104." *Id.* 8:4-8, FIG. 1. Specifically, "voice activity detector (VAD) 106 ... receives two input signals 105, with one of the signals defined to hold a stronger speech signal," and generates "control signal 107 ... to activate the signal separation process only when speech is occurring." *Id.* 8:33-40.[5]

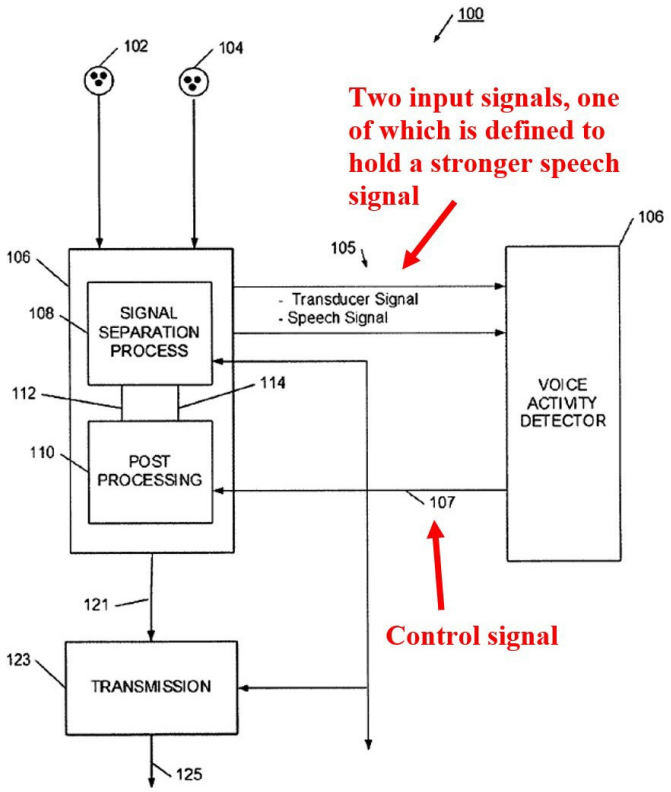


FIG. 1

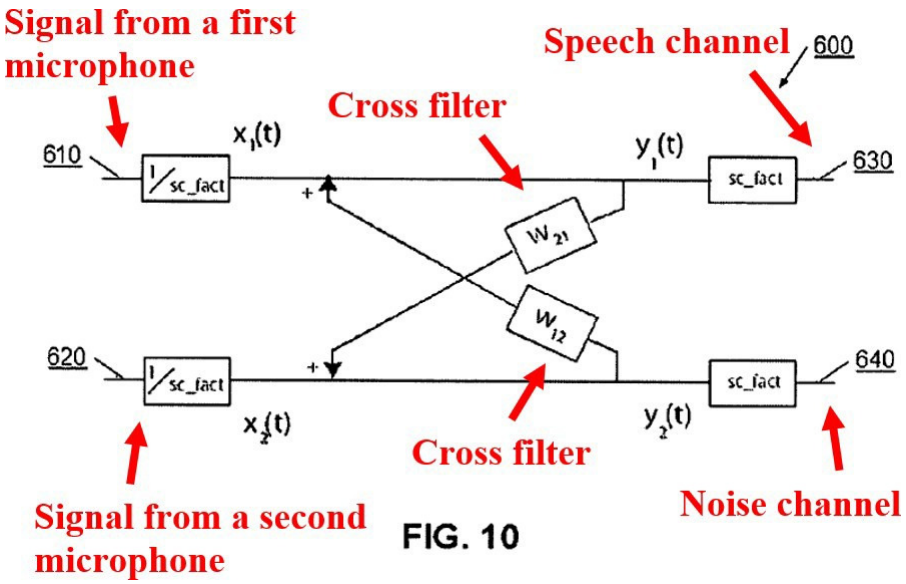
Ex. 1006, FIG. 1

Visser's "signal separation process" 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. Visser's "ICA or BSS processing

[5] All emphasis added.

function" receives "signals X_1 and X_2 ... from channels 610 and 620," each of the signals "typically ... com[ing] from at least one microphone." *Id.* 17:36-39, FIG. 10. Further, Visser's "ICA or BSS processing function" generates (i) "channel 630 of separated signals U_1" (i.*e.,* "speech channel") that "<u>contains predominantly desired signals,</u>" namely speech, and (ii) "channel 540 of separated signals U_2" (*i.e.*, "noise channel") that "<u>contains predominantly noise signals</u>." *Id.* 17:39-44.

Visser's system generates the "speech channel" by combining "input signal X_1" and "input signal X_2" ("input signal X_2" having been filtered by "cross filter w_{12}"). *Id.* FIG. 10. Further, the system generates the "noise channel" by combining "input signal X_2" and "input signal X_1" ("input signal X_1" having been filtered by "cross filter w_{21}"). *Id.*



<u>Ex. 1006, FIG. 10</u>

These channels are provided as inputs to "VAD 106" (*e.g.*, "input signals105") to determine "when speech is present." *Id.* 8:33-36. As Visser describes, "cross filters W_{21} and W_{12} can have sparsely distributed coefficients over time to capture a long period of time delays." *Id.* 17:56-64. 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. In Visser's "ICA process," each filter "ha[s] an adaptable and adjustable filter coefficient." *Id.* 16:63-65. Specifically, "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; Ex. 1003 ¶¶68-77.

C. Combination of Avendano and Visser

A POSITA would have found it obvious to combine Avendano and Visser. Ex. 1003 ¶¶101-125. Both references come from the same field of endeavor of enhancing speech and attenuating noise based on voice activity detection. Ex. 1005 Abstract, 1:24-26, 3:13-26, 3:42-60; Ex. 1006 Abstract, 1:19-23, 6:57-7:25, 8:4-47. Further, 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. Ex. 1005 Abstract, 1:24-26, 3:13-7:9; Ex. 1006 Abstract, 1:19-23, 6:57-7:25, 8:4-47, 17:29-50. Further still, 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." Ex. 1006, 16:65-17:4, FIG. 9; Ex. 1003 ¶104. For example, Avendano determines the presence of speech based on a comparison between (i) a first processed signal ("cardioid primary signal (C_f)" directed to the front) and (ii) a second processed signal ("cardioid secondary signal (C_b)" directed to the back). Section III.A; Ex. 1005, 4:47-7:9. This comparison is particularly suitable for determining presence of speech, as the source of speech ("audio source 102") is positioned on a front side of "audio device 104," and thus "cardioid primary signal (C_f)" would have a greater response to speech than "cardioid secondary signal (C_b)." Ex. 1003 ¶¶105-106.

Specifically, as taught by Avendano, the differences in the "levels" between the two processed signals is represented by a ratio (ILD) between energy levels of these two processed signals. Ex. 1005, 5:49-6:34, 6:58-7:3. If the ratio is sufficiently high (*e.g.,* when the energy level of front-facing "cardioid primary signal C_f" is sufficiently higher than that of back-facing "cardioid secondary signal C_b"), this is indicative of a presence of speech activity. *Id.* 6:58-7:3.

Further, Avendano's "cardioid primary signal (C_f)" and "cardioid secondary signal (C_b)" are each 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. Section III.A; Ex. 1005, 5:15- 35.

Visser determines the presence of speech using principles similar to those taught by Avendano. For example, Visser's "VAD 106" determines the presence of speech based on two processed signals, "with one of the signals defined to hold a stronger speech signal." Section III.B; Ex. 1006, 8:33-35. Specifically, 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.*, generated by "ICA or BSS processing"). Ex. 1006, 17:36-44.

Like Avendano, Visser's processed signals are generated by applying a delay to a signal received from one of the physical microphones (*e.g.,* using a cross filter having "a delay gain factor for the time delay between the output signal and the feedback input signal"), and combining the delayed signal with a signal received from the other physical microphone. *Id.* 17:39-62, FIG. 10.

Visser additionally describes that the separation between the "speech channel" and the "noise channel" is further enhanced by generating and adapting the filters according to an "ICA process." *Id.* 9:20-62, 17:29-25:9, FIGS. 10-13. Such a 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.

Accordingly, a POSITA would have been motivated to 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." Ex. 1003 ¶112.

For example, 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, in order to generate a first processed signal representative of speech (*e.g.,* by combining the filtered signal with another microphone signal). Ex. 1005, 5:15-35, FIG. 4a; Ex. 1006, 17:36-44, FIG. 10; Ex. 1003 ¶¶113-115.

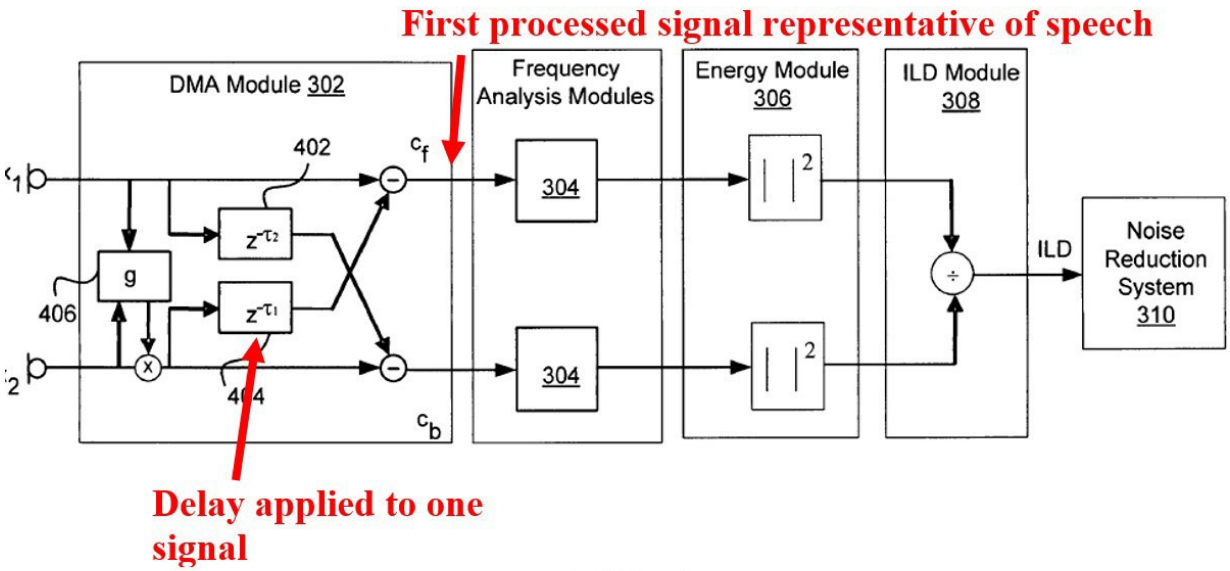


FIG. 4a

Ex. 1005, FIG. 4a

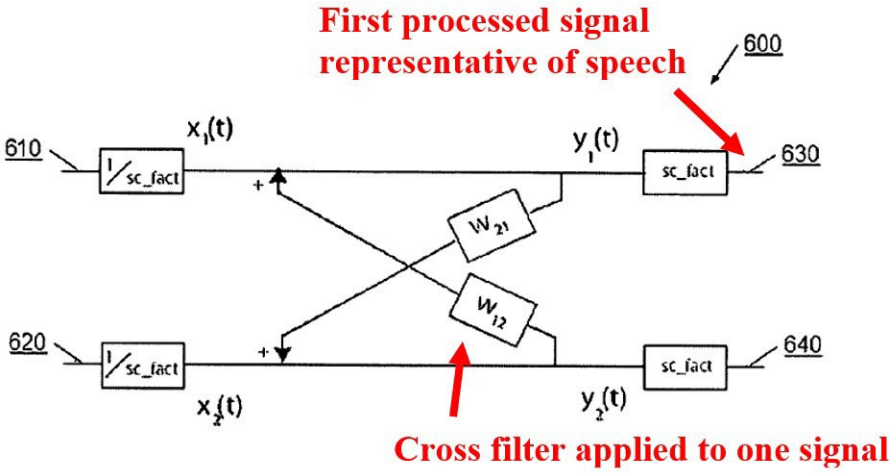


FIG. 10

Ex. 1006, FIG. 10

Similarly, a POSITA would have recognized that, like Avendano's "delay node 402," Visser's "cross filter w_{21}" is configured to filter 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). Ex. 1005, 5:15-35, FIG. 4a; Ex. 1006, 17:36-44, FIG. 10; Ex. 1003 ¶¶116-118.

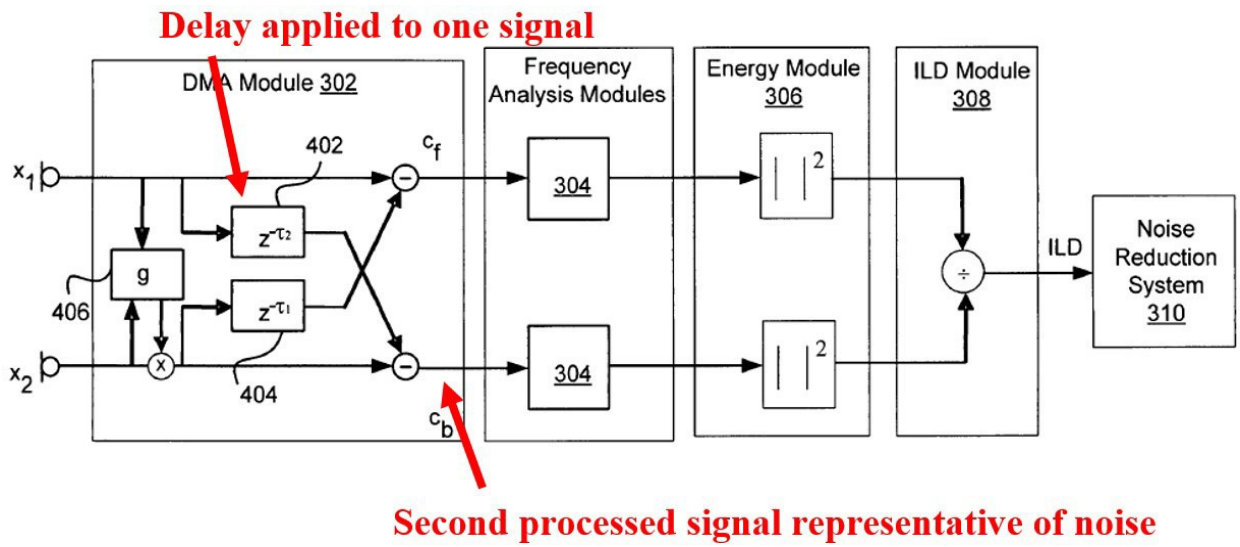


FIG. 4a

Ex. 1005, FIG. 4a

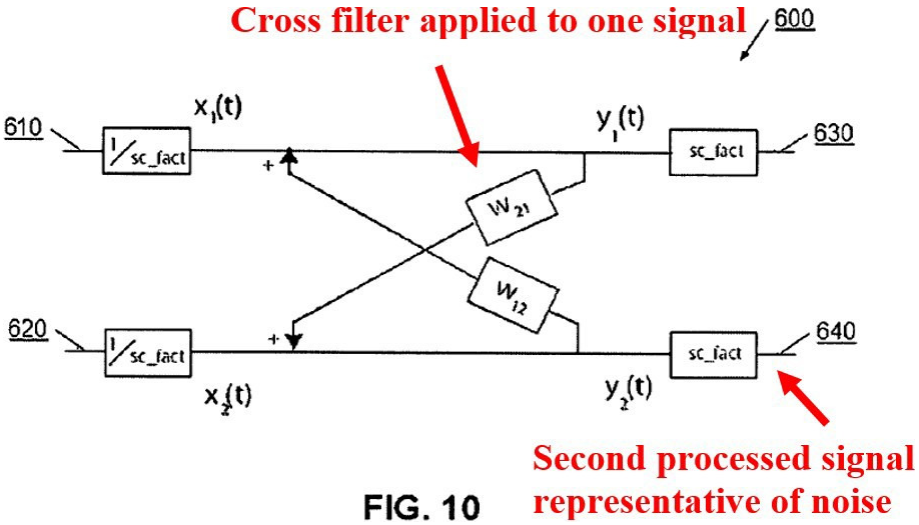


FIG. 10

Ex. 1006, FIG. 10

From this disclosure, and as an example of applying additional filtering to Avendano's signals, a POSITA would have found it obvious to incorporate Visser's "cross filters w_{21} and w_{12}" into Avendano's system. Ex. 1003 ¶119. Such a modification would merely involve using a known technique (*e.g.,* using Visser's "cross filters") to improve a similar device in a similar way (*e.g.,* to further enhance the separation of speech and noise, such that voice activity is detected more accurately). *Id.*

In this example, Visser's cross filter w_{12} would have been applied to Avendano's second signal x_2 (*e.g.,* to further increase speech content in the first virtual microphone C_f). *Id.* ¶120. Further, Visser's cross filter w_{21} would have been applied to Avendano's first signal x_1 (*e.g.,* to further decrease speech content in the second virtual microphone C_b). *Id.* An example of this modification is shown below:

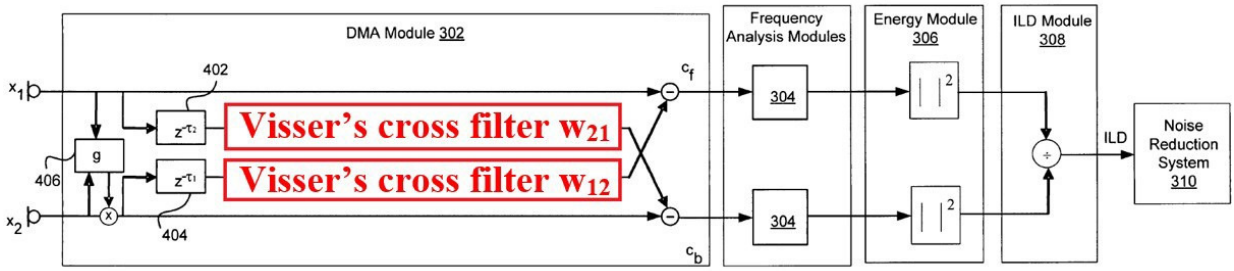


FIG. 4a

Ex. 1005, FIG. 4a (modified)

Further, a POSITA would have had a reasonable expectation of success in incorporating Visser's "cross filters" into Avendano's system. Ex. 1003 ¶¶121-125. For

example, Avendano's "DMA Module 302" and Visser's "cross filters" are both configured to (i) receive similar types of input signals (*e.g.,* signals from two physical microphones), and (ii) output signals for a similar purpose (*e.g.,* to produce two processed signals in which one processed signal has a greater degree of speech than the other). Ex. 1005, 5:15-6:34, FIG. 4a; Ex. 1006, 17:36-44, FIG. 10; Ex. 1003 ¶¶121-123. Given these similarities, limited modifications would have been necessary to incorporate Visser's "cross filters" into Avendano's system and other operations in Avendano's system would not have been impacted by the combination. Ex. 1003 ¶124.

Thus, 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. *Id.* ¶125.

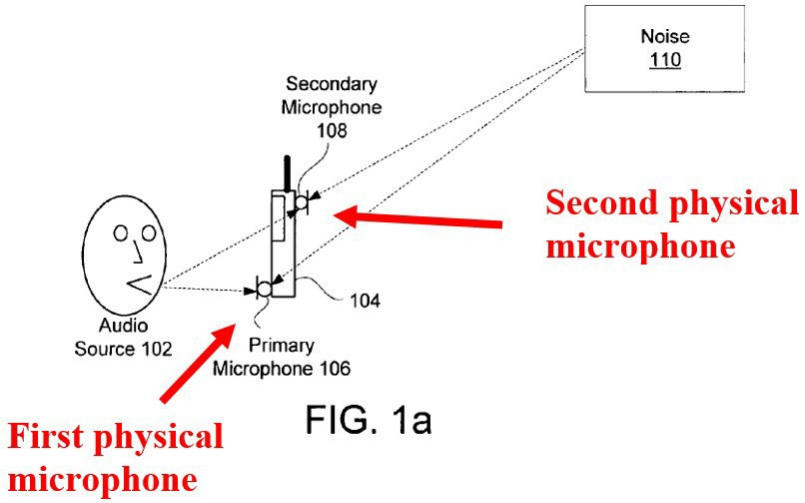
D. Claim 1

[1pre]: A method comprising:

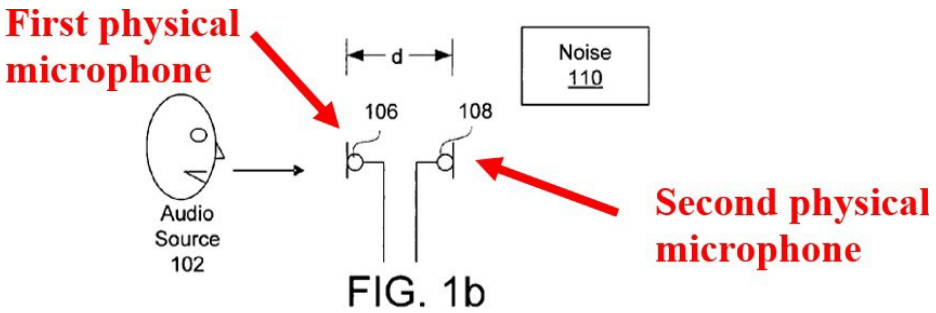
To the extent that the preamble of claim 1 is limiting, Avendano discloses a method. Ex. 1005, FIGS. 1a-7, 3:12-7:9, 9:43-10:25 ("an exemplary method for utilizing ILD of omni-direction microphones for noise suppression and speech enhancement."); Ex. 1003 ¶126.

[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.

In Avendano, the first physical microphone is "primary microphone 106" and the second physical microphone is "secondary microphone 108." Ex. 1005, 3:27-55, FIGS. 1a-1b; Ex. 1003 ¶¶127-131.



Ex. 1005, FIG. 1a



Ex. 1005, FIG. 1b

Further, in Avendano, the first signal is "primary acoustic signal (X_1)" and the second signal is "secondary acoustic signal (X_2)." *Id.* 4:20-27, FIGS. 4a-4b. Further, the

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

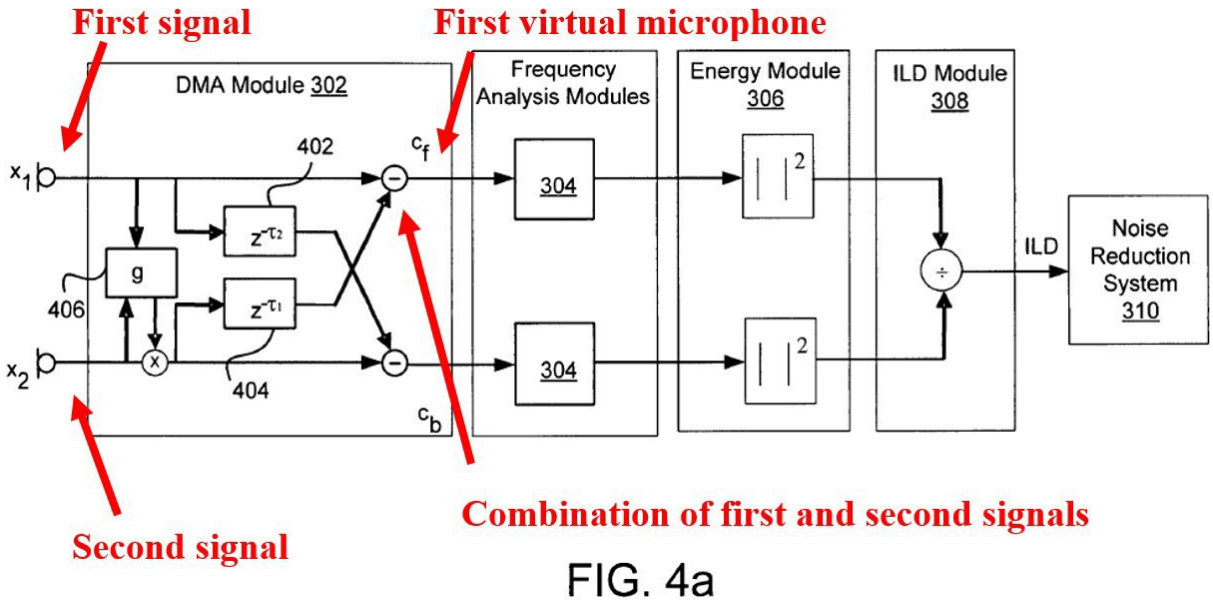


FIG. 4a

Ex. 1005, FIG. 4a

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

Ex. 1005, 5:25-30

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

In Avendano, the filter is "delay node 402," which describes a relationship for speech between the first physical microphone and the second physical microphone. Ex. 1005, 4:28-34, 5:19-35, FIGS. 4a-4b; Ex. 1003 ¶¶133-138.

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." Ex. 1005, 1:33-36; Ex. 1003 ¶135. Specifically, in Avendano, a first physical microphone is positioned closer to a source of speech ("audio source 102"), and a second physical microphone is positioned farther from the source of speech. Section III.D.[1a]; Ex. 1005, 3:45-49, FIGS. 1a-1b. Accordingly, the "relationship for speech" between Avendano's two physical microphones is, at least in part, that one microphone will receive speech prior to the other microphone. Ex. 1003 ¶135. Avendano's "delay node 402" delays the first signal, which would otherwise include speech content before the second signal. Ex. 1005, 5:15-35. Accordingly, Avendano's "delay node 402" describes, at least in part, a temporal relationship for speech between the first and second signals. Ex. 1003 ¶136.

Further, Avendano's delay nodes are implemented using filters. Ex. 1005, 8-38-51 ("[t]o implement a fractional delay, allpass filters 416 and 418 ... are applied to the signals").

This also is consistent with the '611 patent, which describes that a time delay can be implemented using a "delay filter." Ex. 1001, FIG. 3 ("Delay filter $z^{-\gamma}$").

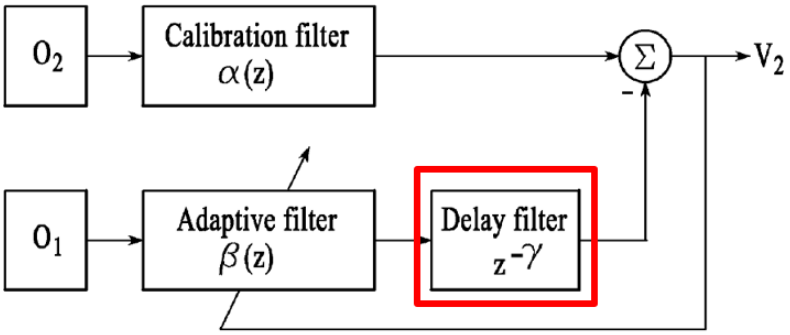


FIG.3

Ex. 1001, FIG. 3

To the extent that Avendano alone does not describe each of the features of [1b], it would have been obvious to modify Avendano's system in view of Visser to include these features. Ex. 1003 ¶¶139-144.

For example, it would have been obvious to incorporate Visser's "ICA or BSS processing function" into Avendano's system. Section III.C. According to this modification, Avendano's system would include Visser's cross filters w_{12} and w_{21}, each of which describes a relationship for speech between first and second physical microphones. Ex. 1003 ¶¶140-141.

According to one example modification, Visser's cross filter w_{12} would have been applied to Avendano's second signal x_2 (*e.g.,* to further increase speech content in the first virtual microphone C_f), and Visser's cross filter w_{21} would have been applied

to Avendano's first signal x_1 (*e.g.,* to further decrease speech content in the second virtual microphone C_b). *Id.*

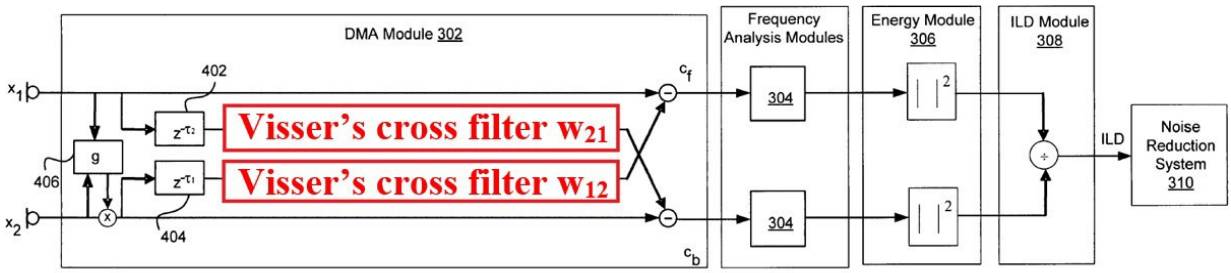


FIG. 4a

Ex. 1005, FIG. 4a (modified)

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. Ex. 1006, 15:43-16:2. Visser's system leverages the different responses of the physical microphones to speech by generating cross filters w_{12} and w_{21} (*e.g.*, using independent component analysis (ICA)) to form "speech channel 630" (which contains predominately speech) and "noise channel 640" (which contains predominately noise). *Id.* 8:16-18, 16:3-28, 17:29-44, FIG. 10. According to an "ICA process," Visser's filters are "adapt[ed] during operation," such that the filters better separate speech from noise. *Id.* 16:8-11, 21:17-18:44, FIG. 12. Specifically, the "speech separation process ... may be adaptive and learn according to the specific acoustic environment," adapting "to particular microphone placement, the acoustic environment, or a particular user's speech." *Id.* 9:8-12.

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. Ex. 1003 ¶144.

[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.

In Avendano, the 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. Ex. 1005, 5:15-35, FIGS. 4a-4b; Ex. 1003 ¶¶145-152.

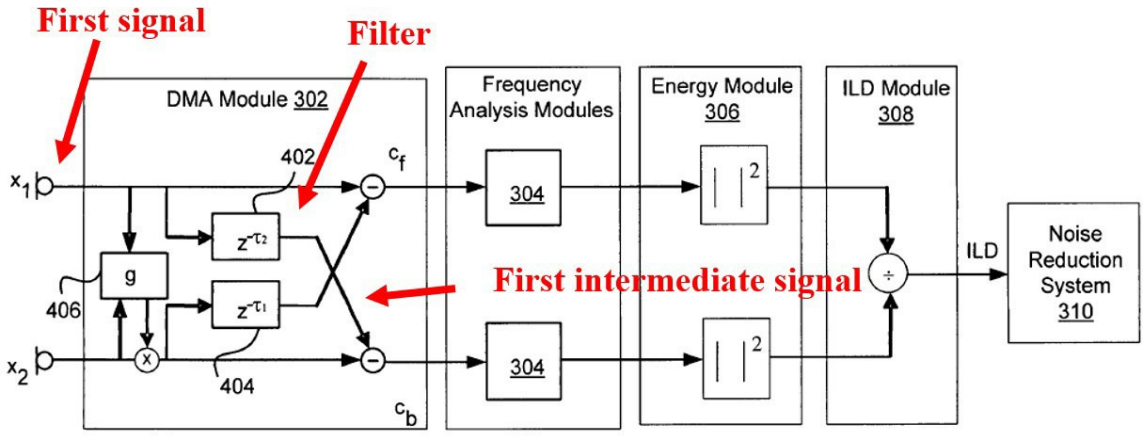


FIG. 4a

Ex. 1005, FIG. 4a

Further, in Avendano, a second virtual microphone is "cardioid secondary signal (C_b)," which is formed by subtracting the first intermediate signal from the

second signal. *Id.*

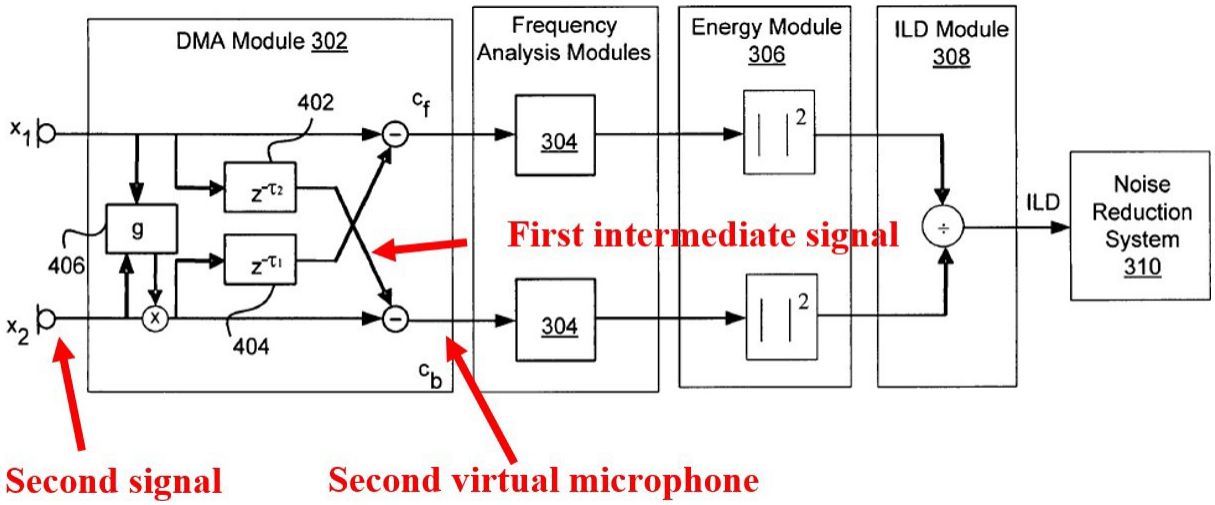


FIG. 4a

Ex. 1005, FIG. 4a

Avendano's FIG. 4b implements this subtraction by summing (i) the second signal x_2, and (ii) the inverse of the first intermediate signal (represented using a summation node and a negative sign by a node input). Ex. 1003 ¶149.

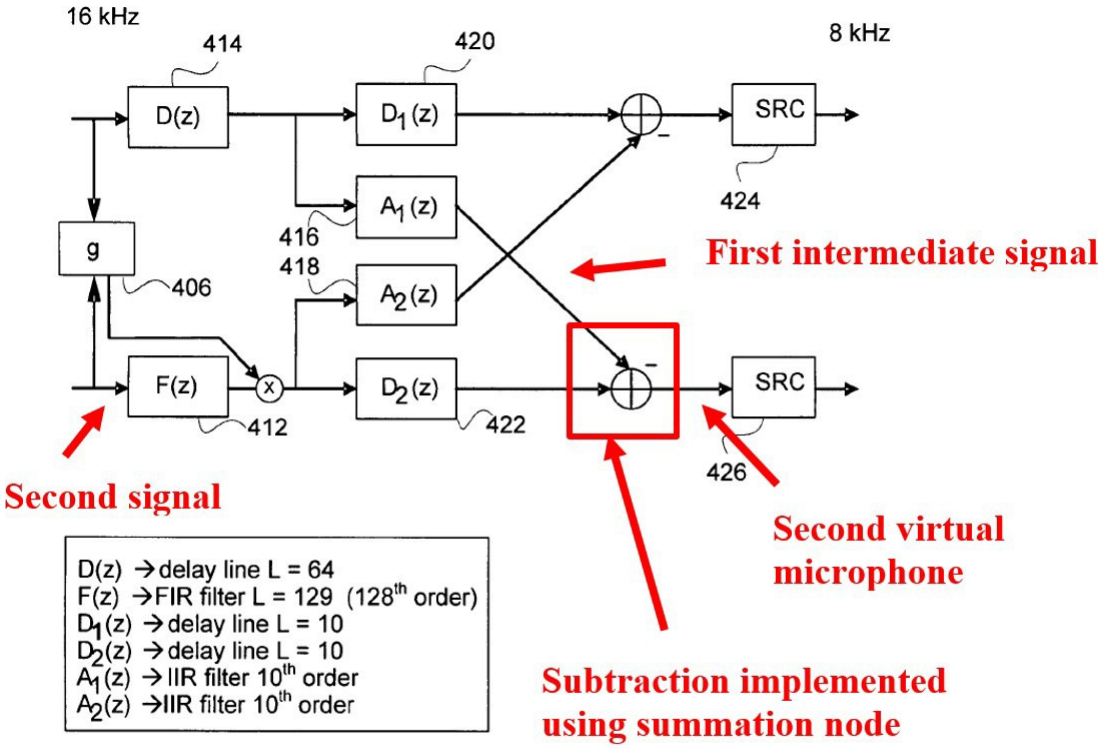


FIG. 4b

Ex. 1005, FIG. 4b

This is consistent with the '611 patent, which describes that "processing paths" are "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." Ex. 1001, 21:26-52, FIGS. 25-26; Ex. 1003 ¶150.

Further, FIG. 3 of the '611 patent indicates that a summation (signified by "Σ") includes summing a first value with the inverse of a second value (signified using a negative sign by an input to "Σ") to form a second virtual microphone V_2 (*i.e.*, subtracting the second value from the first value). Ex. 1003 ¶151.

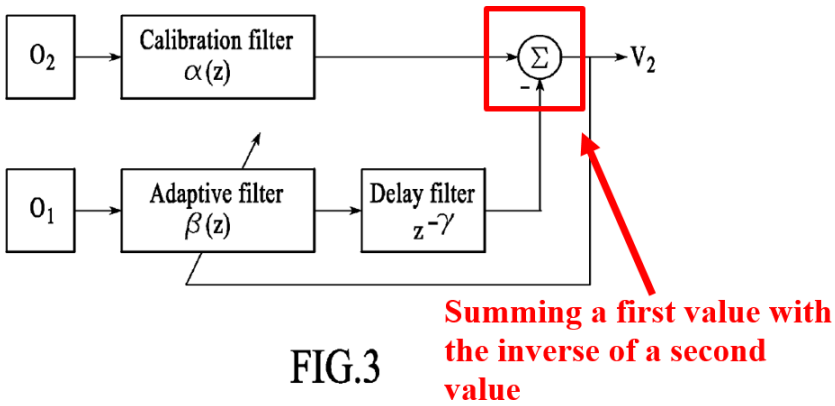


FIG.3

Ex. 1001, FIG. 3

Accordingly, as in the '611 patent, Avendano forms 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. Ex. 1003 ¶152.

To the extent that Avendano alone does not describe each of the features in [1c], it would have been obvious to modify Avendano's system in view of Visser to include these features. *Id.* ¶¶153-162.

As discussed above, 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. Section III.C. According to one example modification, Visser's cross filter w_{12} would have been applied to Avendano's second signal x_2 (*e.g.,* to further increase speech content in the first virtual microphone C_f), and Visser's cross filter w_{21} would have been applied to Avendano's first signal x_1 (*e.g.,* to further decrease speech content in the second virtual microphone C_b). *Id.*

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. *Id.*

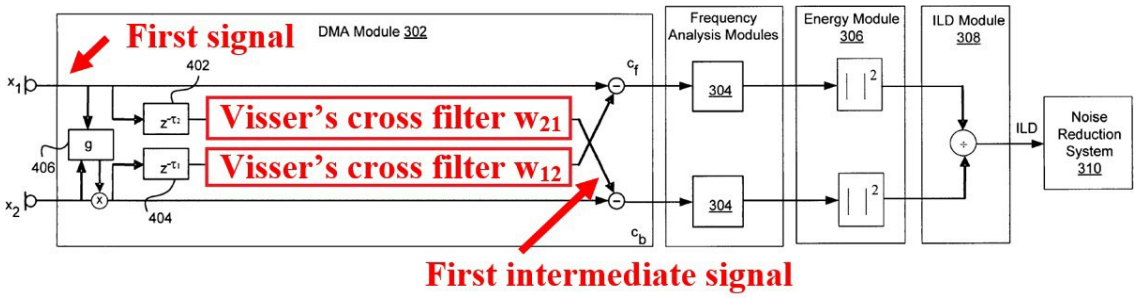


FIG. 4a

Ex. 1005, FIG. 4a (modified)

Further, according to the modification, a second virtual microphone is "cardioid secondary signal (C_b)," which is formed by subtracting the first intermediate signal from a second signal x_2. Ex. 1005, FIG. 4a.

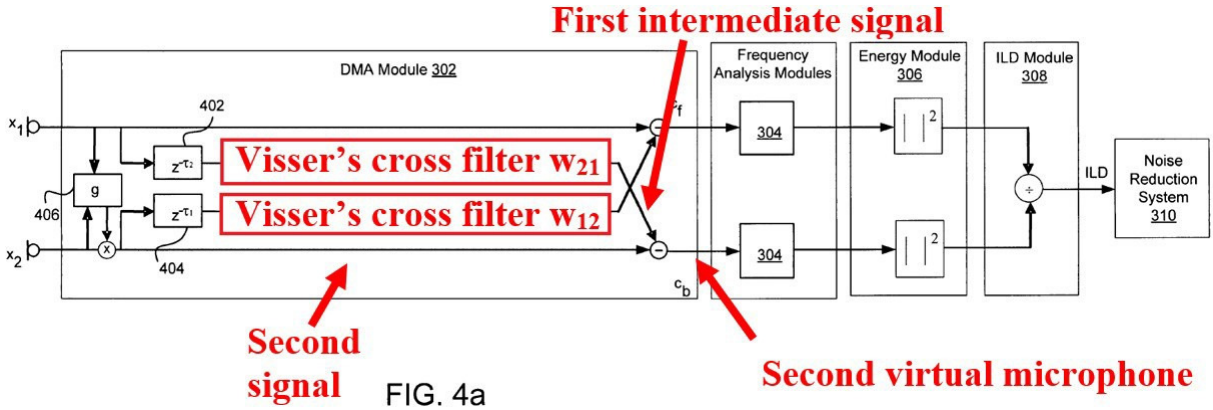


FIG. 4a

Ex. 1005, FIG. 4a (modified)

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

In Avendano, the energy of the first virtual microphone is E_f, which is calculated by integrating the first virtual microphone over time. Ex. 1005, 5:49-61; Ex. 1003 ¶¶163-166.

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

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

Ex. 1005, 5:61

The energy of the second virtual microphone is E_b, which is calculated by integrating 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**

Ex. 1005, 6:5

Further, the energy ratio is "Inter-Level Difference" (ILD), which 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

Ex. 1005, 6:16

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

Avendano compares the energy ratio (ILD) to a "threshold" to determine the presence or absence of speech. Ex. 1005, 6:35-7:9; Ex. 1003 ¶¶167-171.

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

Ex. 1005, 6:61

If the ratio (ILD) is "smaller than a threshold value (e.g., threshold=0.5) above which speech is expected to be," a value λ_l is set to zero (*e.g.,* indicating an absence of speech). *Id.* 6:58-7:3. However, if the ratio (ILD) "starts to rise (e.g., because speech is present within the large ILD region), λ_l increases" (*e.g.,* set to one, indicating a presence of speech). *Id.*

E. Claim 2

[2]: wherein the first virtual microphone and the second virtual microphone are distinct virtual directional microphones.

Avendano discloses a first virtual microphone ("cardioid primary signal (C_f)") and a second virtual microphone ("cardioid secondary signal (C_b)"). Sections III.D.[1a], [1c]. The first and second virtual microphones are distinct from one another and represent virtual directional microphones (*e.g.,* directed to "two different directional patterns about the audio device"). Ex. 1005 FIG. 4a, 4:27-52, 5:15-35; Ex. 1003 ¶¶172-175.

F. Claims 3 and 4

[3]: wherein the first virtual microphone and the second virtual microphone have approximately similar responses to noise.

[4]: wherein the first virtual microphone and the second virtual microphone have approximately dissimilar responses to speech.

Avendano's first virtual microphone has a directional pattern directed to the front of "audio device 104," and the second virtual microphone has a directional pattern directed to the back of "audio device 104." Sections III.D.[1a], [1c]; Ex. 1005, 4:47-52, 5:25-35, FIG. 6; Ex. 1003 ¶¶176-187.

Further, in Avendano, noise is typically generated in the "background" or "far field," rather than from a location near the audio device, and "may include reverberations and echoes." Ex. 1005, 2:49, 3:35-41, FIG. 1a.

A POSITA would have recognized that, according to Avendano's configuration, Avendano's virtual microphones would have approximately similar responses to noise. Ex. 1003 ¶¶179-182.

For example, although Avendano describes two virtual microphones having respective directional patterns, neither of these virtual microphones is directed to a "background" or "far field." Ex. 1005, 2:49, FIG. 1a. Accordingly, neither virtual microphone would be more sensitive to noise generated in the "background" or "far field" than the other. Ex. 1003 ¶180.

Further, according to Avendano's configuration, an audio device would not merely detect noise propagating directly from a noise source to the audio device, but would also detect reverberations and echoes of that noise from multiple other locations relative to the audio device. Ex. 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. Ex. 1003 ¶181.

A POSITA also would have recognized that, according to Avendano's configuration, Avendano's virtual microphones would have approximately dissimilar responses to speech. *Id.* ¶¶183-187.

In Avendano, speech is typically generated near the front of the device. Ex. 1005, 3:42-55, FIG. 1a. According to this configuration, Avendano's virtual

microphones would have approximately dissimilar responses to speech (*e.g.,* first virtual microphone is directed towards the source of speech, whereas the second virtual microphone is directed away from the source of speech). Ex. 1003 ¶187.

G. Claims 5 and 6

[5]: applying a calibration to at least one of the first signal and the second signal.

[6]: 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.

Avendano applies a calibration ("gain factor, g") to at least the second signal "to equalize the signal levels" of the first and second signals. Ex. 1005, 5:36-39, 9:54-59; FIGS. 4a-4b; Section III.D.[1a]; Ex. 1003 ¶¶188-195.

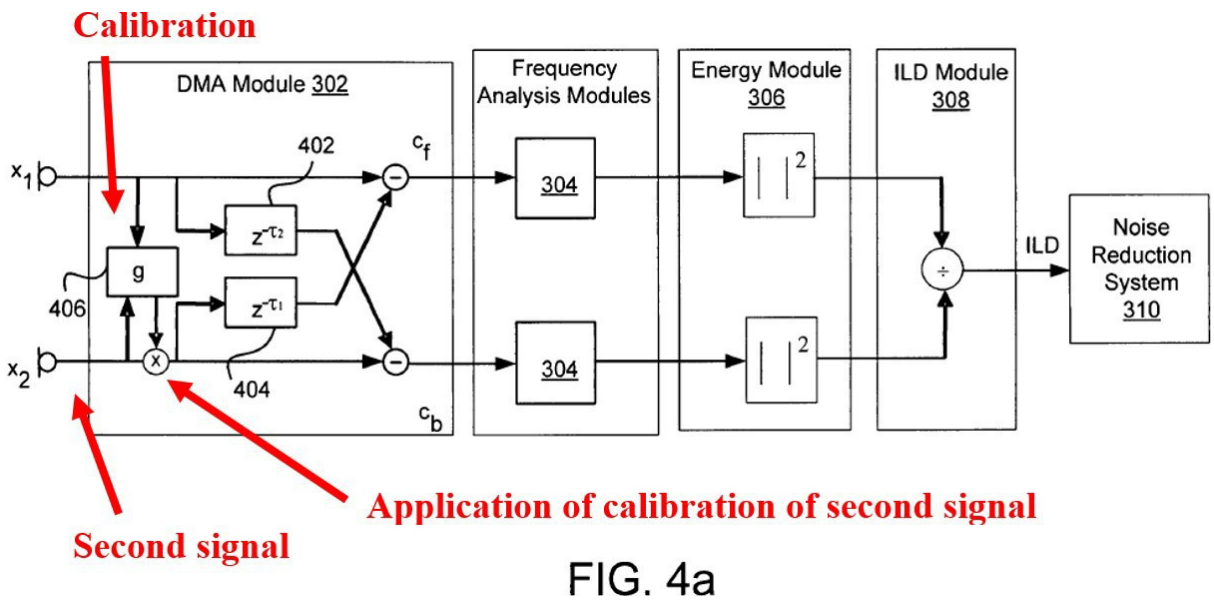


FIG. 4a

Ex. 1005, FIG. 4a

Accordingly, Avendano's 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. Ex. 1003 ¶195.

H. Claim 7

[7]: applying a delay to the first intermediate signal.

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. Section III.D.[1c]; Ex. 1005, 5:15-35, FIGS. 4a-4b; Ex. 1003 ¶¶196-201.

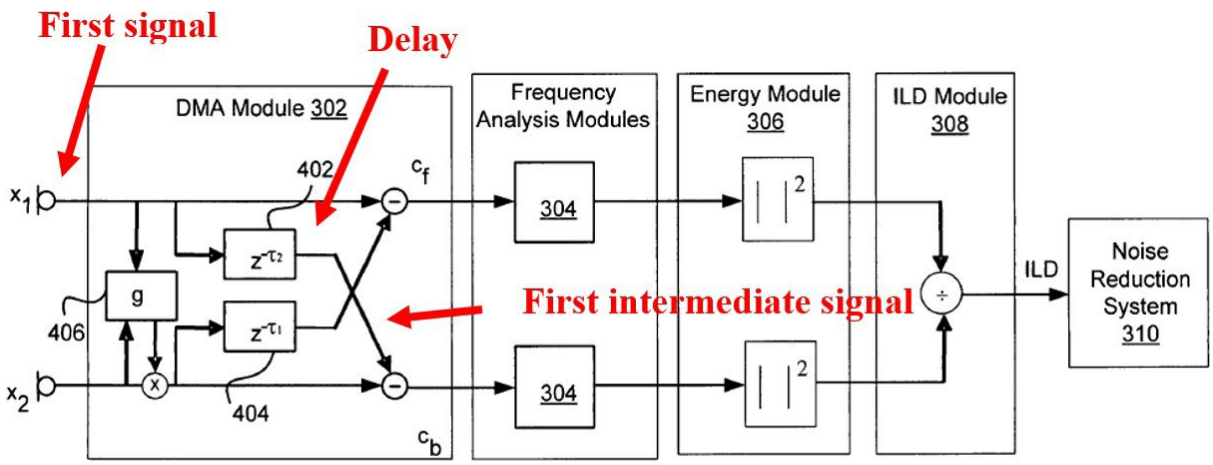


FIG. 4a

Ex. 1005, FIG. 4a

Accordingly, in Avendano, a delay has been applied to the first intermediate signal. Ex. 1003 ¶199. Further, Visser's "cross filters w_{12} and w_{21}" can be "gain factors with only 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." Ex. 1006, 17:58-62. 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}. Ex. 1003 ¶201.

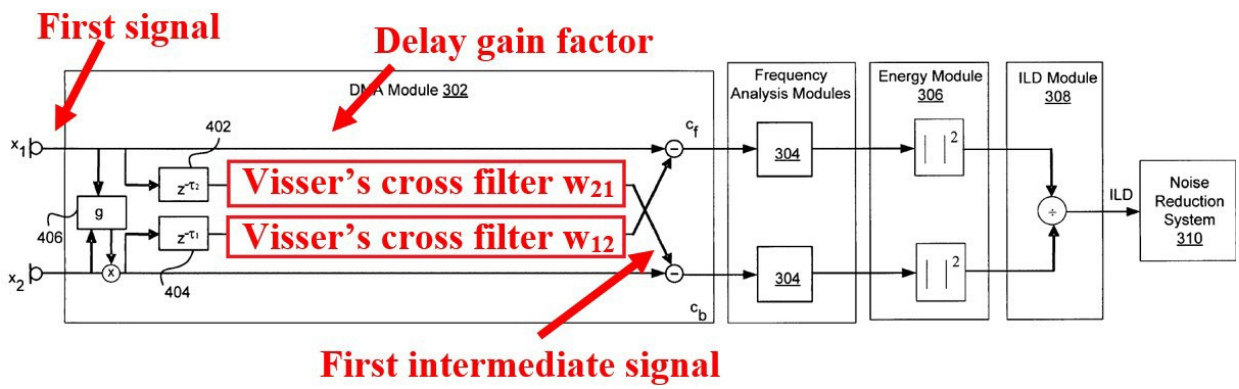


FIG. 4a

Ex. 1005, FIG. 4a (modified)

I. Claim 25

[25]: generating a vector of the energy ratio versus time.

Avendano's energy ratio (ILD) is generated as a vector over time (t). Ex. 1005, 6:8-34; Section III.D.[1d]; Ex. 1003 ¶¶202-204.

Energy ratio

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

Ex. 1005, 6:16

J. Claim 26

[26]: wherein the first and second physical microphones are omnidirectional microphones.

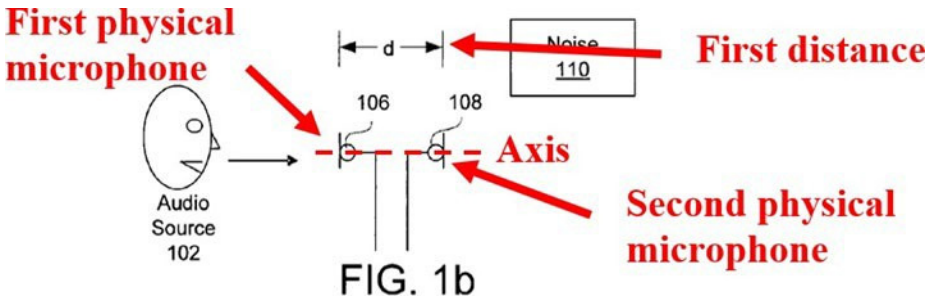
Avendano's physical microphones are "omni-directional microphone[s]." Ex. 1005 Abstract, 2:3-19, 3:33-35, 9:39-52; Ex. 1003 ¶¶205-207; Section III.D.[1a].

K. Claims 27 and 28

[27] 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.

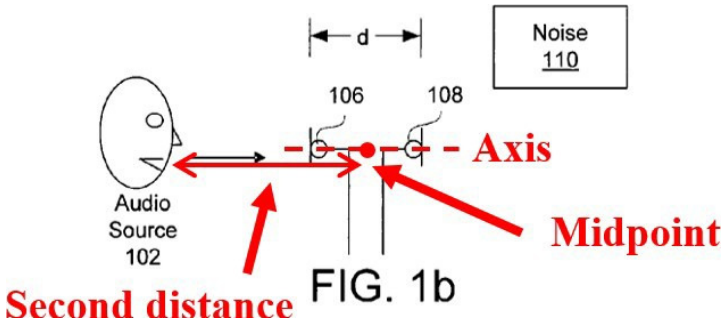
[28] wherein a midpoint of the axis is a second distance from a mouth of the speaker, wherein the mouth is located in a direction defined by an angle relative to the midpoint.

Avendano's first and second physical microphones are positioned along an axis, and separated by a first distance (d). Ex. 1005, FIG. 1b; Section III.D.[1a]; Ex. 1003 ¶¶208-214.



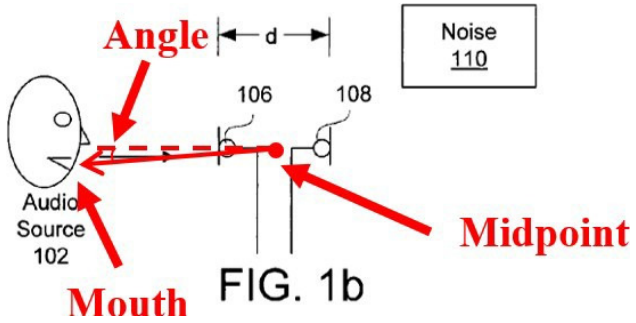
Ex. 1005, FIG. 1b

The midpoint of the axis is a second distance away from a mouth of a speaker. *Id.*



Ex. 1005, FIG. 1b

The mouth of the speaker is located in a direction defined by an angle relative to the midpoint. *Id.*



Ex. 1005, FIG. 1b

IV. GROUND 2: AVENDANO, VISSER, AND BISGAARD (CLAIMS 8-16, 23, 24)

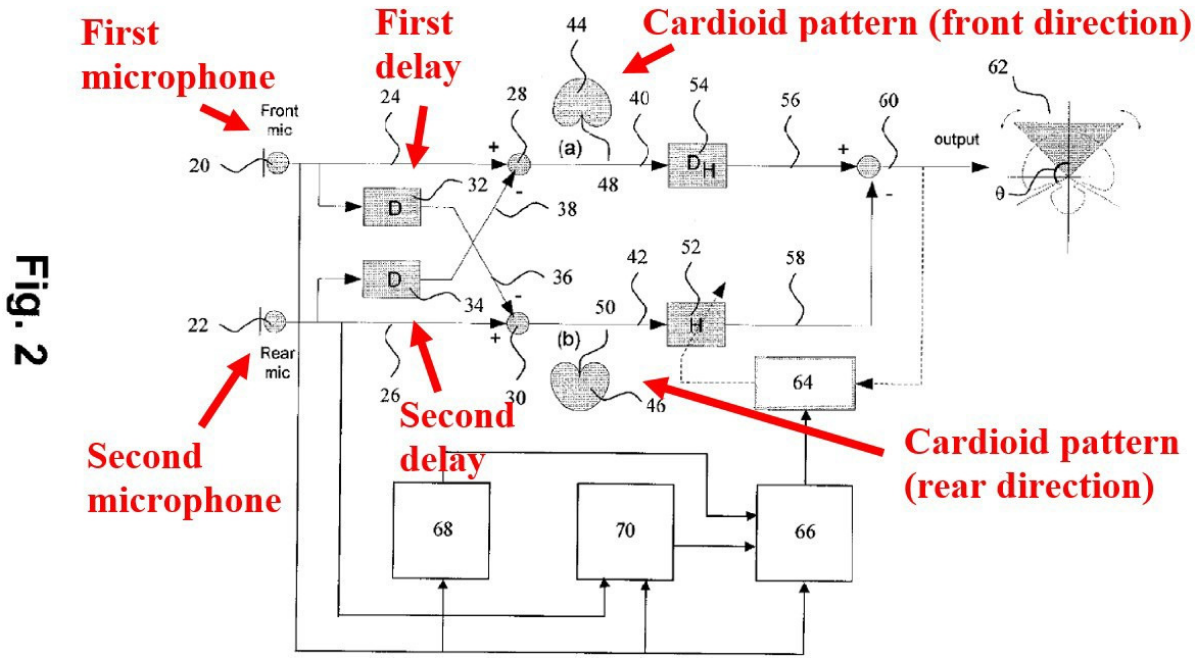
A. Bisgaard Overview

Bisgaard describes "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." Ex. 1010 ¶ [0002]; Ex. 1011, 1:3-5.

Bisgaard's system obtains signals from two "microphones 20, 22," and

processes the signals to obtain (i) "cardioid pattern 44" pointing towards a front of the device and (ii) "cardioid pattern 46" pointing towards a rear of the device." Ex. 1010, ¶¶ [0041]-[0042], FIG. 2; Ex. 1011, 5:18-31, FIG. 2; Ex. 1003 ¶79.

Further, Bisgaard includes "delay[s] 32, 34," which "delay[] the digitized sound signal [received from 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." Ex. 1010 ¶ [0042], FIG. 2; Ex. 1011, 5:21-24, FIG. 2; Ex. 1003 ¶¶78-81.



Ex. 1010, FIG. 2

B. Combination of Avendano, Visser, and Bisgaard

A POSITA would have found it obvious to combine Avendano and Visser with Bisgaard. Ex. 1003 ¶¶215-222. Like Avendano, Bisgaard receives signals from two

microphones, and processes the signals to obtain (i) a cardioid signal directed in a forward direction (*e.g.*, towards a user), and (ii) another cardioid signal directed in a backwards direction (*e.g.*, away from the user). Ex. 1005, 4:42-5:35, FIGS. 4a-4b, 6; Ex. 1010 ¶¶ [0041]-[0042], FIG. 2; Ex. 1011, 5:18-42, FIG. 2. Further, like Avendano, Bisgaard forms each of the cardioid signals by delaying the signal from one microphone, and combining it with the signal from the other microphone. *Id.*

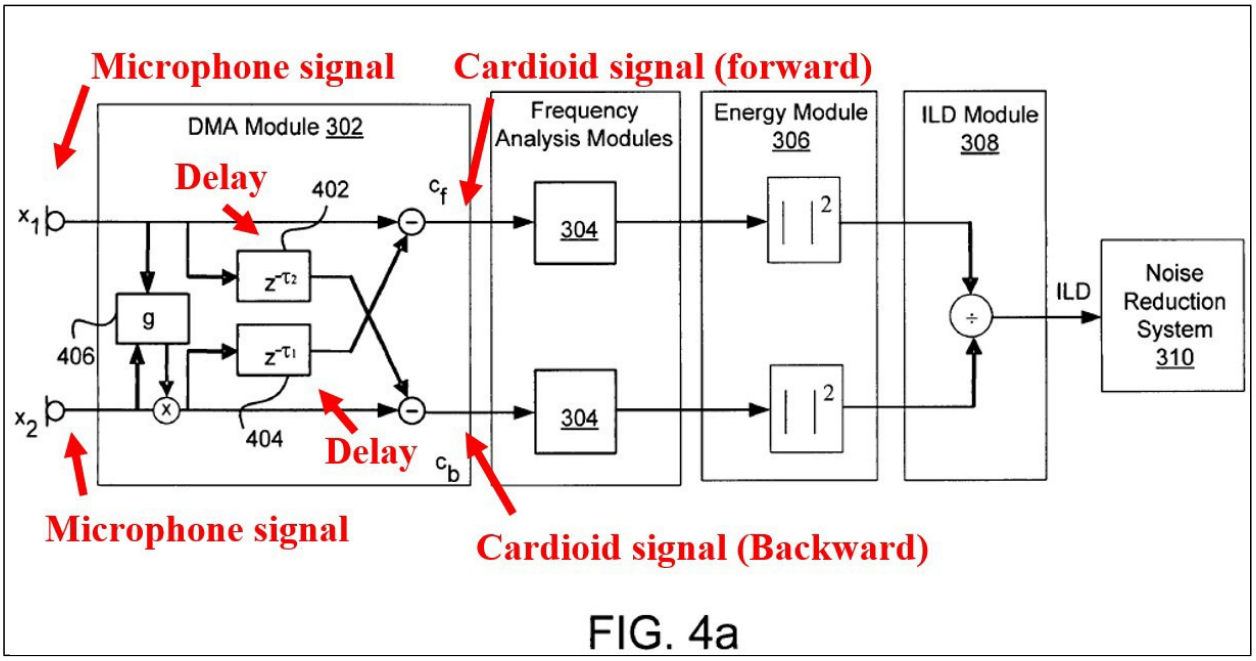
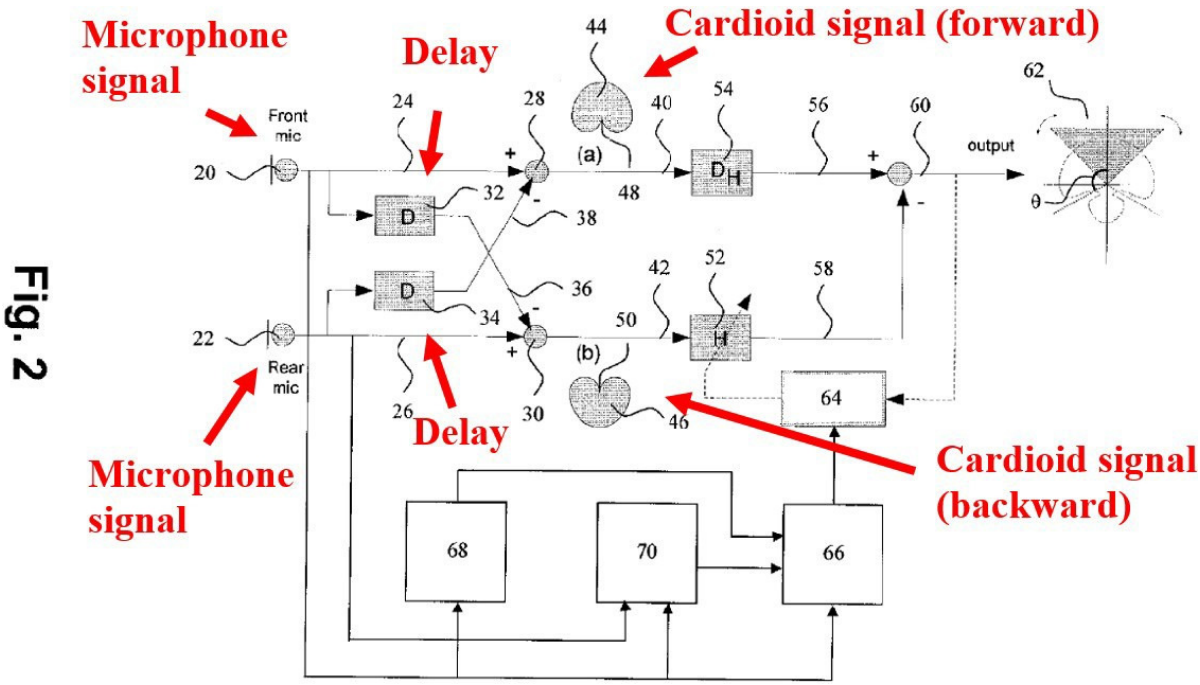


FIG. 4a

Ex. 1005, FIG. 4a



Ex. 1010, FIG. 2

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). Ex. 1003 ¶219. Accordingly, a POSITA would have looked to Bisgaard for details related to implementing such a delay in Avendano's system. *Id.*

From this disclosure, and 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 ¶220.

Further, given the similarities between Avendano's "delay nodes" and Bisgaard's "delays," a POSITA would have had a reasonable expectation of success in incorporating Bisgaard's "delays" in Avendano's system. Ex. 1003 ¶221. For example, as discussed above, Avendano's "delay nodes" and Bisgaard's "delays" are configured to receive similar input signals and generate similar outputs. Ex. 1005, 4:42-5:35, FIGS. 4a-4b, 6; Ex. 1010 ¶¶ [0041]-[0042], FIG. 2; Ex. 1011, 5:18-42, FIG. 2. Given these similarities, limited modifications would have been necessary to incorporate Bisgaard's "delays" into Avendano's system and other operations in Avendano's system would not have been impacted by the combination. Ex. 1003 ¶221.

Thus, 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. *Id.* ¶222.

C. Claim 8

[8] 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.

As Avendano describes, "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." Ex. 1005, 1:33-36. Further, "embodiments may use a combination of energy level differences and time delays to discriminate speech." *Id.* 3:56-58. Namely, Avendano delays a first signal x_1 by "delay node 402," and delays a second signal x_2 by "delay node 404." Section III.A; Ex. 1005, 15:35, FIG. 4a; Ex. 1003 ¶¶223-230.

A POSITA would have recognized that, to account for the "space difference between the microphones," the delays would need to be proportional to a time difference between the arrivals of the speech at the first and second physical microphones. Ex. 1003 ¶225. For example, a POSITA would have recognized that, if the "space difference between the microphones" were to increase, the delay would likewise proportionally increase to account for the increase in propagation time, and vice versa. *Id.*

To the extent that Avendano alone does not describe the features of [8], it would have been obvious to modify Avendano's system in view of Bisgaard to

include these features. *Id.* ¶¶226-230. A POSITA would have found it obvious to combine Avendano and Visser with Bisgaard (*e.g.*, by incorporating Bisgaard's "delays" into Avendano's system). Section IV.B.

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." Ex. 1010 ¶ [0042]; Ex. 1011, 5:22-24.

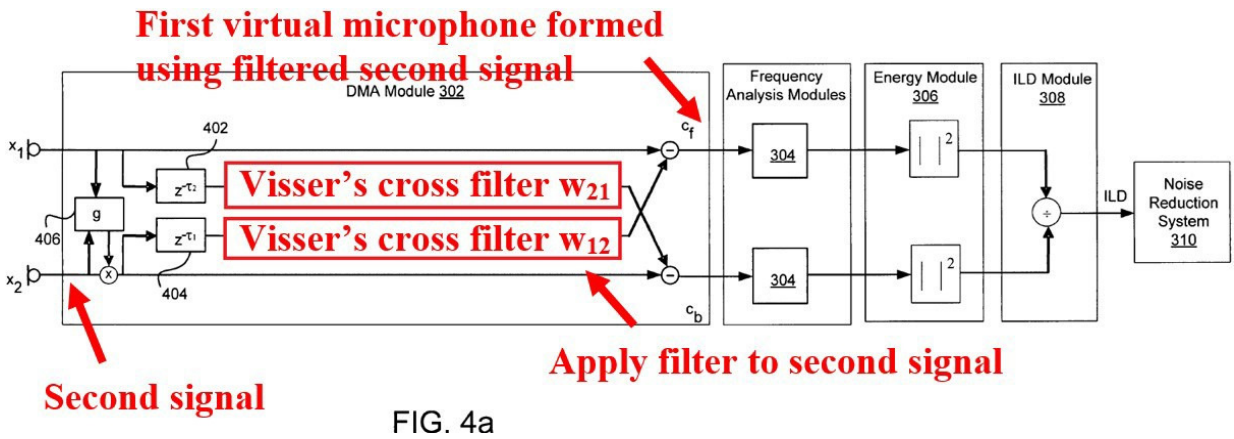
A POSITA would have recognized that, to account for "the amount of time used by a sound signal to propagate in the 0° azimuth direction from [a] front microphone ... to [a] rear microphone," the sound signal would be delayed by an amount of time that is proportional to a time difference between arrival of the speech at one physical microphone and arrival of the speech at another physical microphone. Ex. 1003 ¶¶229-230.

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, and vice versa. Ex. 1010 ¶[0042]; Ex. 1011, 5:22-24; Ex. 1003 ¶230.

D. Claim 9

[9]: wherein the forming of the first virtual microphone comprises applying the filter to the second signal.

In the combination, Visser's cross filter w_{12} would be applied to Avendano's second signal x_2. Sections III.C-D; Ex. 1003 ¶¶231-233. 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.*



Ex. 1005, FIG. 4a (modified)

E. Claim 10

[10]: wherein the forming of the first virtual microphone comprises applying the calibration to the second signal.

Avendano applies a calibration ("gain factor, g") to the second signal x_2. Section III.G; Ex. 1003 ¶¶234-236.

Further, Avendano's calibrated second signal x_2 is used to form the first virtual microphone C_f (*e.g.,* by combining the calibrated second signal x_2 with the first signal x_1). Ex. 1005, 4:47-52, 5:25-35.

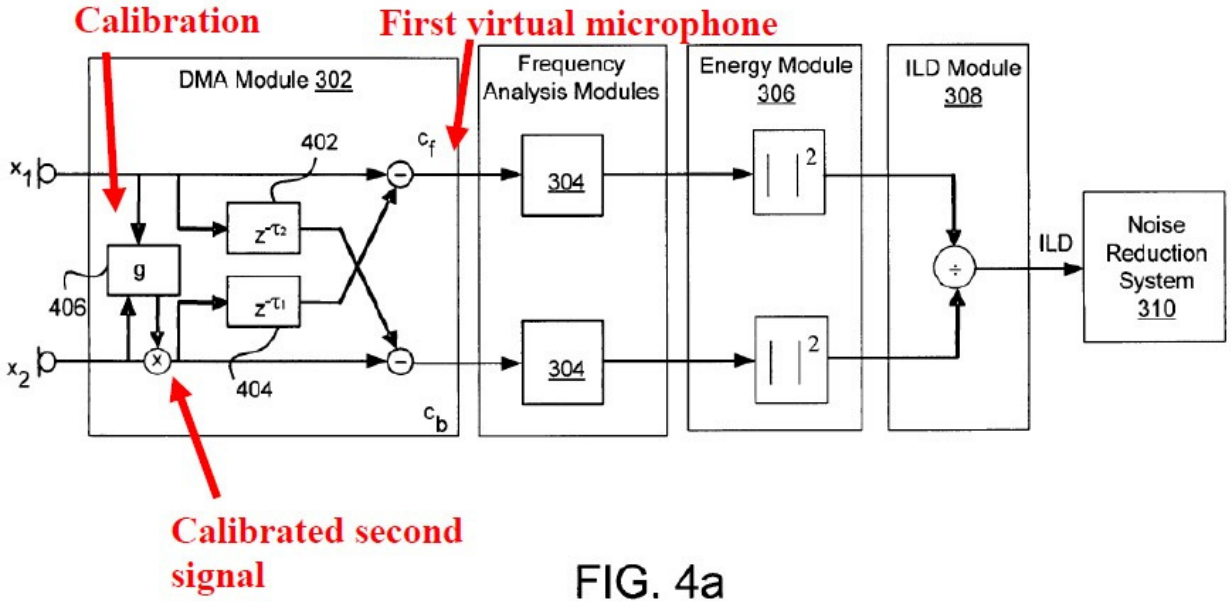


FIG. 4a

Ex. 1005, FIG. 4a

F. Claim 11

[11]: wherein the forming of the first virtual microphone comprises applying the delay to the first signal.

In Avendano's system, "DMA Module 302" forms the first virtual microphone C_f. Ex. 1005, 4:20-5:39, FIGS. 3, 4a. Further, in a "practical implementation of the DMA module 302" shown in FIG. 4b, a delay is applied to the first signal x_1 via "delay node 414." *Id.* 8:38-51, FIG. 4b; Ex. 1003 ¶¶237-239.

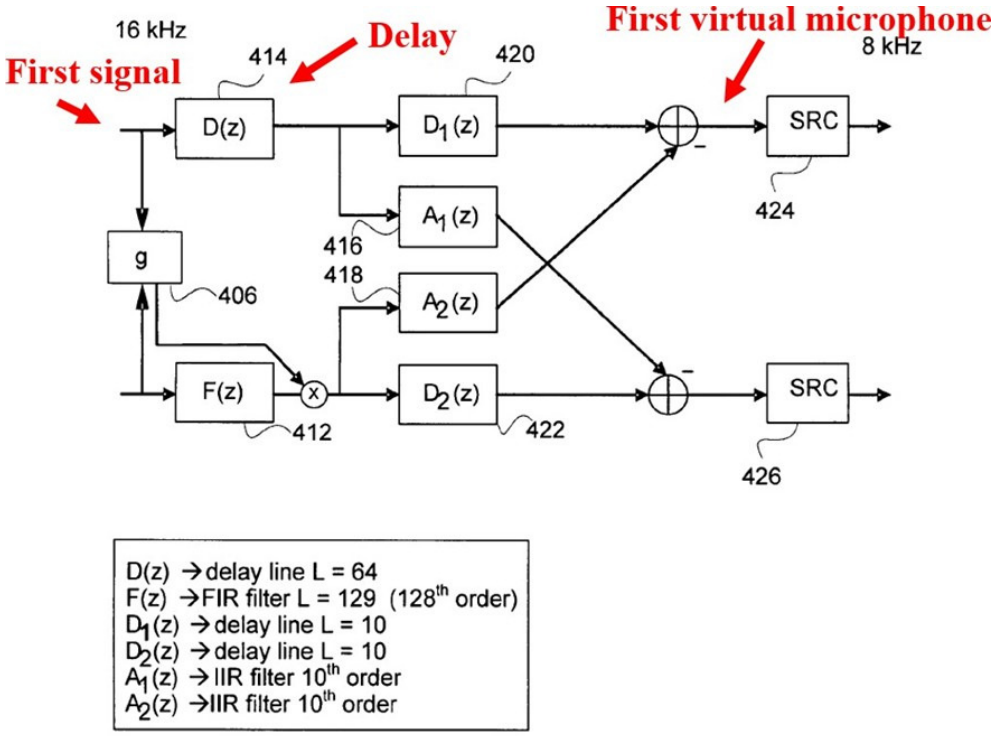


FIG. 4b

Ex. 1005, FIG. 4b

G. Claim 12

[12] wherein the forming of the first virtual microphone by the combining comprises subtracting the second signal from the first signal.

Avendano's first virtual microphone C_f is formed by subtracting second signal x_2 from first signal x_1. Ex. 1005, 5:15-35, FIG. 4a; Ex. 1003 ¶¶240-241.

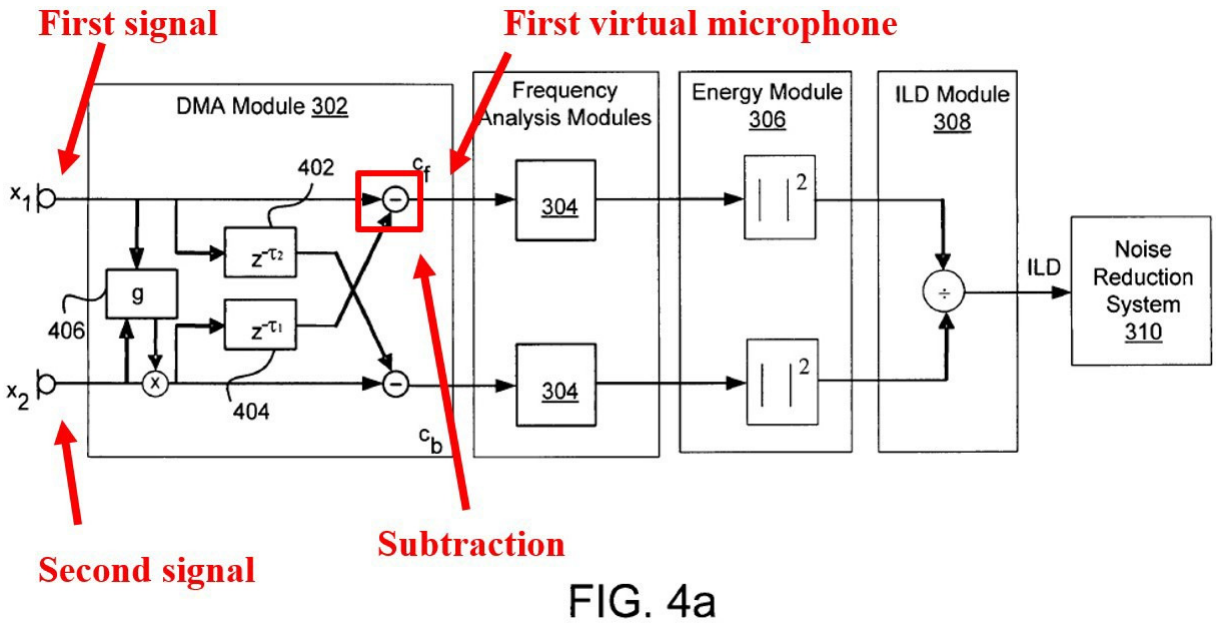


FIG. 4a

Ex. 1005, FIG. 4a

H. Claim 13

[13]: wherein the filter is an adaptive filter.

Visser's cross filters w_{12} and w_{21} are adaptive filters. Ex. 1006, 8:16-18, 16:3-28, 16:57-25:9, FIGS. 5, 7, 10-13; Ex. 1003 ¶¶242-244. For example, Visser's cross filters w_{12} and w_{21} are adapted in an "ICA or BSS processing function." Ex. 1006, 17:29-32. An "ICA process" includes filters with "adaptable and adjustable filter coefficient[s]," which are adapted using "learning stage 752." *Id.* 16:61-17:4; 21:17-58, FIG. 12.

I. Claim 14

[14]: 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.

Visser's cross filters w_{12} and w_{21} are adapted to filter signals received from two physical microphones ("[i]nput signals X_1 and X_2") to generate (i) a first virtual microphone that would "contain[] predominantly desired signals," namely speech ("speech channel"), and (ii) a second virtual microphone that would "contain[] predominantly noise signals" ("noise channel"). Section IV.H; Ex. 1006, 17:36-44, 16:61-17:4, 21:17-58; Ex. 1003 ¶¶245-250.

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. Ex. 1003 ¶250.

J. Claim 15

[15]: wherein the adapting comprises applying a least-mean squares process.

Visser's "adaptive least mean square (NLMS) algorithm" is used to "build[] a linear filter model." Ex. 1006, 23:13-25. 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"). Ex. 1003 ¶¶251-254.

K. Claim 16

[16]: generating coefficients of the filter during a period when only speech is being received by the first physical microphone and the second physical microphone.

Visser generates coefficients of the filter during a period when only speech is being received. Ex. 1003 ¶¶255-261. For example, Visser "turn[s] off" an "ICA module" for adapting filters "when desired speech is not present, ... thereby enabling adaptation only when such adaptation will be able to achieve a separation improvement." Ex. 1006, 9:12-43. As Visser explains, this "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," "adds significant robustness to the methodology," and conserves "processing and battery power." *Id.*

L. Claim 23

[23]: wherein the filter is a static filter.

Avendano's "delay node 402" describes a relationship for speech between the first and second physical microphones. Section III.D.[1b]; Ex. 1005, 4:28-34, 5:19-35, FIGS. 4a-4b; Ex. 1003 ¶¶262-274. A POSITA would have recognized that, if the first and second physical microphones 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. Ex. 1003 ¶264. Accordingly, a POSITA would have found it obvious to implement Avendano's delay as a static filter. *Id.*

To the extent that Avendano alone does not describe a static filter, the combination of Avendano and Visser (and Bisgaard) would have rendered this feature obvious. *Id.* ¶¶265-271. A POSITA would have found it obvious to implement Visser's cross filters w_{12} and w_{21} as static filters in at least some circumstances. *Id.* ¶¶267-271.

Visser's "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." Ex. 1006, 16:23-27. For example, Visser's filters can have "filter values ... or taps," and in cases that the device "has only a limited range of operating conditions" (*e.g.,* limited changes in "the distance from each microphone to the speaker's mouth"), "default values" can be selected for the "taps ... to account for the expected operating arrangement." *Id.* 22:7-25, FIG. 12. As Visser explains, "the default values <u>may</u> adapt over time and according to environment conditions." *Id.*

However, a POSITA would have recognized that such an adaption would not necessarily be required (*e.g.,* 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). *Id.* 16:23-27, 22:7-25, FIG. 12; Ex. 1003 ¶¶268-271. Accordingly, a POSITA would have found it obvious to implement the delay as a static filter under these circumstances. *Id*.

To the extent that the combination of Avendano and Visser does not describe a static filter, the combination of Avendano, Visser, and Bisgaard would have rendered this feature obvious. Ex. 1003 ¶¶272-274.

Specifically, Bisgaard provides details regarding how to implement Avendano's 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." Sections IV.B-C; Ex. 1010, [0042]; Ex. 1011, 5:22-24; Section III.D.[1b]; Ex. 1005, 4:28-34, 5:19-35, FIGS. 4a-4b.

A POSITA would have recognized that, if the first and second microphones do not move relative to one another or to 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. Ex. 1003 ¶274.

Accordingly, a POSITA would have found it obvious to implement the delay as a static filter. *Id.*

M. Claim 24

[24]: 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.

A POSITA would have recognized that, 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. Section IV.C; Ex. 1010 ¶[0042]; Ex. 1011, 5:22-24; Ex. 1003 ¶¶275-279.

Further, a POSITA would have recognized that, given constant environment conditions, sound would propagate through an environment at a constant speed. Ex. 1003 ¶278. 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. *Id*. 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. *Id.*

Accordingly, to determine a time difference between arrivals of the speech at two physical microphones, 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), (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 (*e.g.,* representing a proportional relationship between the first distance and the second distance). *Id.* ¶279.

V. GROUND 3: AVENDANO, VISSER, BISGAARD, AND HOU (CLAIMS 17- 19)

A. Hou Overview

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

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

Further, Hou's system receives "a second electronic sound signal" from "microphone 504," delays the "second electronic sound signal" using "delay

unit 516," and "estimates the minimum for the second electronic sound signal" using "second minimum estimate unit 510." *Id.*

Further, "divide unit 512 produces a quotient by dividing the first minimum estimate by the second minimum estimate," where "[t]he quotient represents a scaling amount that is sent to a multiplication unit 514." *Id.* 5:42-45.

Further, "[t]he second electronic sound signal is then multiplied with the scaling amount to produce a compensated sound signal." *Id.* 5:45-47. As Hou describes, "[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.

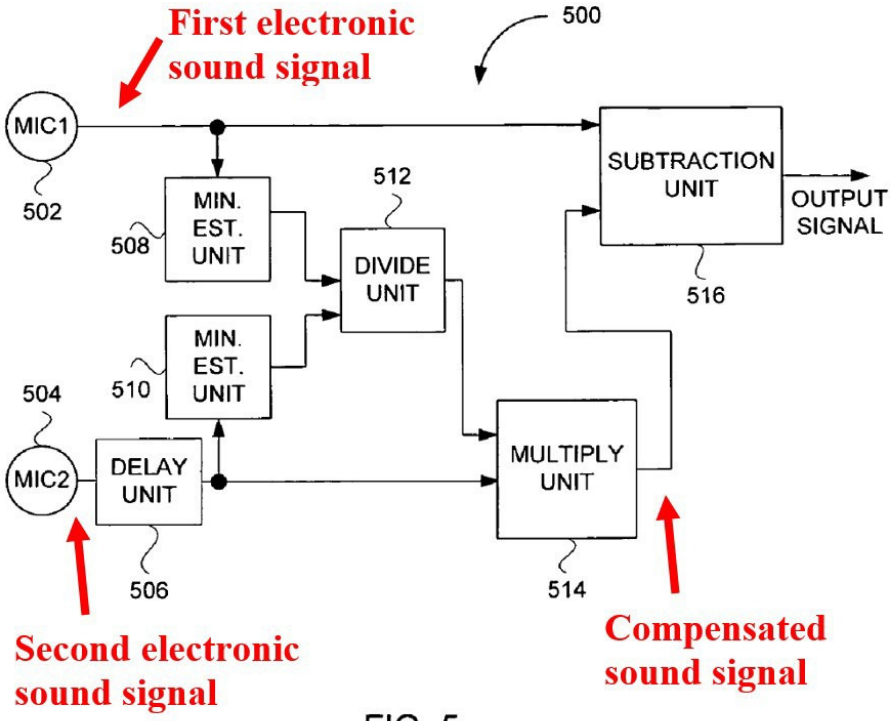


FIG. 5

Ex. 1008, FIG. 5

Further, "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.

As Hou describes, generating a compensated sound signal ensures 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.* 9:28-41; Ex. 1003 ¶¶82-91.

B. Combination of Avendano, Visser, Bisgaard, and Hou

A POSITA would have found it obvious to combine Avendano, Visser, and Bisgaard with Hou. Ex. 1003 ¶¶280-294.

First, both Avendano and Hou come from the same field of endeavor of enhancing speech and attenuating noise. Ex. 1005 Abstract, 1:24-26, 3:13-26, 3:42-60; Ex. 1008 Abstract, 2:44-52, 4:40-49; Ex. 1003 ¶282.

Second, both Avendano and Hou describe enhancing speech and attenuating noise using similar techniques, including equalizing the signal levels of two signals and generating directional signals based on the equalized signals. Ex. 1003 ¶¶283-288.

For example, Avendano's "primary microphone 106 is much closer to [an] audio source 102 than the secondary microphone 108," and thus "the intensity level is

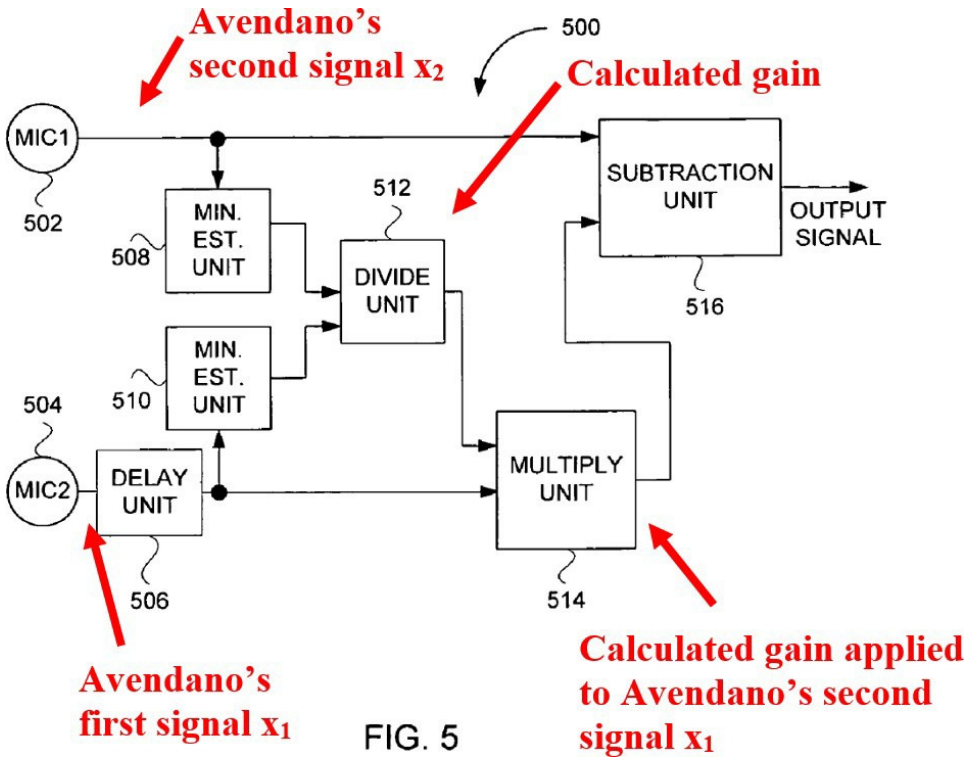
higher for the primary microphone 106 resulting in a larger energy level during a speech/voice segment." Ex. 1005, 3:27-49; Section III.A. To account for the differences in intensity levels, Avendano applies "gain factor, g" to a second signal "to equalize the signal levels" of the first and second signals. Ex. 1005, 5:36-39, 9:54-59; FIGS. 4a-4b. Such equalization is beneficial, as "systems can suffer loss of performance when the microphone signals have different levels." *Id.* 5:37-39. Further, Avendano uses the equalized signal to generate directional signals, such as "cardioid primary signal (C_f)" (front) and "cardioid secondary signal (C_b)" (back). *Id.* 41-52, 5:25-35, FIGS. 4a-4b, 6.

Hou's "two-microphone directional processing system 500" 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." Ex. 1008, 5:47-50; Ex. 1003 ¶287. Similarly, Hou explains that 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." Ex. 1008, 1:48-2:2.

Like Avendano, Hou processes the compensated or corrected signal 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.

A POSITA would have recognized that Avendano teaches the importance of equalizing signal levels of signals generated by two respective microphones, and that Hou discloses a specific "two-microphone directional processing system 500" for performing precisely this function. Ex. 1003 ¶289. Accordingly, a POSITA would have been motivated to implement Hou's signal compensation or correction components in Avendano to further improve the equalization of the signal output by the microphones (*e.g.,* to generate a signal "hav[ing] robust directionality" and to enhance "directional noise suppression"). Ex. 1008, 5:50-56, 9:28-41; Ex. 1003 ¶289.

As 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." Ex. 1003 ¶290. In particular, Hou's "first minimum estimate unit 508" would receive Avendano's signal x_2, and Hou's "second minimum estimate unit 510" would receive Avendano's signal x_1. *Id*. Further, Hou's "divide unit 512" would calculate a gain as the quotient of the output of "first minimum estimate unit 508" and the output of "second minimum estimate unit 510." *Id*. Further still, the calculated gain would be applied to Avendano's signal x_1 by "multiplication unit 514." *Id*.



Ex. 1008, FIG. 5

Although Hou multiplies one of the signals by the calculated gain, a similar equalization effect could be achieved by multiplying the other signal by the inverse of the calculated gain. Ex. 1003 ¶291.

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). Ex. 1005, 4a; Ex. 1003 ¶292.

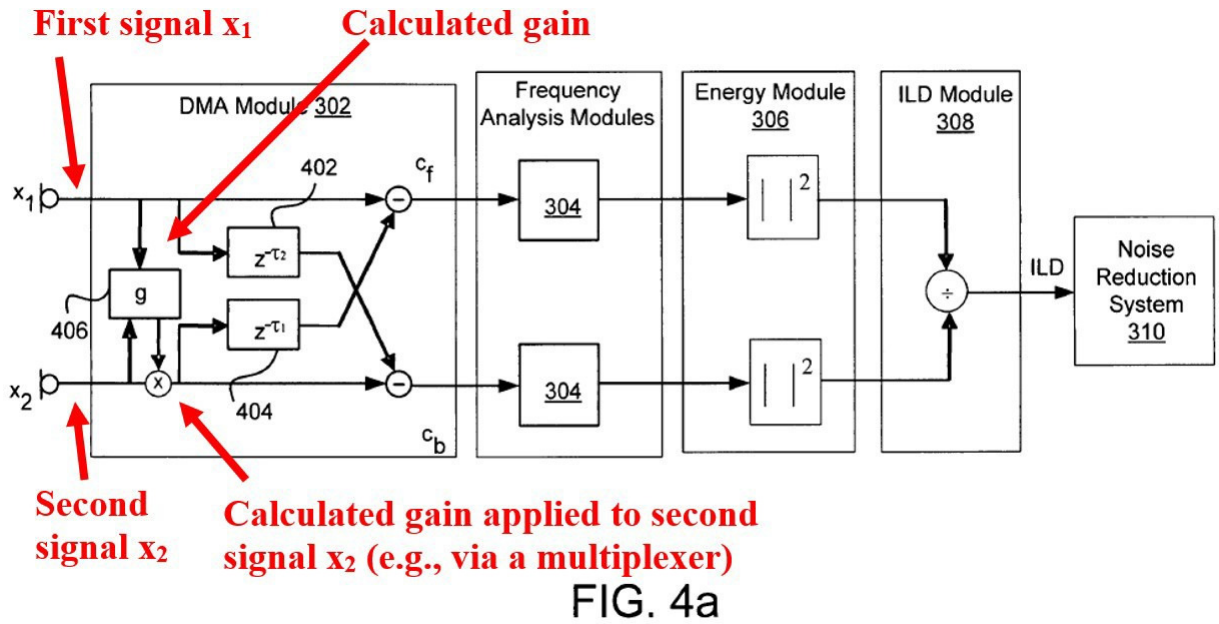


FIG. 4a

Ex. 1005, FIG. 4a

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. Ex. 1003 ¶293. For example, Avendano's "gain factor, g" and Hou's "two-microphone directional processing system 500" receive similar input signals (*e.g.,* from two microphones), and generate similar outputs (*e.g.,* "equalized" or "compensated" signals). Ex. 1005, 5:36-39, 9:54-59; FIGS. 4a-4b; Ex. 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 and other operations in

Avendano's system would not have been impacted by the combination. Ex. 1003 ¶293.

Thus, 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. *Id.* ¶294.

C. Claim 17

[17]: 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.

Hou's "first minimum estimate unit 508" would 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; Ex. 1008, 5:38-39, FIG. 5; Ex. 1003 ¶¶295-299.

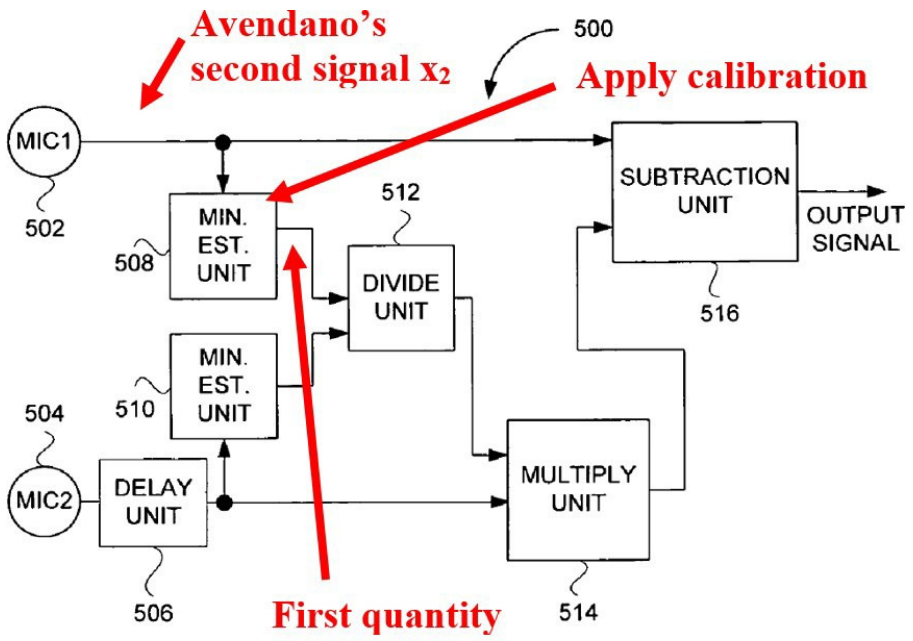
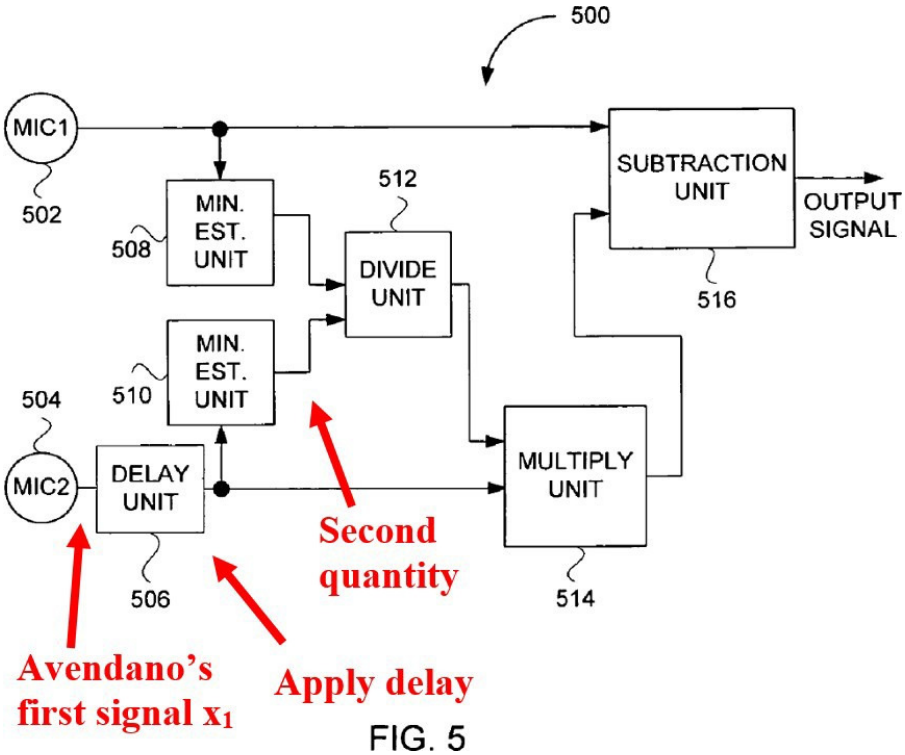


FIG. 5

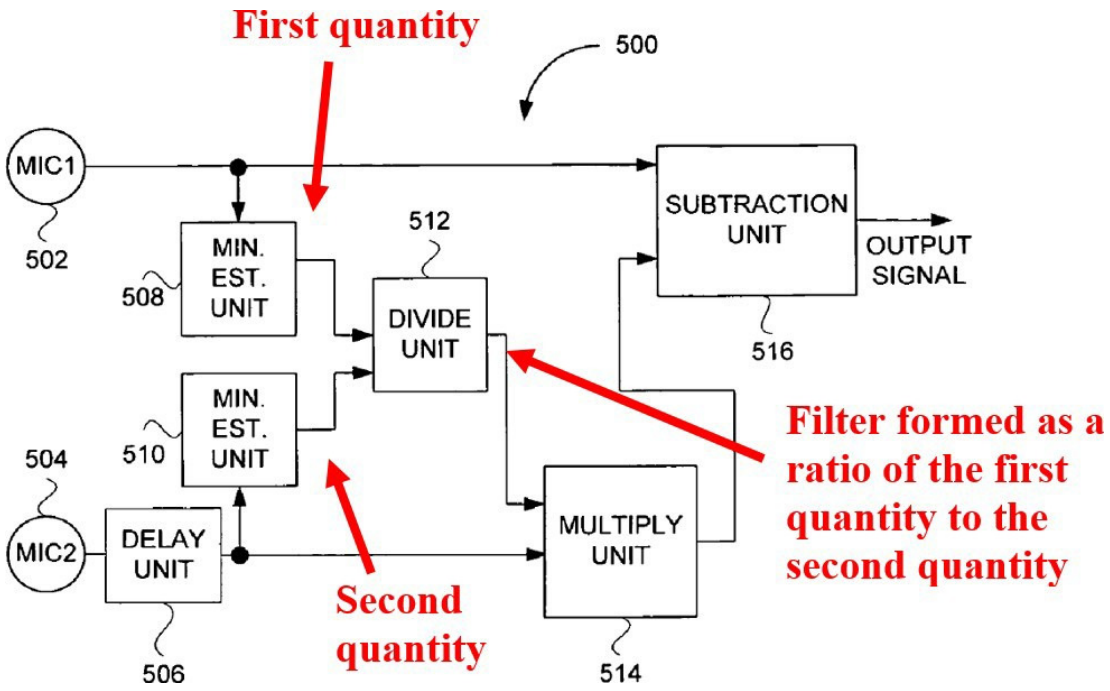
Ex. 1006, FIG. 5

Further, Hou's "delay unit 506" would 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_1 to generate a second quantity. Section V.B; Ex. 1008, 5:36-38, FIG. 5; Ex. 1003 ¶298.



Ex. 1006, FIG. 5

Further, Hou's "divide unit 512" forms a filter as a ratio of the first quantity to the second quantity. Section V.B; Ex. 1008, 5:42-45, FIG. 5. Specifically, Hou's "divide unit 512 produces a quotient by dividing the first minimum estimate by the second minimum estimate," "[t]he quotient represent[ing] a scaling amount that is sent to a multiplication unit 514." Ex. 1008, 5:42-45.



Ex. 1006, FIG. 5

D. Claims 18 and 19

[18]: wherein the generating of the energy ratio comprises generating the energy ratio for a frequency band.

[19]: wherein the generating of the energy ratio comprises generating the energy ratio for a frequency subband.

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." Ex. 1005, 4:55-5:2; Ex. 1003 ¶¶300-311; Section III.D.[1d].

Specifically, Avendano's "frequency analysis module 304 separates the cardioid signals into frequency bands," and performs "sub-band analysis on the acoustic signal [to] determine[] what individual frequencies are present in the

complex acoustic signal during a frame." *Id.*

"Once the frequencies are determined," Avendano's "energy module 306" calculates the energy ratio ("ILD") based on signals that are output by "frequency analysis module 304." Ex. 1005, 5:3-10. Specifically, the energy ratio is calculated "based on bandwidth of the cochlea channel" (*e.g.,* frequency band(s) or subband(s) for which "frequency analysis module 304 ... mimics the frequency analysis of the cochlea"). *Id.*; Ex. 1003 ¶¶303, 309.

Accordingly, Avendano generates the energy ratio for a frequency band or subband. Ex. 1003 ¶¶304, 310.

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

Energy ratio

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

Ex. 1005, 6:16

VI. GROUND 4: AVENDANO, VISSER, BISGAARD, HOU, AND FREQUENCY ART (BYRNE, BURNETT, AND/OR BERGLUND) (CLAIMS 20-22)

A. Byrne Overview

Byrne describes "[t]he long-term average speech spectrum (LTASS) ... for 12 languages," including "[LTASS] values for five samples of English." Ex. 1009 Abstract, 2112-2120, FIG. 1. For these samples, the average of the sound pressure level (SPL)—corresponding to an intensity of sound—peaked at approximately 500 Hz, and included significant spectral components in frequency ranges that include 500 Hz (*e.g.*, in frequencies greater than 200 Hz, between 250-1250 Hz, and between 200-3000 Hz). *Id.*; Ex. 1003 ¶¶92-95.

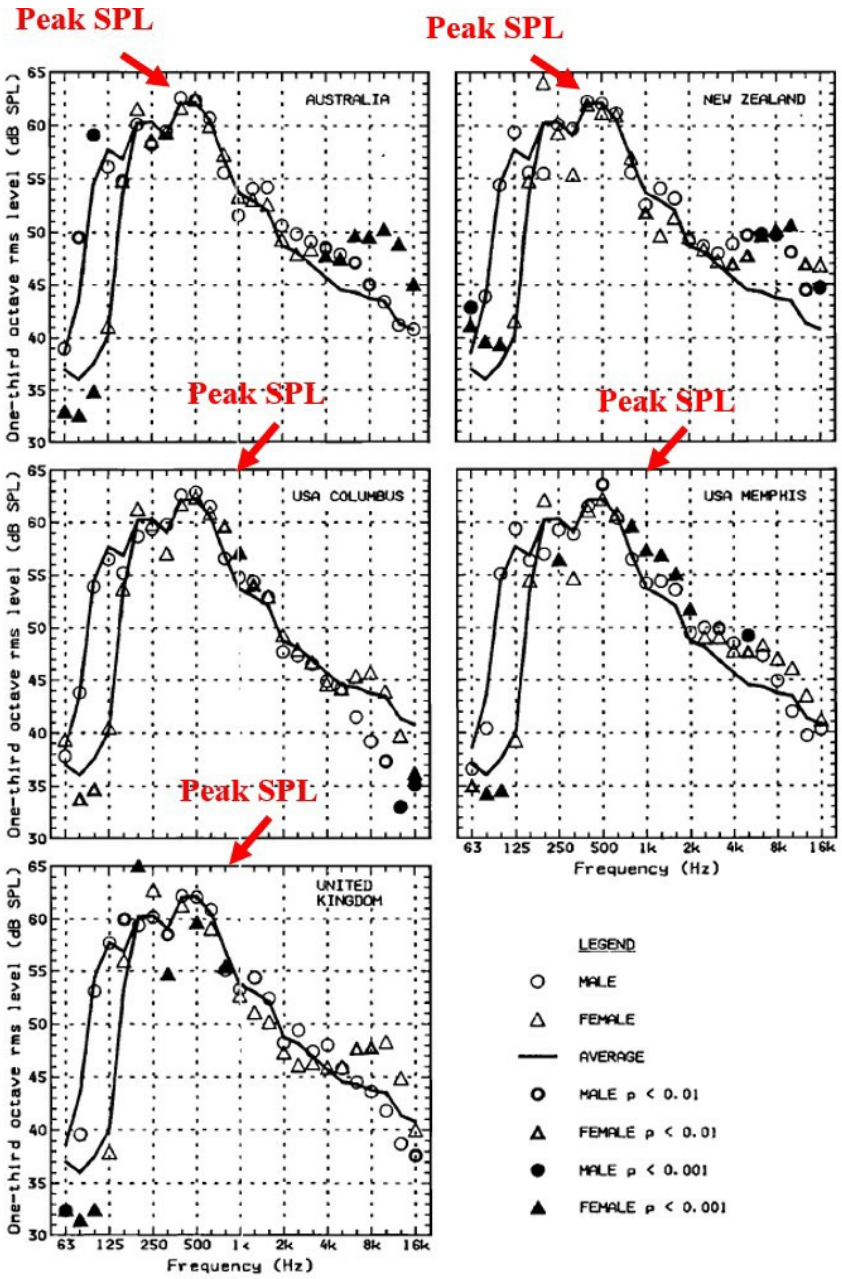


FIG. 1. Male and female long-term average speech spectrum (LTASS) values for five samples of English. Solid line shows LTASS average across 17 speech samples (all samples except Arabic), males and females separately for frequencies below 160 Hz, combined for higher frequencies.

Ex. 1009, FIG. 1

Further, Byrne's LTASS values vary for different types of speech (*e.g.,* speech uttered by different genders, according to different languages, etc.). *Id.* FIGS. 1-9, 2112-2120.

B. Burnett Overview

Burnett describes "[s]ystems and methods ... for detecting voiced and unvoiced speech in acoustic signals having varying levels of background noise." Ex. 1012 Abstract. Burnett's system "group[s] ... utterances by their spectral characteristics," which enables the system to "work better in noisy environments." *Id.* ¶[0063]. Specifically, Burnett's system "bandpass[es] 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." *Id.* Example frequency bands include (i) 500-4000 Hz corresponding to "k" in "kick" (ii) 1700-4000 Hz corresponding to "sh" in "she," (iii) 300-2500 Hz corresponding to "/i/ ('ee')," and (iv) 900-1200 Hz corresponding to "/a/ ('ah')." *Id.* ¶[0064]; Ex. 1003 ¶¶96-97.

C. Berglund Overview

Berglund describes "[t]he source of human exposure to low-frequency noise and its effects." Ex. 1013 Abstract.

For example, Berglund's Figure 3 shows the spectral components of common noises experienced by passengers of "road transportation vehicles":

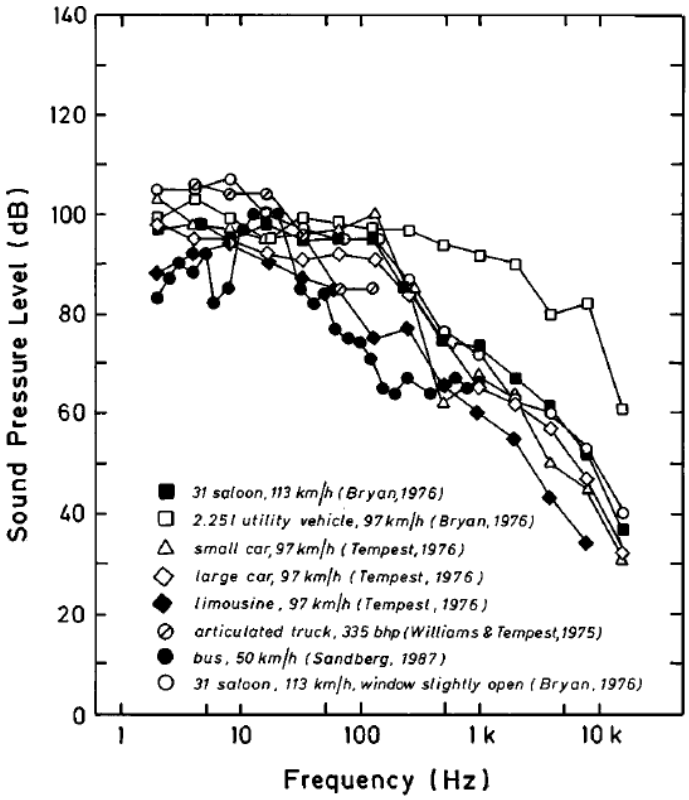


FIG. 3. Passenger noise exposure in road transportation vehicles as a function of frequency.

Ex. 1013, FIG. 3

Berglund's 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.,* less than approximately 250 Hz). Ex. 1013, 2985-2987, FIGS. 1, 3; Ex. 1003 ¶¶98-100.

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

A POSITA would have found it obvious to combine Avendano, Visser, Bisgaard, and Hou with one or more of Byrne, Burnett, and Berglund. Ex. 1003 ¶¶312-323.

A POSITA would have recognized that the processes described by Avendano (and in combination with Visser, Bisgaard, and Hou) are configured to analyze audible speech in specific frequency range(s). Ex. 1003 ¶314; Sections III.D, V.D; Ex. 1005, 4:55-5:10. Accordingly, a POSITA would have found it obvious to perform these processes in frequency range(s) that are known to have significant spectral components of speech in order to detect the presence of speech. Ex. 1003 ¶314. For example, in view of Byrne, a POSITA would have found it obvious to perform these processes 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. *Id.* Example frequency ranges include (i) frequencies greater than 200 Hz, (ii) frequencies between 250-1250 Hz, and (iii) frequencies between 200-3000 Hz. *Id.*

A POSITA also would have found it obvious to filter out frequency range(s) that are known to have significant spectral components of noise to reduce the likelihood of false positives in a speech detection process. *Id.* ¶315. For example, in view of Berglund, a POSITA would have found it obvious to filter out frequencies corresponding to common noises experienced by passengers of "road transportation vehicles," such as in the "low frequency noise" range (*e.g.,* less than approximately 250 Hz). Ex. 1013, FIG. 3; Ex. 1003 ¶315.

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. Ex. 1003 ¶316.

Further, a POSITA would have found it obvious to tune the frequency range to accommodate the specific operational requirements of the system. *Id.* ¶¶317- 323.

For example, 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. *Id.* ¶318. 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. *Id.*

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." Ex. 1012, [0063]. As Burnett describes, 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.* In

particular, a system can use bandpass filters to capture signals in certain frequency ranges to detect unvoiced speech (*e.g.,* 500-4000 Hz to detect "k" in "kick," and 1700-4000 Hz to detect "sh" in "she"), and in other frequency ranges to detect voiced speech (*e.g.,* 300-2500 Hz to detect "ee," and 900-1200 Hz to detect "ah"). *Id.* In view of Burnett, a POSITA would have found it obvious to tune the frequency range, depending on the specific unvoiced speech and/or voiced speech that is to be detected. Ex. 1003 ¶¶319-322.

E. Claim 20

[20]: wherein the frequency subband includes frequencies higher than approximately 200 Hertz (Hz).

A POSITA would have found it obvious to perform the processes described by Avendano, Visser, Bisgaard, and Hou 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. Section VI.D; Ex. 1003 ¶¶324-329. This would have resulted in performing the processes in a frequency subband that includes frequencies higher than approximately 200 Hz. *Id.*

For example, as taught by Byrne, the spectral components of English language speech are predominantly in a frequency range greater than 200Hz. Ex. 1009, FIG. 1; Ex. 1003 ¶326.

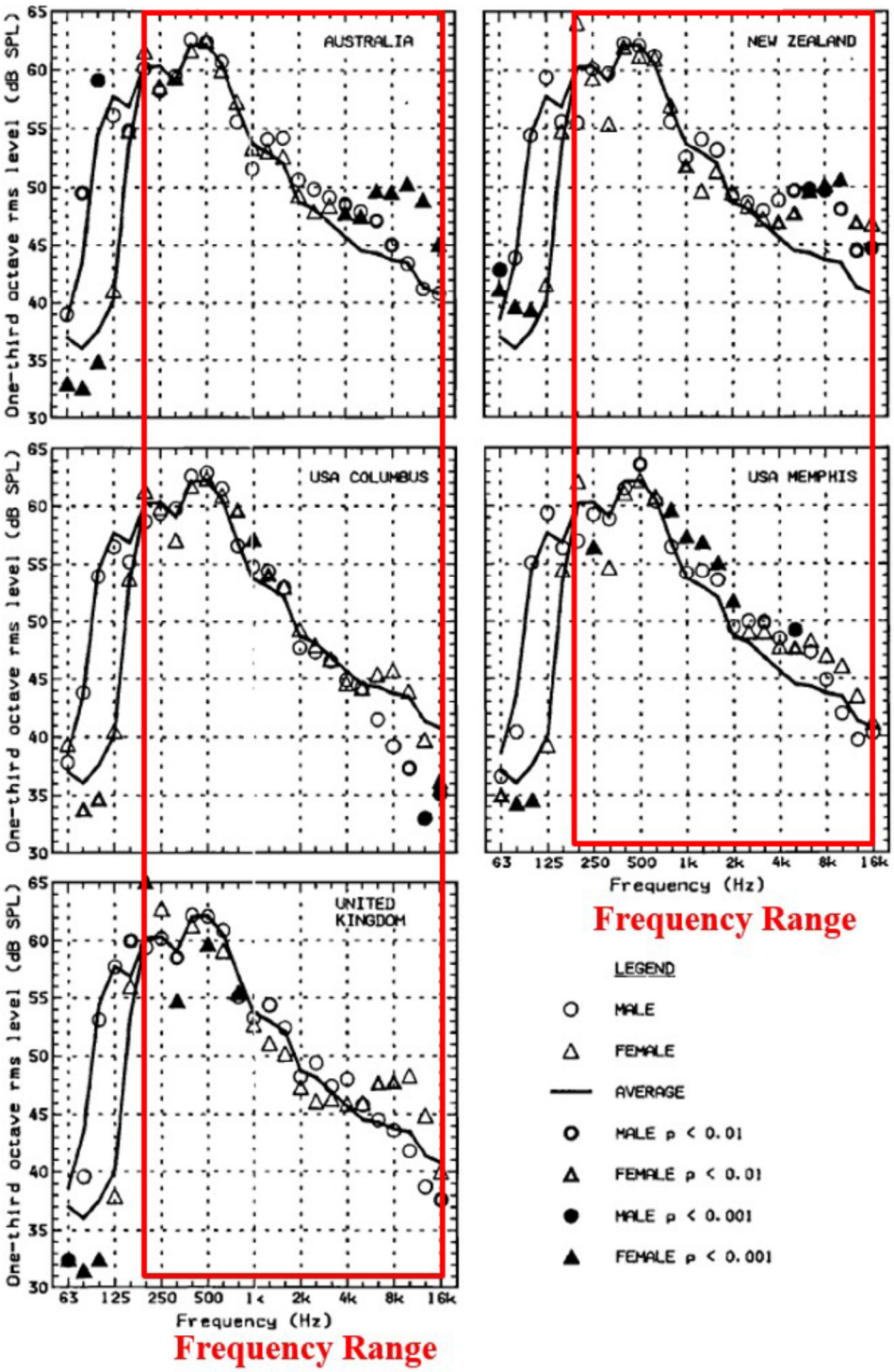


FIG. 1. Male and female long-term average speech spectrum (LTASS) values for five samples of English. Solid line shows LTASS average across 17 speech samples (all samples except Arabic), males and females separately for frequencies below 160 Hz, combined for higher frequencies.

Ex. 1009, FIG. 1

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.,* less than approximately 250 Hz). Ex. 1009, 2985-2987, FIGS. 1, 3; Ex. 1003 ¶327.

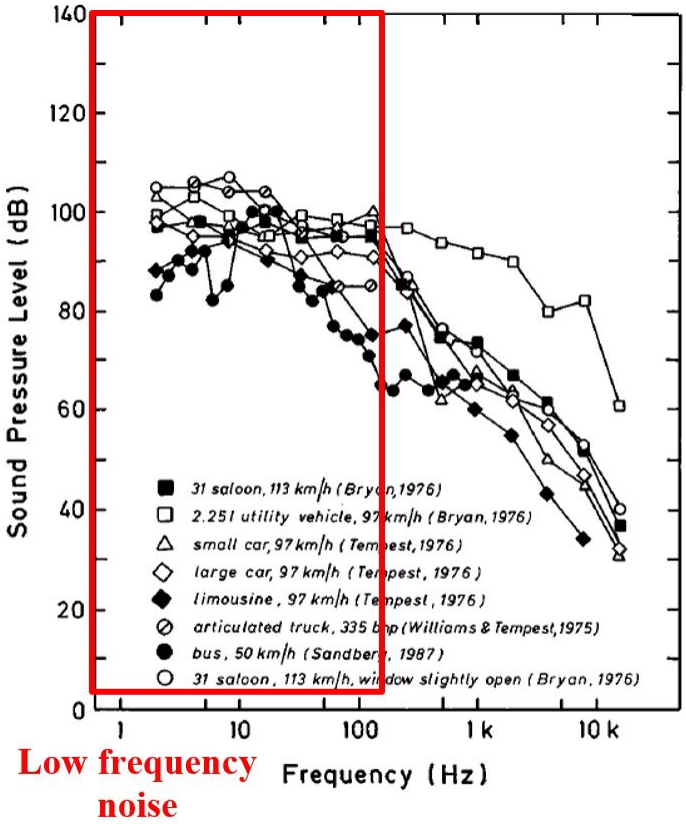


FIG. 3. Passenger noise exposure in road transportation vehicles as a function of frequency.

Ex. 1013, FIG. 3

Further, in view of Burnett, a POSITA would have found it obvious to perform the processes in frequency range(s) known to have "voiced and unvoiced speech" to improve the separation of "voiced and unvoiced speech from background acoustic noise." Ex. 1012 ¶[0063]; Ex. 1003 ¶¶328-329. As described by Burnett, these

frequency ranges would include frequencies higher than approximately 200 Hz (*e.g.,* 500-4000 Hz to detect "k" in "kick," 1700-4000 Hz to detect "sh" in "she," 300-2500 Hz to detect "ee," and 900-1200 Hz to detect "ah"). Ex. 1012 ¶[0064].

F. Claims 21 and 22

[21]: wherein the frequency subband includes frequencies in a range from approximately 250 Hz to 1250 Hz.

[22]: wherein the frequency subband includes frequencies in a range from approximately 200 Hz to 3000 Hz.

When evaluating whether a prior range anticipates a species, courts evaluate whether the claimed species is "critical." *ClearValue, Inc. v. Pearl River Polymers, Inc.*, 668 F.3d 1340, 1344-45 (Fed. Cir. 2012). "'[W]here 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." *E.I. DuPont de Nemours & Co. v. Synvina C.V.*, 904 F.3d 996, 1006 (Fed. Cir. 2018) (internal citations omitted).

There is no evidence in the '611 patent that either of (i) "a range from approximately 250 Hz to 1250 Hz" or (ii) "a range from approximately 200 Hz to 3000 Hz" is critical. For example, in the specification, the **only** explanation regarding these frequency ranges 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, <u>but many others are possible and useful</u>.

Ex. 1001, 6:42-46.

As the '611 patent concedes, the frequency ranges 250-1250 Hz and 200-3000 Hz are but two of "many" frequency ranges that are "possible" and "useful." *Id.* The '611 patent does not explain how using these frequency ranges would produce any new and unexpected results relative to the prior art. Ex. 1003 ¶¶332-334, 347-349. In fact, the '611 patent does not explain how using these frequency ranges would be more advantageous than using any other specific frequency ranges or even "the entire frequency band of interest." *Id.* Further, the '611 patent does not provide any guidance regarding the selection of "possible" and "useful" frequency ranges. *Id.* Also, nothing in the prosecution history of the '611 patent asserts any new and unexpected results relative to the prior art that are associated with these frequency ranges. *Id.*

Further still, there is no evidence in the disclosure or the prosecution history of the '611 patent that the prior art taught away from using either of these frequency ranges, or that any other pertinent secondary considerations are associated with either of these frequency ranges. *Id.*

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.,* Ex. 1001, 17:18-36)

demonstrates the criticality of either of the claimed ranges, Petitioner respectfully disagrees, as the description does not correspond to either of the claimed ranges.

For example, the '611 patent uses bandpass filters to capture signals in certain frequency ranges to detect unvoiced speech (*e.g.*, 500-4000 Hz to detect "k" in "kick," and 1700-4000 Hz to detect "sh" in "she"), and in other frequency ranges to detect voiced speech (e.g., 300-2500 Hz to detect "ee," and 900-1200 Hz to detect "ah"). *Id.* 17:29-36.

However, none of these frequency ranges corresponds to the claimed frequency range of 250-1250 Hz. Instead, the claimed frequency range would arbitrarily include the entirety of the frequency range for some types of speech (*e.g.,* 900-1200 Hz ("ah")), exclude a portion of the frequency range for other types of speech (*e.g.*, 300-2500 Hz ("ee") and 500-4000 Hz ("k" in "kick")), and exclude the entirety of the frequency range for yet other types of speech (*e.g.*, 1700-4000 Hz ("sh" in "she")).

Likewise, none of these frequency ranges corresponds to the claimed frequency range of 200-3000 Hz. Instead, the claimed frequency range would arbitrarily include the entirety of the frequency range for some types of speech (*e.g.,* 900-1200 Hz ("ah") and 300-2500 Hz ("ee")), and exclude a portion of the frequency range for other types of speech (*e.g.*, 500-4000 Hz ("k" in "kick") and 1700-4000 Hz ("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. Ex. 1003 ¶¶330-359.

However, even if this description in the ’611 patent were to demonstrate the criticality of either of the claimed ranges, Burnett nevertheless includes this same description verbatim. Ex. 1012 ¶¶[0063]-[0064]. As such, the claimed ranges would still be disclosed by Burnett.

Further, Avendano in combination with Byrne, Burnett, and/or Berglund (and in the further combination with Visser, Bisgaard, and Hou) also discloses the claimed ranges. Ex. 1003 ¶¶335-344, 350-359. By virtue of the inclusive term “including,” claims 21 and 22 do not preclude a frequency subband from including frequencies in addition to those between approximately 250-1250 Hz and 200-3000 Hz, respectively. 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. Sections VI.D-E. Such a frequency subband would include frequencies in a range from approximately 250-1250 Hz and 200-3000 Hz, among others. Ex. 1003 ¶¶339, 354.

Nevertheless, to the extent that claims 21 and 22 preclude a frequency subband from including frequencies other than those between approximately

250-1250 Hz and 200-3000 Hz, respectively, as described above, a POSITA also would have found it obvious to perform the processes described by Avendano, Visser, and Bisgaard in these frequency subbands. Ex. 1003 ¶¶340-344, 355-359.

Frequencies in a Range from Approximately 250 Hz to 1250 Hz:

A POSITA would have found it obvious to perform the processes in frequency range(s) known to have significant spectral components of speech (*e.g.,* to more readily detect the presence of human speech). Section VI.D; Ex. 1003 ¶336.

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). Ex. 1009, FIG. 1; Ex. 1003 ¶¶337-338.

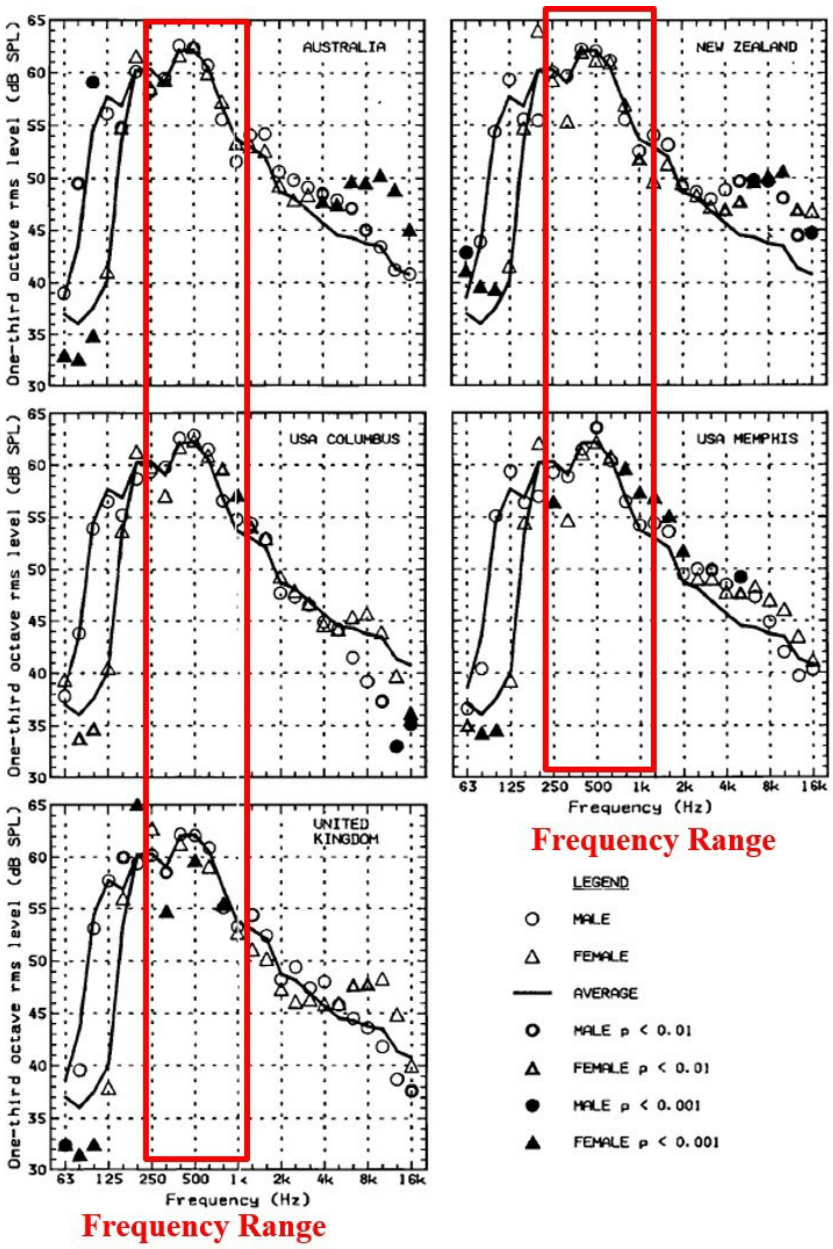


FIG. 1. Male and female long-term average speech spectrum (LTASS) values for five samples of English. Solid line shows LTASS average across 17 speech samples (all samples except Arabic), males and females separately for frequencies below 160 Hz, combined for higher frequencies.

Ex. 1009, FIG. 1

As taught by Burnett, certain types of speech can be detected in a frequency range of 250-1250 Hz (*e.g.*, "ah" in frequencies 900-1200 Hz. Ex. 1012 ¶ [0064].

Further, a POSITA would have found it 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. Section VI.D; Ex. 1013, 2985-2987, FIGS. 1, 3; Ex. 1003 ¶341.

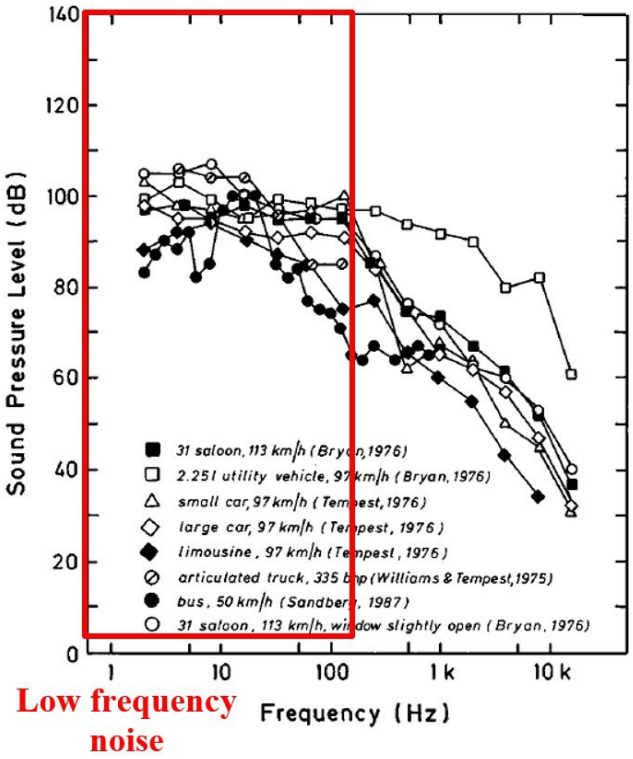


FIG. 3. Passenger noise exposure in road transportation vehicles as a function of frequency.

Ex. 1013, FIG. 3

Further, a POSITA would have found it obvious to tune the frequency range, depending on the specific use case at hand. Section VI.D. For example, 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 (*e.g.*, to improve the speed, accuracy, and/or efficiency in detecting voice activity). Ex. 1003 ¶342.

Additionally, a POSITA would have found it obvious to use other frequency ranges, depending on the type of speech that is being detected, the type of noise that is being filtered out, and the expected spectral distributions thereof. *Id.*

Frequencies in a Range from Approximately 200 Hz to 3000 Hz:

A POSITA would have found it obvious to perform the processes in frequency range(s) known to have significant spectral components of speech (*e.g.,* to more readily detect the presence of human speech). Section VI.D; Ex. 1003 ¶351.

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). Ex. 1009, FIG. 1; Ex. 1003 ¶352.

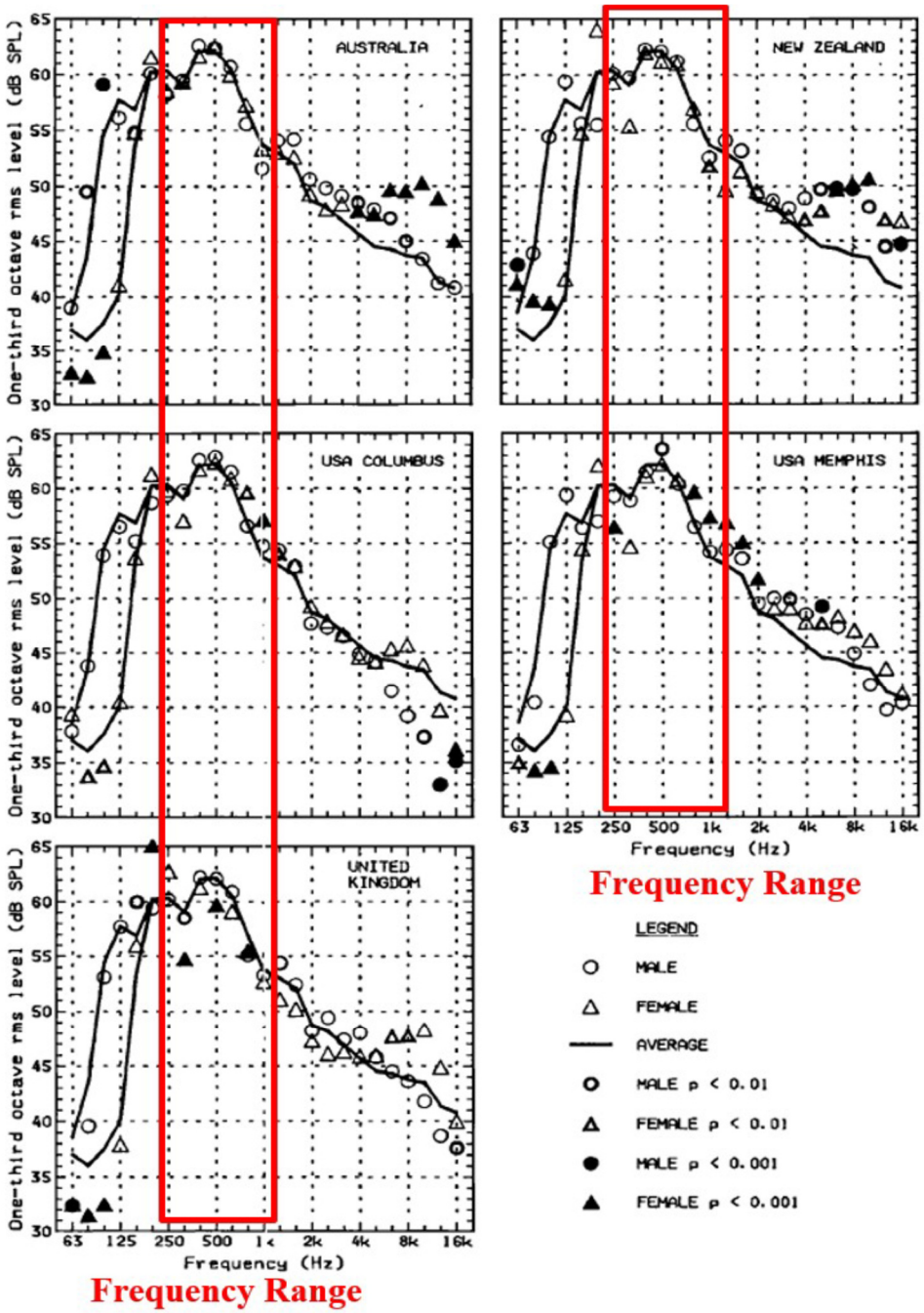


FIG. 1. Male and female long-term average speech spectrum (LTASS) values for five samples of English. Solid line shows LTASS average across 17 speech samples (all samples except Arabic), males and females separately for frequencies below 160 Hz, combined for higher frequencies.

Ex. 1009, FIG. 1

As taught by Burnett, certain types of speech can be detected in a frequency range of 200-3000 Hz (*e.g.,* "ee" in frequencies 300-2500 Hz). Ex. 1012 ¶[0064].

Further, a POSITA would have found it 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. Section VI.D; Ex. 1013, 2985-2987, FIGS. 1, 3; Ex. 1003 ¶356.

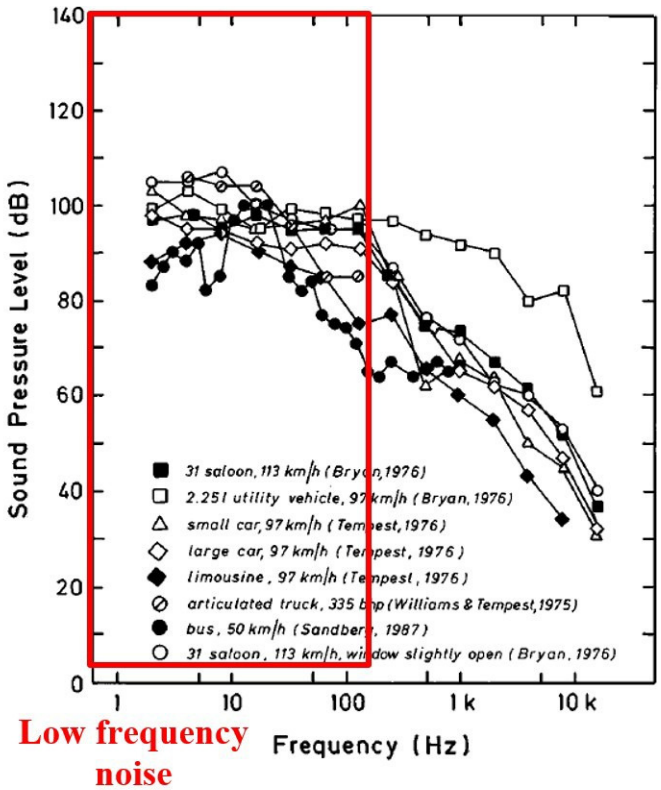


FIG. 3. Passenger noise exposure in road transportation vehicles as a function of frequency.

Ex. 1013, FIG. 3

Further, a POSITA would have found it obvious to tune the frequency range, depending on the specific use case at hand. Section VI.D. For example, 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 (*e.g.*, to improve the speed, accuracy, and/or efficiency in detecting voice activity). Ex. 1003 ¶357. Additionally, a POSITA also would have found it obvious to use other frequency ranges, depending on the type of speech that is being detected, the type of noise that is being filtered out, and the expected spectral distributions thereof. *Id.*

VII. INSTITUTION IS APPROPRIATE HERE

A. Co-Pending Litigation Does Not Warrant Discretionary Denial Under § 314(A) (*Fintiv*)

Efficiency, fairness, and the merits support institution. *Apple Inc. v. Fintiv, Inc.*, IPR2020-00019, Paper 11 (PTAB Mar. 20, 2020) ("*Fintiv*").

1. Factor 1: Potential Stay

On February 28, 2023, Patent Owner ("PO") sued Petitioner for infringement of the '611 patent in *Jawbone Innovations, LLC v. Sony Electronics Inc.*, No. 2:23-cv-01161-CCC-LDW (D.N.J.) ("the Litigation"). There are currently 16 instituted IPRs covering every patent claim asserted against Petitioner in the February 28, 2023 complaint, including the claims of the '611 Patent, as follows:

- US10,779,080 (IPR2022-01059, IPR2023-00284);

- US8,019,091 (IPR2022-00649, IPR2023-00253);
- US7,246,058 (IPR2022-00623; IPR2023-00222);
- US8,321,213 (IPR2022-00797, IPR2023-00276, IPR 2023-00279);
- US8,467,543 (IPR2022-01027);
- US8,503,691 (IPR2022-01060, IPR2022-01061);
- US11,122,357 (IPR2022-01124, IPR2023-00251); and
- US8,326,611 (IPR2023-00285 (claims 29-44 of the '611 Patent), IPR2023-286 (claims 1-28 of the '611 Patent)).

On June 12, 2023, Jawbone amended its complaint to add one additional patent.

The District of New Jersey ("DNJ") has a "liberal" policy toward staying cases when IPRs have been instituted, to avoid wasting "a court's time and resources" and because "a USPTO decision could drastically alter the nature of the litigation." *Kirsch Rsch. & Dev., LLC v. GAF Materials, LLC*, Civil Action No. 20-13683-JMV, 2021 U.S. Dist. LEXIS 111667, at *5 (D.N.J. June 15, 2021). This is particularly true when, as here, the litigation is at its early stages. *Id.* at *13. Given that the Litigation is at its inception and all asserted claims (other than the claims in a single patent added months after the original complaint) are being challenged in instituted IPRs, the likelihood of the grant of a stay weighs against discretionary denial.

2. Factor 2: Proximity of Trial to FWD

The DNJ has not set a trial date. Thus, this factor weighs against discretionary denial. *Global Tel-Link Corp. v. HLFIP Holding, Inc.*, IPR2021-00444, Paper 14, at 16-19 (PTAB July 22, 2021) ("The lack of a scheduled trial date … typically weighs against exercising discretion to deny the Petition."); *GAF Materials LLC v. Kirsch Rsch. & Dev., LLC*, IPR2021-00192, Paper 14, at 12-13 (PTAB May 25, 2021); *Skechers U.S.A., Inc. v. Nike, Inc.*, IPR2021-00160, Paper 10, at 9-11 (PTAB May 19, 2021).

3. Factor 3: Investment in Parallel Proceeding

PO filed its complaint on February 28, 2023, and served it on March 16, 2023. The Litigation, ECF Nos. 1, 5. The parties have not conducted any discovery, and an initial scheduling conference will not occur until July 18, 2023. *Id.* ECF No. 18. Thus, the parties have invested very little in the Litigation, and little is expected to be invested in the Litigation over the next several months (if the case is not stayed). Even if a scheduling order subsequently imposes some discovery deadlines over the following 6 months (*e.g.*, fact discovery, claim construction), much work will still remain, including expert reports, expert discovery, dispositive motions, pretrial motions, and trial. This "remaining investment" significantly outweighs any investment made thus far. *Samsung Elecs. Am. Inc. v. Snik LLC*, IPR2020-01428, Paper 10, at 11 (PTAB

Mar. 9, 2021); *Huawei Techs. Co. v. WSOU Invs., LLC*, IPR2021-00226, Paper 10, at 13-14 (PTAB June 10, 2021).

4. Factor 4: Overlapping Issues

If this IPR is instituted and the Litigation is stayed, Petitioner could not pursue in the Litigation any invalidity ground raised or that could have been reasonably raised in this IPR. *Cal. Inst. of Tech. v. Broadcom Ltd.*, 25 F.4th 976, 991 (Fed. Cir. 2022). Thus, duplicative efforts or conflicting decisions are unlikely.

Petitioner hereby stipulates that, if this IPR is instituted but the Litigation is not stayed, Petitioner will not argue in the Litigation invalidity of the '611 patent based on the grounds in this petition. This factor weighs against discretionary denial. *Sand Revolution II, LLC v. Cont'l Intermodal Grp.-Trucking LLC*, IPR2019-01393, Paper 24, at 11-12 (PTAB June 16, 2020) (informative).

5. Factor 5: Parties in Parallel Proceedings

The parties are the same, but trial in the Litigation is highly unlikely to start before this IPR reaches a final written decision, and most likely will occur long after a final written decision. Thus, this factor is neutral. *See Google LLC v. Jawbone Innovations LLC*, IPR2022-00630, Paper 10, at 17 (PTAB Sept. 13, 2022).

6. Factor 6: Other Circumstances

To Petitioner's knowledge, other than Amazon's co-pending IPR petition on the same grounds that Petitioner is seeking to join in its accompanying motion for

joinder (IPR2023-00286) and a now-terminated petition by Apple advancing the same grounds (IPR2022-01085), the '611 patent has never been compared to this combination of references by the PTO, any court, or any jury. The petition's merits are particularly strong, which favors institution. *Fintiv*, IPR2020-00019, Paper 11, at 18.

Accordingly, the *Fintiv* analysis favors institution.

B. The Existence of Previously Filed IPR Petitions Does Not Warrant Discretionary Denial (*General Plastic*)

Google filed IPR petitions challenging the '611 Patent based on different prior art on February 22, 022 (IPR2022-00604, institution denied) and May 16, 2022 (IPR2022-00889, institution denied). Apple also filed two petitions, both of which were terminated following settlement (IPR2022-1085, IPR2022-1495), and Amazon has filed two petitions in IPRs that Sony now seeks to join (IPR2023-00285, IPR 2023-00286). Those petitions do not warrant denial here under *General Plastic Indus. Co. v. Canon Kabushiki Kaisha*, IPR2016-01357, Paper 19 at 15-16 (PTAB Sept. 6, 2017) (precedential). The Board has found that the *General Plastic* factors are "not particularly relevant" in this situation, i.e., where a different petitioner files a "me-too" or "copycat" petition with a timely motion for joinder. *Central Security Group*, IPR2019-01609, Paper 11, at 8; *Celltrion, Inc. v. Genentech, Inc.*, IPR2018-01019, Paper 11, at 9-11 (PTAB Oct. 30, 2018).

Even if the Board were to consider the *General Plastic* factors, discretionary denial is unwarranted. Sony and the prior petitioners are not the same party and have no significant relationship. They are not co-defendants. They are competitors accused of infringement—in different actions pending in different courts—based on sales of different products. This weighs against denial. *NetNut Ltd. v. Bright Data Ltd.*, IPR2021-00465, Paper 11, at 9 (PTAB Aug. 12, 2021) (declining to extend *General Plastic* and *Valve* to different petitioner with no relationship to previous petitioners); *Mercedes-Benz USA, LLC v. Carucel Invs. L.P.*, IPR2019-01404, Paper 12, at 11-12 (PTAB Jan. 22, 2020).

The remaining factors consider timing issues, *e.g.*, when Petitioner became aware of prior art, whether the PO or Board responded to prior petitions, and the length of time that has passed since an earlier petition. But the sequential IPR filings here were caused by PO's sequential district court filings and therefore should not be held against Petitioner. Specifically, PO sued Google, Apple and Amazon in 2021 (6:21-cv-00985-ADA, 6:21-cv-00984-ADA (W.D. Tex.), 2:21-cv-00435 (E.D. Tex.)), but waited nearly 18 months before suing Petitioner in February 2023. The Board should not deny Petitioner's IPR petition merely because PO sued other defendants earlier. *Toshiba*, IPR2018-01538, Paper 11, at 21-22 (PO's complaint about multiple IPRs "is not persuasive when the volume appears to be a direct result of its own litigation activity"); *id.* at 22 (it was not unfair for a petitioner to file close to one-year

bar date even when co-defendants in related litigation "with distinct interests" and "distinct litigation strategies" filed earlier).

VIII. DISCRETIONARY DENIAL UNDER § 325(d) IS ALSO NOT APPROPRIATE

The Office has not previously considered "the same or substantially the same prior art or arguments." 35 U.S.C. § 325(d). Here, all grounds rely on Avendano, which the PTO never considered during original examination of the '611 patent.

Further, though this Petition presents the same grounds as in Amazon's petition (IPR2023-00286), Petitioner is filing this petition to preserve its ability to maintain a petition on the merits in the event that Amazon terminates its involvement. If Petitioner is joined to the Amazon IPR, Petitioner will take an understudy role. Thus, the proposed grounds will not be cumulative of references previously considered by the Board.

IX. MANDATORY NOTICES

A. Real Parties-In-Interest (37 C.F.R. § 42.8(b)(1))

Sony Electronics Inc., Sony Group Corporation, Sony Corporation, and Sony Corporation of America are the real parties in interest.

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

To the best knowledge of Petitioner, the '611 Patent is or has been involved in the following district court litigations and petitions for *inter partes* review:

Name	Number	Court	Filed
Jawbone Innovations, LLC v. Amazon.com, Inc.	5:22-cv-06727	N.D. Cal.	Nov. 29, 2021 (transferred from E.D. Tex. on Nov. 1, 2022)
Jawbone Innovations, LLC v. Amazon.com, Inc.	2:21-cv-00435	E.D. Tex.	Nov. 29, 2021
Jawbone Innovations, LLC v. Google LLC	3:23-cv-00466	N.D. Cal.	Sept. 23, 2021 (transferred from W.D. Tex. on Feb. 1, 2023)
Jawbone Innovations, LLC v. Google LLC	6:21-cv-00985	W.D. Tex.	Sept. 23, 2021
Petition for Inter Partes Review (Amazon)	IPR2023-00285	PTAB	Nov. 28, 2022
Petition for Inter Partes Review (Amazon)	IPR2023-00286	PTAB	Nov. 28, 2022
Jawbone Innovations, LLC v. Meta Platforms, Inc.	6:23-cv-00158	W.D. Tex.	Feb. 28, 2023
Jawbone Innovations, LLC v. ZTE Corporation	2:23-cv-00082	E.D. Tex.	Feb. 28, 2023
Jawbone Innovations, LLC v. Panasonic Holdings	2:23-cv-00081	E.D. Tex.	Feb. 28, 2023
Jawbone Innovations, LLC v. Guangdong Oppo Mobile Telecommunications Corp., Ltd.	2:23-cv-00079	E.D. Tex.	Feb. 28, 2023
Jawbone Innovations, LLC v. LG Electronics, Inc.	2:23-cv-00078	E.D. Tex.	Feb. 28, 2023
Jawbone Innovations, LLC v. HTC Corporation	2:23-cv-00077	E.D. Tex.	Feb. 28, 2023
Jawbone Innovations, LLC v. Sony Electronics Inc.	2:21-cv-00186	D.N.J.	Feb. 28, 2023

C. Lead and Backup Counsel (37 C.F.R. § 42.8(b)(3))

Amazon provides the following designation of counsel:

LEAD COUNSEL	BACKUP COUNSEL
Gregory S. Gewirtz Registration No. 36,522 LERNER DAVID LLP 20 Commerce Drive Cranford, NJ 07016 Telephone: 908.645.5000 Facsimile: 908.654.7866 E-mail: ggewirtz@lernerdavid.com litigation@lernerdavid.com	Russell W. Faegenburg Registration No. 77,876 LERNER DAVID LLP 20 Commerce Drive Cranford, NJ 07016 Telephone: 908.645.5000 Facsimile: 908.654.7866 E-mail: rfaegenburg@lernerdavid.com litigation@lernerdavid.com

D. Service Information (37 C.F.R. § 42.8(b)(4))

Service information for lead and backup counsel is provided in the designation of lead and backup counsel above. Petitioner hereby consents to service by email at the following email address:

litigation@lernerdavid.com

E. Payment of Fees (37 C.F.R. § 42.103)

The fee set forth in 37 C.F.R. § 42.15(a) for this petition has been paid. Any additional fees due in connection with this petition may be charged to Deposit Account 121095.

X. CONCLUSION

Petitioner respectfully requests cancellation of the challenged claims.

Respectfully submitted,

LERNER DAVID LLP

July 6, 2023 *<u>/Gregory S. Gewirtz/</u>*
Gregory S. Gewirtz (Reg. No. Reg. No. 36,522)
Counsel for Petitioner Sony Electronics Inc.

CERTIFICATE OF COMPLIANCE

Pursuant to 37 C.F.R. § 42.24(d), the undersigned certifies that foregoing **PETITION FOR *INTER PARTES* REVIEW OF CLAIMS 1-28 OF U.S. PATENT NO. 8,326,611**, exclusive of the parts exempted as provided in 37 C.F.R. § 42.24(a), contains 13,381 words and therefore complies with the type-volume limitations of 37 C.F.R. § 42.24(a).

Respectfully submitted,

LERNER DAVID LLP

July 6, 2023

/Gregory S. Gewirtz/
Gregory S. Gewirtz (Reg. No. 36,522)
Counsel for Petitioner Sony Electronics Inc.

CERTIFICATE OF SERVICE

I hereby certify that a true and correct copy of the foregoing **PETITION FOR *INTER PARTES* REVIEW OF CLAIMS 1-28 OF U.S. PATENT NO. 8,326,611** and **EXHIBITS 1001-1016** are being served on July 6, 2023, via Federal Express overnight mail on counsel of record for U.S. Patent No. 8,326,611 as addressed below:

Mark Leonardo
Nutter McClennen & Fish LLP
Seaport West
155 Seaport Blvd.
Boston, MA 02210-2604

Peter Lambrianakos
FABRICANT LLP
411 Theodore Fremd Avenue
Suite 206 South
Rye, NY 10580

Respectfully submitted,

LERNER DAVID LLP

July 6, 2023

/Gregory S. Gewirtz/
Gregory S. Gewirtz (Reg. No. 36,522)
Counsel for Petitioner Sony Electronics Inc.