

UNITED STATES PATENT AND TRADEMARK OFFICE

In re *Inter Partes* Review of:)
U.S. Patent No. 10,779,080)
Issued: September 15, 2020)
Application No.: 13/948,160)
Filing Date: July 22, 2013)

For: Dual Omnidirectional Microphone Array

FILED VIA P-TACTS

**DECLARATION OF DR. MICHAEL JOHNSON IN SUPPORT OF
PETITION FOR INTER PARTES REVIEW OF
U.S. PATENT NO. 10,779,080**

Declaration of Dr. Michael Johnson in Support of
Petition for *Inter Partes* Review of USP No. 10,779,080

TABLE OF CONTENTS

I.	INTRODUCTION -----	1
II.	EXPERIENCE AND QUALIFICATIONS -----	2
III.	UNDERSTANDING OF RELEVANT LEGAL PRINCIPLES-----	5
A.	Claim Construction -----	5
B.	Obviousness -----	6
IV.	MATERIALS REVIEWED AND CONSIDERED -----	8
V.	LEVEL OF SKILL IN THE ART AND PERSPECTIVE APPLIED IN THIS DECLARATION -----	10
VI.	BACKGROUND -----	12
A.	In 1982, Griffiths and Jim Publish Their Seminal Article on the GSC. -----	12
B.	The GSC Was Applied to Microphone Arrays to Reduce Noise-----	15
C.	A 2001 Book Edited by Brandstein Illustrates How to Apply a GSC to Microphone Arrays to Reduce Noise. -----	16
D.	Adaptive Noise Removal Methods Were Well Known -----	23
E.	Near-field Compensation Techniques Were Well Known.-----	26
F.	GSCs Incorporated Voice Activity Detection -----	28
VII.	THE '080 PATENT -----	29
A.	Specification -----	29
B.	Prosecution History -----	33

Declaration of Dr. Michael Johnson in Support of
Petition for *Inter Partes* Review of USP No. 10,779,080

C.	The Earliest Priority Date of the '080 Patent Claims Is June 13, 2007.-----	34
D.	The References Are Prior Art.-----	34
E.	The Asserted References Are Analogous Art. -----	35
VIII.	GROUND OF UNPATENTABILITY -----	35
A.	Ground 1: Brandstein in View of Ikeda -----	35
1.	Motivation to Combine -----	36
a)	Motivation to Combine Teachings of Brandstein's Embodiments -----	36
b)	Motivation to Combine Brandstein and Ikeda -----	39
2.	Reasonable Expectation of Success -----	44
a)	Reasonable Expectation of Success in Combining Teachings of Brandstein's Embodiments-----	44
b)	Reasonable Expectation of Success in Combining Brandstein and Ikeda-----	46
3.	Claim 7 -----	48
a)	[7pre] "A system, comprising:"-----	48
b)	[7a] "a first virtual microphone formed from a first combination of a first microphone signal and a second microphone signal, wherein the first microphone signal is generated by a first physical microphone and the second microphone signal is generated by a second physical microphone;" -----	49

Declaration of Dr. Michael Johnson in Support of
Petition for *Inter Partes* Review of USP No. 10,779,080

- c) [7b] “a second virtual microphone formed from a second combination of the first microphone signal and the second microphone signal, wherein the second combination is different from the first combination,” ----- 50
- d) [7c] “wherein the first virtual microphone has a first linear response to speech and first linear response to noise,” ----- 53
- e) [7d] “the first linear response to speech being substantially similar across a plurality of frequencies for a speech source located within a predetermined angle relative to an axis of the microphone array and devoid of a null,” ----- 56
- f) [7e] “wherein the second virtual microphone has a second linear response to speech that has a single null oriented in a direction toward a source of the speech and a second linear response to noise,” ----- 60
- g) [7f] “wherein the second linear response to noise is substantially similar to the first linear response to noise,” and [7h] “the second linear response to speech is substantially dissimilar to the first linear response to speech,” ----- 65
- h) [7g] “one or both of the first linear response to noise and the second linear response to noise being nonzero in a direction toward a source of noise, and” ----- 69
- i) [7i] “wherein the speech is human speech; and” ----- 69
- j) [7j] “an adaptive noise removal application coupled to the first and second virtual

Declaration of Dr. Michael Johnson in Support of
Petition for *Inter Partes* Review of USP No. 10,779,080

	microphones and generating denoised output signals”-----	70
k)	[7k] “by forming a plurality of combinations of signals output from the first virtual microphone and the second virtual microphone,”-----	75
l)	[7l] “by filtering and summing the plurality of combinations of signals in the time domain, and”-----	75
m)	[7m] “by a varying linear transfer function between the plurality of combinations of signals,”-----	77
n)	[7n] “wherein the denoised output signals include less acoustic noise than acoustic signals received at the first and second physical microphones.”-----	78
4.	Claim 8: “The system of claim 7 and further comprising: a microphone array, the first and second physical microphones positioned in [sic] the microphone array.”-----	80
5.	Claim 9: “The system of claim 7, wherein the single null is a region of the second linear response to speech having a measured response level that is lower than the measured response level of any other region of the second linear response to speech.”-----	80
6.	Claim 11: “The system of claim 7 and further comprising: a communications channel coupled with the processing component and including one or more of a wireless channel, a wired channel, and a hybrid wireless/wired channel.”-----	82

Declaration of Dr. Michael Johnson in Support of
Petition for *Inter Partes* Review of USP No. 10,779,080

7. Claim 12: “The system of claim 11 and further comprising: a communication device wirelessly coupled with the wireless channel of the communications channel.” ----- 84
8. Claim 13: “The system of claim 7, wherein the second microphone signal is multiplied by a ratio, wherein the ratio is a ratio of a third distance to a fourth distance, the third distance being between the first physical microphone and the speech source and the fourth distance being between the second physical microphone and the speech source.” ----- 84
9. Claim 14 ----- 90
 - a) [14pre] “A system, comprising:” ----- 90
 - b) [14a] “a first virtual microphone comprising a first combination of a first microphone signal and a second microphone signal,” ----- 90
 - c) [14b] “the first virtual microphone having a first linear response to speech and a first linear response to noise,” ----- 90
 - d) [14c] “the first linear response to speech being substantially similar across a plurality of frequencies for a speech source located within a predetermined angle relative to an axis of a microphone array,” ----- 90
 - e) [14d] “wherein the first microphone signal is output from a first physical microphone and the second microphone signal is output from a second physical microphone;” ----- 90

Declaration of Dr. Michael Johnson in Support of
Petition for *Inter Partes* Review of USP No. 10,779,080

- f) [14e] “a second virtual microphone comprising a second combination of the first microphone signal and the second microphone signal,”----- 91
- g) [14f] “the second virtual microphone having a second linear response to speech and a second linear response to noise,” ----- 91
- h) [14g] “the second linear response to noise being substantially similar to the first linear response to noise,”----- 91
- i) [14h] “one or both of the first linear response to noise and the second linear response to noise being non- zero in a direction toward a source of noise, and”----- 91
- j) [14i] “the second linear response to speech being substantially dissimilar to the first linear response to speech,” ----- 92
- k) [14j] “wherein the second combination is different from the first combination,”----- 92
- l) [14k] “wherein the first virtual microphone and the second virtual microphone are distinct virtual directional microphones; and”----- 92
- m) [14l] “a processing component coupled to the first and second virtual microphones, the processing component including an adaptive noise removal application receiving acoustic signals from the first virtual microphone and the second virtual microphone, filtering and summing the acoustic signals in the time domain, applying a varying linear transfer function between the acoustic signals, and generating an output signal,”----- 93

Declaration of Dr. Michael Johnson in Support of
Petition for *Inter Partes* Review of USP No. 10,779,080

n)	[14m]“wherein the output signal is a denoised acoustic signal.”-----	94
10.	Claim 16: “The system of claim 14 and further comprising: a communications channel coupled with the processing component and including one or more of a wireless channel, a wired channel, and a hybrid wireless/wired channel.” -----	95
11.	Claim 17: “The system of claim 16 and further comprising: a communication device wirelessly coupled with the wireless channel of the communications channel.” -----	95
12.	Claim 18: “The system of claim 14, wherein the acoustic signals from the first virtual microphone, the second virtual microphone or both are delayed.” -----	95
13.	Claim 19: “The system of claim 18, wherein the delay is raised to a power that is proportional to a time difference between arrival of the speech at the first virtual microphone and arrival of the speech at the second virtual microphone.”-----	97
14.	Claim 20: “The system of claim 19, wherein the power is proportional to a sampling frequency multiplied by a quantity equal to a third distance subtracted from a fourth distance, the third distance being between a first physical microphone and the speechsource, the fourth distance being between a second physical microphone and the speech source, and the first and second physical microphones are positioned in the microphone array.” -----	99
15.	Claim 1 -----	101
a)	[1pre]“A system, comprising:” -----	101

Declaration of Dr. Michael Johnson in Support of
Petition for *Inter Partes* Review of USP No. 10,779,080

- b) [1a] “a microphone array including a first physical microphone outputting a first microphone signal and a second physical microphone outputting a second microphone signal;” ----- 101
- c) [1b] “a processing component coupled to the microphone array and generating a virtual microphone array including a first virtual microphone and a second virtual microphone,” ----- 101
- d) [1c] “the first virtual microphone including a first combination of the first microphone signal and the second microphone signal, the second virtual microphone including a second combination of the first microphone signal and the second microphone signal wherein the second combination is different from the first combination,” ----- 102
- e) [1d] “wherein the first virtual microphone and the second virtual microphone have substantially similar responses to noise and substantially dissimilar responses to speech; and” ----- 102
- f) [1e] “an adaptive noise removal application coupled to the processing component and generating denoised output signals by forming a plurality of combinations of signals output from the first virtual microphone and the second virtual microphone, by filtering and summing the plurality of combinations of signals in the time domain, and by a varying linear transfer function between the plurality of combinations of signals,” ----- 103

Declaration of Dr. Michael Johnson in Support of
Petition for *Inter Partes* Review of USP No. 10,779,080

g)	[1f] “wherein the denoised output signals include less acoustic noise than acoustic signals received at the microphone array.” -----	103
16.	Claim 2: “The system of claim 1, wherein the acoustic noise comprises noise content and the acoustic signals comprise speech content.” -----	104
17.	Claim 3: “The system of claim 2, wherein the speech content comprises human speech.” -----	104
18.	Claim 5: “The system of claim 1 and further comprising: a communications channel coupled with the processing component and including one or more of a wireless channel, a wired channel, and a hybrid wireless/wired channel.” -----	104
19.	Claim 6: “The system of claim 5 and further comprising: a communication device wirelessly coupled with the wireless channel of the communications channel.” -----	104
B.	Ground 2: Brandstein in combination with Ikeda and McCowan -----	105
1.	Motivation to Combine -----	105
1.	Expectation of success -----	108
2.	Claim 13: “The system of claim 7, wherein the second microphone signal is multiplied by a ratio, wherein the ratio is a ratio of a third distance to a fourth distance, the third distance being between the first physical microphone and the speech source and the fourth distance being between the second physical microphone and the speech source.” -----	108

Declaration of Dr. Michael Johnson in Support of
Petition for *Inter Partes* Review of USP No. 10,779,080

3.	Claim 19: “The system of claim 18, wherein the delay is raised to a power that is proportional to a time difference between arrival of the speech at the first virtual microphone and arrival of the speech at the second virtual microphone.”-----	109
4.	Claim 20: “The system of claim 19, wherein the power is proportional to a sampling frequency multiplied by a quantity equal to a third distance subtracted from a fourth distance, the third distance being between a first physical microphone and the speechsource, the fourth distance being between a second physical microphone and the speech source, and the first and second physical microphones are positioned in the microphone array.” -----	110
C.	Ground 3: Brandstein in combination with Ikeda and Yang -----	110
1.	Motivation to Combine -----	111
2.	Reasonable expectation of Success -----	111
3.	Claim 10: “The system of claim 7 and further comprising: a voice activity detector (VAD) coupled with the processing component and operative to generate voice activity signals.”-----	113
4.	Claim 15: “The system of claim 14 and further comprising: a voice activity detector (VAD) coupled with the processing component and operative to generate voice activity signals.”-----	114
5.	Claim 4: “The system of claim 1 and further comprising: a voice activity detector (VAD) coupled with the processing component and operative to generate voice activity signals.”-----	114
IX.	CONCLUSION -----	114

Declaration of Dr. Michael Johnson in Support of
Petition for *Inter Partes* Review of USP No. 10,779,080

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

I. INTRODUCTION

1. I have been retained by Meta Platforms, Inc. (“Petitioner”) to provide my opinions concerning the validity of claim 1-20 of U.S. Patent No. 10,779,080 (“the ’080 patent”).

2. I have been simultaneously engaged by Petitioner to provide a declaration for IPR petition challenging U.S. Patent No. 8,503,691.

3. This Declaration summarizes the opinions I have formed to date. I reserve the right to modify my opinions, if necessary, based on further review and analysis of information that I receive subsequent to the filing of this report, including in response to positions taken by Jawbone Innovations LLC (“Jawbone”) or its experts that I have not yet seen, including any secondary consideration evidence that Jawbone or its expert may consider and present.

4. It is my opinion that claims 1-20 of the ’080 patent are invalid based on the following grounds:

Ground	Claims Challenged	References
1	1-3, 5-9, 11-14, 16-20	Brandstein, Ikeda
2	13, 19, 20	Brandstein, Ikeda, McCowan
3	4, 10, 15	Brandstein, Ikeda, Yang

II. EXPERIENCE AND QUALIFICATIONS

5. I am currently the Associate Dean for Undergraduate Education and Student Success of the University of Kentucky (UK) Pigman College of Engineering in Lexington, Kentucky, and have served in that position since July 2023. I also currently serve as Department Chair of Electrical and Computer Engineering (ECE) at UK and have been a Full Professor in ECE since 2016. Prior to that, I was a faculty member at Marquette University in Milwaukee, Wisconsin from 2000 to 2016, during which time I also spent two separate sabbatical years as a Visiting Professor at Tsinghua University in Beijing, China.

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

7. The focus of my research work is Speech and Audio Signal Processing, and I regularly conduct and publish original research articles within this field, with

Declaration of Dr. Michael Johnson in Support of
Petition for *Inter Partes* Review of USP No. 10,779,080

a publication record including 50 peer-reviewed journal articles and more than 140 total peer-reviewed publications over the past two decades. My specific areas of expertise include digital signal processing, speech enhancement and noise reduction, automatic speech recognition, speech production and kinematics, bioacoustics, and machine learning. I have received multiple research grant awards for my work in these areas, including from the National Science Foundation and the National Institutes of Health.

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

9. My professional service includes active membership in the Institute for Electrical and Electronic Engineers (IEEE), the IEEE Signal Processing Society, the Acoustical Society of America (ASA), and the Electrical and Computer Engineering Department Heads Association (ECEDHA). I served as a member of the IEEE Signal Processing Society Speech and Language Technical Committee from 2014 to 2017, and regularly serve as a reviewer for journals, conferences, and workshops in the area of speech and signal processing, including IEEE Transactions on Acoustics,

Declaration of Dr. Michael Johnson in Support of
Petition for *Inter Partes* Review of USP No. 10,779,080

Speech, and Signal Processing, IEEE Transactions on Signal Processing, the Journal of the Acoustical Society of America, Speech Communication, the International Conference on Acoustics, Speech, and Signal Processing, the IEEE Workshop on Automatic Speech Recognition, and Interspeech. I have reviewed funding proposals for a number of public and private agencies and foundations, including for multiple programs and directorates at the National Science Foundation. In addition to these manuscript and grant proposal reviewing activities, I also currently serve as Area Editor for the journal Speech Communication in the area of speech enhancement.

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

11. A detailed record of my professional qualifications, including a list of publications, awards, and professional activities, is set forth in my curriculum vitae, which is attached as Exhibit 1004.

III. UNDERSTANDING OF RELEVANT LEGAL PRINCIPLES

A. Claim Construction

12. I understand that claim construction begins with the language of the claims. I understand that words are given their “ordinary and customary meaning,” which is “the meaning that the term would have to a person of ordinary skill in the art in question at the time of the invention, i.e., as of the effective filing date of the patent application.” I further understand that specification and prosecution history are sources of intrinsic evidence that informs the meaning of the claim language. Extrinsic evidence, such as expert testimony and dictionaries, may be useful in educating the court regarding the field of the invention or helping determine what a person of ordinary skill in the art (“POSITA”) would understand claim terms to mean. I understand that extrinsic evidence in general is viewed as less reliable than intrinsic evidence.

13. I understand that in this proceeding, any claim term that is not construed should be given its plain and ordinary meaning. For purposes of this IPR only, I have been asked to assume that the ’080 patent claim terms do not require construction and can be afforded their plain and ordinary meaning.

14. Additionally, I understand that during claim construction in *Jawbone Innovations, LLC v. Google LLC*, C.A. No. 21-cv-00985 (W.D. Tex. Sept. 23, 2021),

the Court determined that the terms “substantially similar” and “substantially dissimilar” are indefinite. I have been asked, for the purposes of this declaration only, to assume the claims are not invalid under § 112. For the purposes of this declaration only, I interpret “substantially similar” responses to include nearly identical responses, and “substantially dissimilar” responses to include very different responses.

B. Obviousness

15. I understand that a prior art reference can render a patent claim obvious to one of ordinary skill in the art if the differences between the subject matter set forth in the patent claim and the prior art are such that the subject matter of the claim would have been obvious at the time the claimed invention was made. In analyzing obviousness, I understand that it is important to consider the scope of the claims, the level of skill in the relevant art, the scope and content of the prior art, the differences between the prior art and the claims, and any secondary considerations.

16. I understand that when the claimed subject matter involves combining pre-existing elements to yield no more than one would expect from such an arrangement, the combination is obvious. I also understand that in assessing whether a claim is obvious one must consider whether the claimed improvement is more than the predictable use of prior art elements according to their established functions. I understand that there need not be a precise teaching in the prior art directed to the

Declaration of Dr. Michael Johnson in Support of
Petition for *Inter Partes* Review of USP No. 10,779,080

specific subject matter of a claim because one can take account of the inferences and creative steps that a person of skill in the art would employ. I further understand that a person of ordinary skill is a person of ordinary creativity, not an automaton.

17. I understand that obviousness cannot be based on the hindsight combination of components selectively culled from the prior art. I understand that in an obviousness analysis, neither the motivation nor the avowed purpose of the inventors controls the inquiry. Any need or problem known in the field at the time of the invention and addressed by the patent can provide a reason for combining elements. For example, I understand that it is important to consider whether there existed at the time of the invention a known problem for which there was an obvious solution encompassed by the patent's claims. I understand that known techniques can have obvious uses beyond their primary purposes, and that in many cases a person of ordinary skill can fit the teachings of multiple pieces of prior art together like pieces of a puzzle.

18. I understand that, when there is a reason to solve a problem and there are a finite number of identified, predictable solutions, a person of ordinary skill has good reason to pursue the known options within his or her technical grasp. I further understand that, if this leads to the anticipated success, it is likely the product not of

Declaration of Dr. Michael Johnson in Support of
Petition for *Inter Partes* Review of USP No. 10,779,080

innovation but of ordinary skill and common sense, which bears on whether the claim would have been obvious.

19. I understand that secondary considerations can include, for example, evidence of commercial success of the invention, evidence of a long-felt need that was solved by an invention, evidence that others copied an invention, or evidence that an invention achieved a surprising result. I further understand that such evidence must have a nexus, or causal relationship to the elements of a claim, in an order to be relevant. I am unaware of any such secondary considerations for the '080 patent.

IV. MATERIALS REVIEWED AND CONSIDERED

20. All of the opinions contained in this Declaration are based on the documents I reviewed and my knowledge and professional judgment. In forming the opinions expressed in this Declaration, while drawing on my experience in the field of speech and audio signal processing, I reviewed the following documents:

TABLE OF EXHIBITS

Exhibit No.	Description
1001	U.S. Patent No. 10,779,080 to Burnett (“the '080 patent”)
1002	File History of U.S. Patent No. 10,779,080

Declaration of Dr. Michael Johnson in Support of
Petition for *Inter Partes* Review of USP No. 10,779,080

Exhibit No.	Description
1005	Excerpts of MICROPHONE ARRAYS: SIGNAL PROCESSING TECHNIQUES AND APPLICATIONS (Michael Brandstein & Darren Ward eds., Springer-Verlag 2001) (“Brandstein”)
1006	Lloyd Griffiths & Charles Jim, <i>An Alternative Approach to Linearly Constrained Adaptive Beamforming</i> , vol. AP-30, no. 1 IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, 27 (Jan. 1982) (“Griffiths-Jim”)
1007	Japanese Unexamined Patent Application Publication No. H11-18186A to Ikeda et al. (“Ikeda”)
1008	Certified Translation of Japanese Unexamined Patent Application Publication No. H11-18186A to Ikeda et al. (“Ikeda”)
1009	Iain A. McCowan et al., <i>Near-Field Adaptive Beamformer for Robust Speech Recognition</i> , vol. 12, no. 1 DIGITAL SIGNAL PROCESSING, 87 (Jan. 2002) (“McCowan”)
1010	U.S. Patent Application Publication No. 2002/0193130 to Yang et al. (“Yang”)
1011	U.S. Patent No. 5,651,071 (“Lindemann”)
1012	U.S. Patent No. 5,627,799 (“Hoshuyama”)
1013	Excerpt of Bernard Widrow & Samuel D. Stearns, <i>Adaptive Signal Processing</i> , Prentice-Hall (1985) (“Widrow”)
1014	Excerpt of Alan V. Oppenheim et al., <i>Discrete-Time Signal Processing</i> , Prentice-Hall, Inc. (2nd ed. 1999) (“Oppenheim”)
1015	U.S. Patent No. 5,471,538 to Sasaki et al. (“Sasaki”)

Declaration of Dr. Michael Johnson in Support of
Petition for *Inter Partes* Review of USP No. 10,779,080

Exhibit No.	Description
1016	U.S. Patent No. 6,370,401 to Baranowski et al. (“Baranowski”)
1017	U.S. Patent No. 6,006,115 to Wingate (“Wingate”)
1018	U.S. Patent No. 5,590,417 to Rydbeck (“Rydbeck”)
1019	U.S. Patent Application Publication No. 2005/0259811 to Kimm et al. (“Kimm”)
1020	U.S. Patent Application Publication No. 2003/0179888 to Burnett et al. (“Burnett II”)
1021	Claim Construction Order, ECF No. 88, <i>Jawbone Innovations, LLC v. Google LLC</i> , C.A. No. 21-cv-00985 (W.D. Tex.)
1023	U.S. Patent No. 7,386,135 (“Fan”)

V. LEVEL OF SKILL IN THE ART AND PERSPECTIVE APPLIED IN THIS DECLARATION

21. I understand that certain issues relating to validity must be judged from the perspective of a person of ordinary skill in the relevant art, as I discuss below.

22. The invention of the '080 patent generally relates to noise suppression systems, devices, and methods for use in acoustic applications.

23. In determining the level of ordinary skill, I have been asked to consider, for example, the types of problems encountered in the field, prior solutions to those problems, the rapidity with which innovations are made, the sophistication of the

Declaration of Dr. Michael Johnson in Support of
Petition for *Inter Partes* Review of USP No. 10,779,080

technology, and the educational level of active workers in the field. Taking those factors into consideration, a person of ordinary skill in the art at the time of the claimed priority date of the '080 patent would have had a minimum of a bachelor's degree in computer engineering, computer science, electrical engineering, mechanical engineering, or a similar field, and approximately three years of industry or academic experience in a field related to acoustics, speech recognition, speech detection, or signal processing. In my opinion, work experience could substitute for formal education, and additional formal education could substitute for work experience.

24. My opinions do not turn on the precise definition, and the challenged claim would be unpatentable from the perspective of any reasonable POSITA.

25. I believe that I would have qualified as a person of at least ordinary skill in the art as of the earliest claimed June 13, 2007 priority date on the face of the '080 patent. I believe that I have a sufficient level of knowledge, experience, and education to provide an expert opinion in the field of the '080 patent.

26. My opinions in this Declaration are based on the perspective of a person of ordinary skill in the art as of June 13, 2007. This is true even if the testimony is stated in the present tense. Each of the statements below reflects my opinion based on my review of the disclosures of the '080 patent, the file history of the '080 patent, and the challenged claims.

VI. BACKGROUND

27. As of the earliest priority date of the '080 patent in 2007, techniques for reducing noise in signals using multiple microphones had been known for decades. One prominent technique was the Generalized Sidelobe Canceler, or GSC. The GSC is sometimes also called the Griffiths-Jim beamformer (“GJBF”), after its inventors.

A. In 1982, Griffiths and Jim Publish Their Seminal Article on the GSC.

28. In 1982, Lloyd Griffiths and Charles Jim published an article in IEEE Transactions on Antennas and Propagation describing “a simple time-varying beamformer which can be used to combine the outputs of an array of sensors.” Ex. 1006, 27. The purpose of the beamformer was “to minimize the effects of noise and interference at the array output” while capturing the target signal. Ex. 1006, 27.

29. Griffiths and Jim called their beamformer a “generalized sidelobe canceling” structure. Ex. 1006, 29. As shown in Fig. 4 of their paper, the signal processor had two main substructures: a top branch that was a “conventional beamformer” designed to capture the target signal plus noise, and a bottom branch that was the “sidelobe canceling path” designed to capture only noise so that the noise could be subtracted or canceled out:

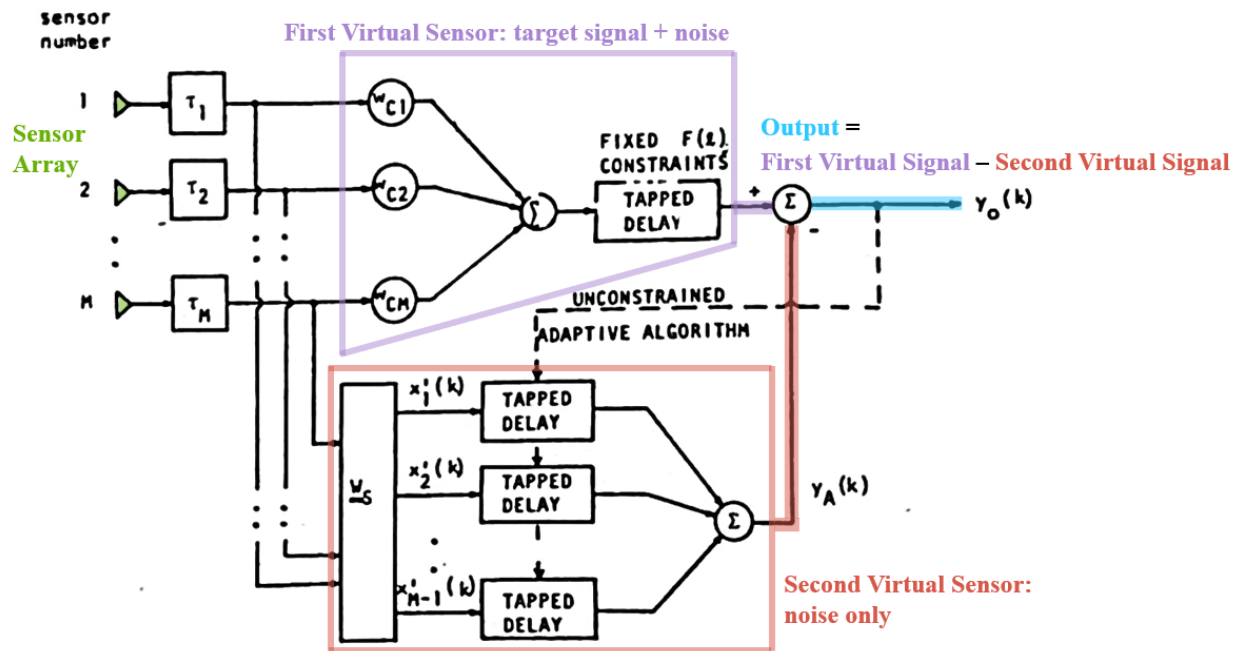


Fig. 4. Generalized sidelobe canceling form of linearly constrained adaptive array processing algorithm.

Ex. 1006, Fig. 4 (annotated)

Ex. 1006, 29–30.¹ In the top branch of the GSC, which I will refer to as the **first virtual sensor**, the outputs of the **array sensors** are combined to form a conventional beamformer. The sensor outputs are delayed by “time-delay steering elements $\tau_1, \tau_2, \dots, \tau_M$ ” and then are combined by multiplying the sensor output signals by factors called “weights” ($w_{c1}, \dots, w_{cM}$ in the paper). These “weights” are also sometimes also called “gains.” The weighted sensor signals are further summed. 1006, 29–30.

¹ I have annotated the figures with color throughout.

Griffiths and Jim denote the output of the first virtual sensor as $y_c'(k)$ in the paper's equations. Ex. 1006, 30. This output contains the target signal and noise.

30. As Griffiths-Jim explains, “[t]he lower path in Fig. 4 is the sidelobe canceling path” Ex. 1006, 30. I refer to this as the **second virtual sensor**. Like the first virtual sensor, the second virtual sensor is formed from a combination of the outputs of the array sensors. However, the combination used for the second virtual sensor differs from the combination used for the first virtual sensor. The lower path includes a blocking matrix (BM) designed to “block the desired signal $s(k)$ from the lower path.” Ex. 1006, 30. The array sensor outputs are combined by filtering and summing (including by blocking the desired signal), and the output of the second virtual sensor is denoted in the paper as $y_A(k)$. Ex. 1006, 30. The second virtual sensor's output “ $y_A(k)$ contains no desired signal terms” and instead “contains only noise and interference terms.” Ex. 1006, 30.

31. By subtracting the noise-only output of the second virtual sensor from the target-and-noise output of the first virtual sensor, the overall **output** of the GSC, $y(k)$, is obtained:

$$y(k) = y_c'(k) - y_A(k). \quad (29)$$

Ex. 1006, 30. The resulting cleaned-up signal contains reduced noise without distorting the desired signal. Ex. 1006, 30 (output due to desired signal satisfies the

constraint defined by paper's equation 9, which defines a constraint for "zero distortion" Ex. 1006, 28.

B. The GSC Was Applied to Microphone Arrays to Reduce Noise

32. In the two decades following Griffiths and Jim's article, the GSC was applied to many signal-processing applications, including microphone arrays to reduce noise in speech applications. For example, U.S. Patent No. 5,651,071 ("Lindemann"), filed in 1993, cites the article and explains that it was known that a GJBF could be used "to improve signal-to-noise ratio for hearing aids." Ex. 1011, 1:40-46, 12:12-14.² As another example, Griffiths-Jim is the first non-patent reference cited in U.S. Patent No. 5,627,799 ("Hoshuyama '799 patent"), filed in 1995, which relates to "interference cancelers, and more particularly to a generalized sidelobe canceler, or adaptive beamformer for an array of sensors such as microphones[.]" Ex. 1012, 1:8-11. Describing what was prior art even at that time, Hoshuyama explains that one way the GSC had been used with microphone arrays was as follows:

According to a prior art microphone array, signals detected by an array of microphones are lowpass filtered and summed together to detect a target signal that arrives in a particular direction. The adaptive microphone array beamformer is one form of the generalized sidelobe canceler as described in an article "An alternative Approach to Linearly Constrained Adaptive Beamforming", Lloyd J. Griffiths and Charles

² Patent citations are in column:line format.

W. Jim, the IEEE Transactions on Antenna and Propagation, Vol. AP-30, No. 1, January 1982, pages 27-34.
(*Id.*, 1:17-26.)

C. A 2001 Book Edited by Brandstein Illustrates How to Apply a GSC to Microphone Arrays to Reduce Noise.

33. Published in 2001, MICROPHONE ARRAYS: SIGNAL PROCESSING TECHNIQUES AND APPLICATIONS (Ex. 1005 (“Brandstein”)), provides “a single complete reference on microphone arrays.” Ex. 1005., Preface. The book became a standard reference for those in the field of audio-signal processing.

34. At the outset of Chapter 5, which discusses robust adaptive beamforming, Brandstein explains that “[a]pplications of beamforming include microphone arrays for speech enhancement.” Ex. 1005, 87 (original page numbering). As Brandstein explains, “[t]he goal of speech enhancement is to remove undesirable signals such as noise and reverberation.” Ex. 1005, 87. Brandstein further explains that, among various known adaptive beamformers, “the Griffiths-Jim beamformer (GJBF), or the generalized sidelobe canceler, is most widely known.” Ex. 1005, 88 (internal citation omitted). In Fig. 5.1, Brandstein depicts the structure of the basic GJBF. Ex. 1005, 88.

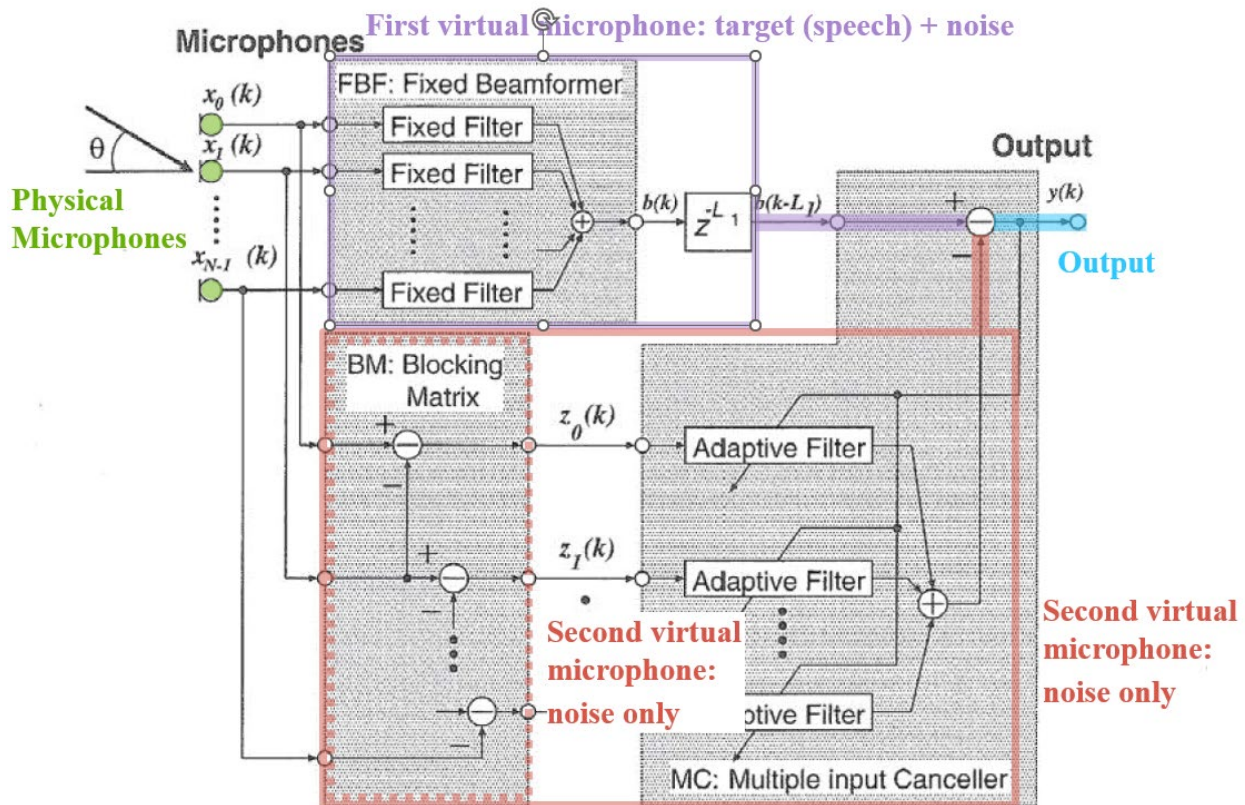


Fig. 5.1. Griffiths-Jim beamformer. It comprises a fixed beamformer (FBF), a multiple-input canceler (MC), and a blocking matrix (BM).

Ex. 1005, Fig. 5.1 (annotated)

As shown, the signals of a first and second physical microphone, $x_0(k)$ and $x_1(k)$, are combined in the top branch, which is a first virtual microphone, by filtering and summing the signals to form a fixed beamformer (FBF). The first virtual microphone captures the target speech signal plus noise.

35. In the bottom branch (with two options for second virtual microphone, as the BM alone or entire bottom branch), the signals of the two physical microphones are combined by summing and filtering the signals in a different way to form

a **second virtual microphone**, which is depicted in the annotated Fig. 5.1 above as either including the BM alone or together with the multiple input canceller (MC). As Brandstein explains, “the BM forms a null in the look direction so that the target signal is suppressed and all other signals are passed through.” Ex. 1005, 88. Brandstein further explains that “[t]he BM was named after its function, which is to block the target signal.” Ex. 1005, 88. As a result, the second virtual microphone captures noise only.

36. The overall **output** is the speech-and-noise output of the first virtual microphone minus the noise-only output of the bottom branch. As a result, “in the subtracter output $y(k)$, the target signal is enhanced and undesirable signals such as ambient noise and interferences are suppressed.” Ex. 1005, 88-89; *see also* 33-34 (explaining that the lower branch of the GJBF is “subtracted from the output of the upper part of the structure to get the noise-reduced output signal.”)

37. The two virtual microphones have very different responses to the target signal (i.e., speech). The first virtual microphone is designed to capture the target speech signal; on the other hand, the second virtual microphone (whether BM only or with MC) is designed to block it. In contrast, they have nearly the same responses to noise so that the final subtraction removes the noise and enhances the target

speech. Brandstein illustrates this in Fig. 5.2, which shows the directivity pattern for the final output of an example GJBF:

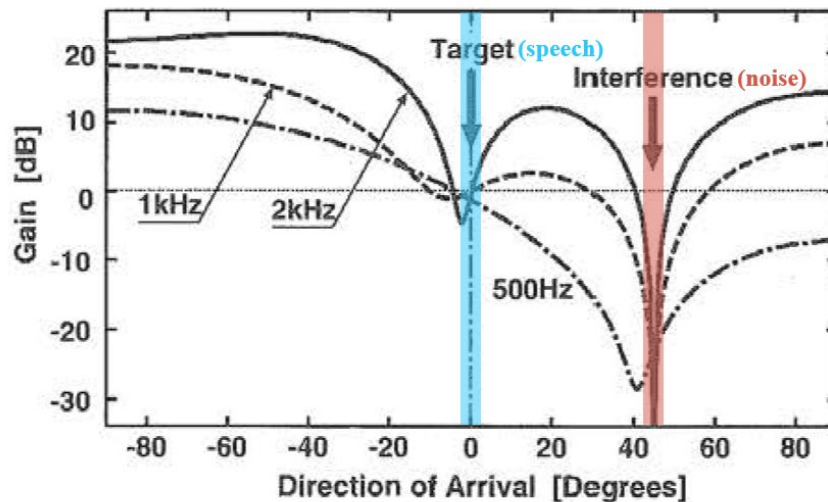


Fig. 5.2. Example directivity pattern of the Griffiths-Jim beamformer.

Ex. 1005, Fig. 5.2 (annotated)

The horizontal axis of the graph shows direction of arrival (“DOA”) measured in degrees relative to the direction perpendicular to the axis of the microphone array. As shown, the **target (speech) signal** is at 0 degrees, while the **interference (noise) signal** is at approximately 45 degrees. The vertical axis of the graph shows gain in decibels: zero gain corresponds to no change in sound pressure or signal power, while a positive gain corresponds to an increase in signal power, and a negative gain corresponds to a decrease in signal power.

38. As highlighted in blue, the target (*speech*) signal is reproduced faithfully with essentially zero gain, reflecting that subtracting the second virtual microphone's noise-only output from the first virtual microphone's speech-and-noise output will produce an unchanged target (*speech*) signal in the direction of the target (*speech*). On the other hand, as highlighted in red, Fig. 5.2 shows a highly negative gain in the direction of the interference (*noise*) signal, reflecting that subtracting the two virtual microphones' outputs cancels the noise. Ex. 1005, 90. As Brandstein explains, “[i]n the direction of the target signal, almost constant gains close to 0 dB are obtained over a wide range of frequencies. On the contrary, in the direction of the interference, a deep null is formed.” Ex. 1005, 90.

39. Brandstein further provides the individual directivity pattern for the FBF and BM components of the GJBF in Fig. 5.3. These directivity patterns show non-zero responses to the interference (*noise*) in the direction of the interference. The directivity patterns also show that BM response has a null in the direction of the target (*speech*) while the FBF response has a non-zero response to the target in that direction.

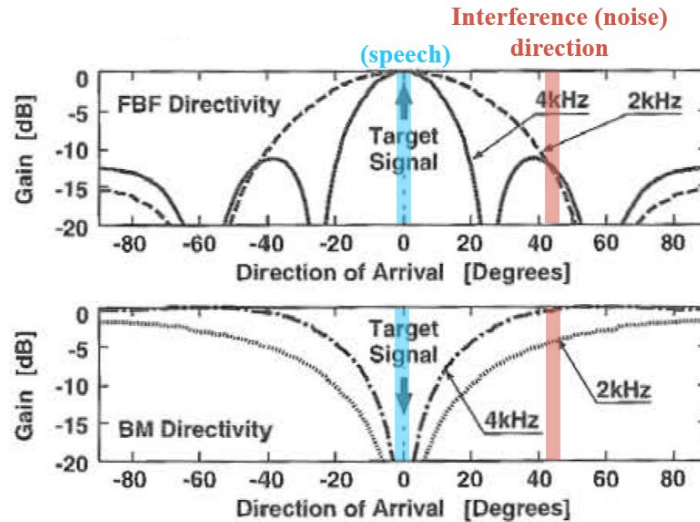


Fig. 5.3. Directivity pattern of a fixed beamformer (FBF) and a blocking matrix (BM).

Ex. 1005, Fig. 5.3 (annotated)

40. According to Brandstein, the basic GJBF depicted in Fig. 5.1 may have errors in the steering-vector, leading to undesirable “leakage” of the target in the BM and potential cancellation of the target signal at the output. Ex. 1005, 90-91, 97. To reduce this undesirable effect, Brandstein teaches ways for the GJBF to (1) track the direction of the target signal, and (2) adapt the BM coefficients only when the target signal is absent, thereby minimizing the effect of target leakage on performance of the beamformer. Ex. 1005, 90-91, 97. To accomplish this, Brandstein incorporates adaptive filters into the bottom branch of the GJBF (containing the *second virtual*

microphone), coupled with an adaptation-mode controller (“AMC”) to detect presence or absence of the target signal. Ex. 1005, 90-91, 97. Brandstein refers to this implementation as the “GJBF-CNA,” which is depicted in Fig. 5.8:

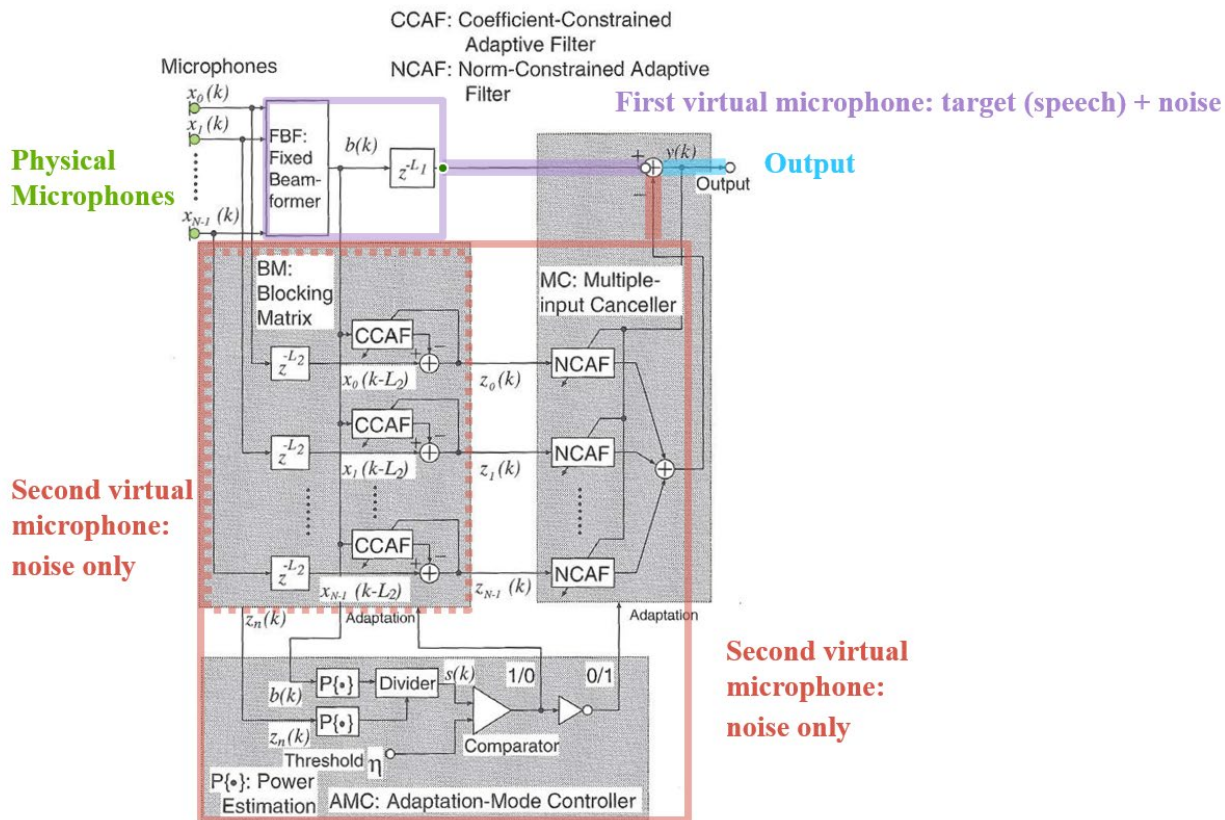


Fig. 5.8. CCAF-NCAF structure with an AMC.

Ex. 1005, Fig. 5.8 (annotated)

41. Brandstein depicts the “sensitivities to the frequency component of the target signal” for the GJBF-CNA, which are plotted in Fig. 5.10 (below). Fig. 5.10 shows how the sensitivities depend on different frequencies (500, 1000, 2000, 3000, and 3500 Hz). Ex. 1005, 100. Brandstein explains that the GJBF-CNA is suitable

for broadband applications such as microphone arrays because the frequency dependency is small throughout a substantial angular region in the direction of the target source, as shown in Fig. 5.10, reproduced below. Ex. 1005, 100.

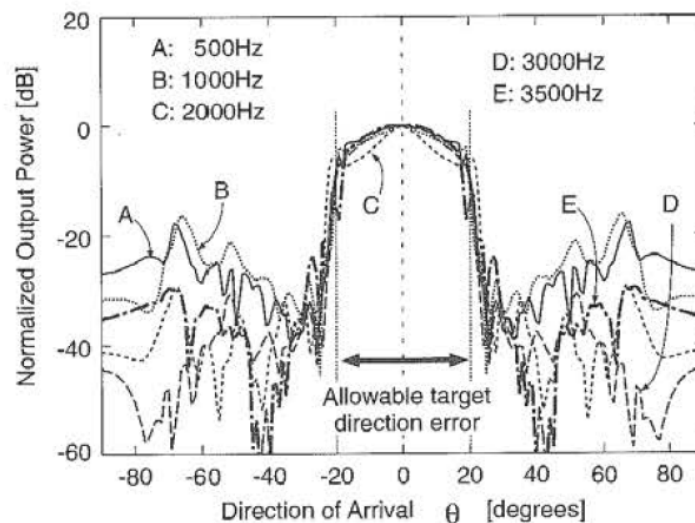
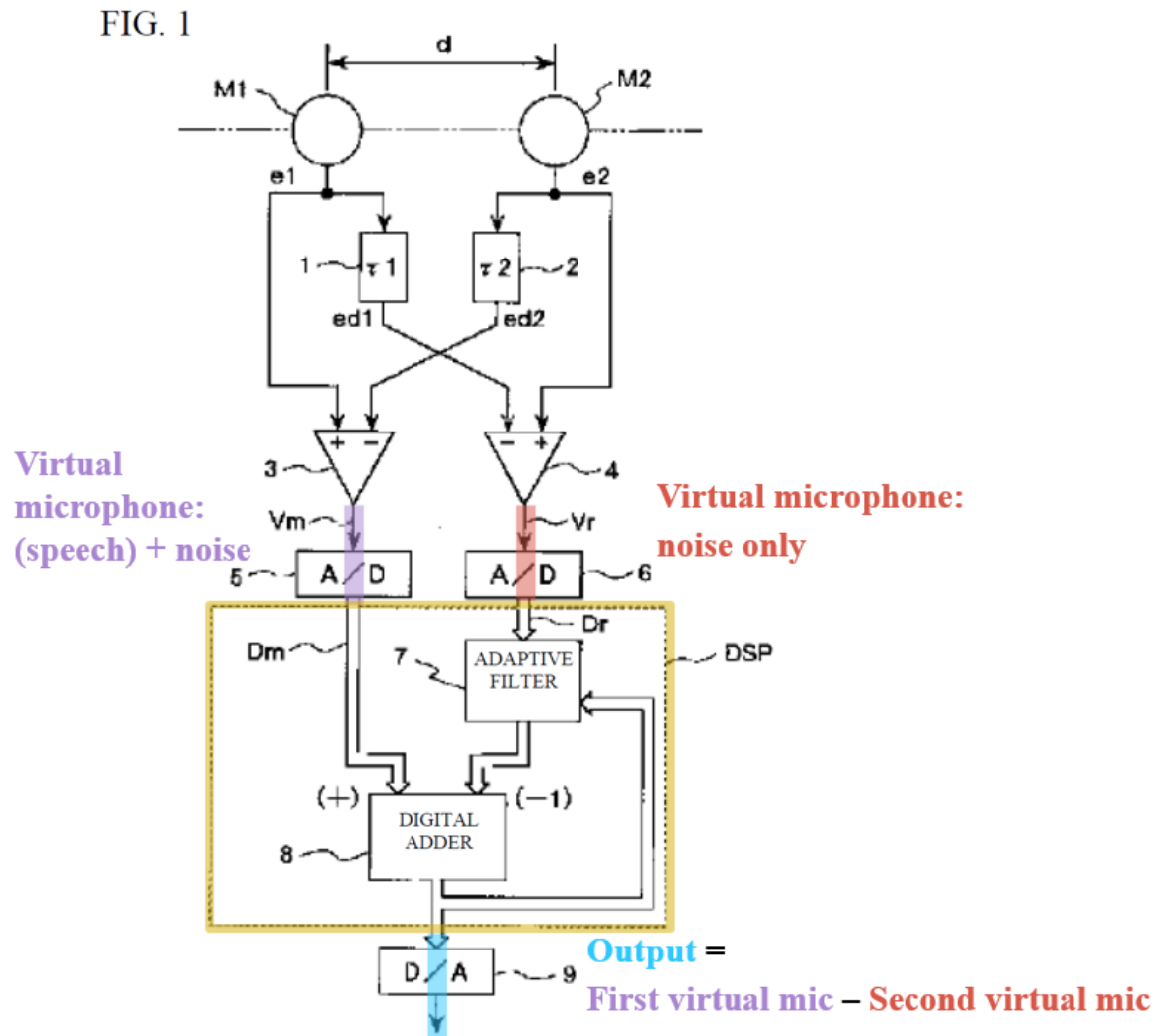


Fig. 5.10. Sensitivities after convergence as a function of DOA at different frequencies.

D. Adaptive Noise Removal Methods Were Well Known

42. By June 13, 2007, it was well-known for adaptive noise removal techniques to include filtering, summing, and transfer functions. For example, Japanese Unexamined Patent Application Publication No. H11- 18186A to Ikeda et al. (“Ikeda”) discloses a microphone apparatus that includes an adaptive noise cancellation system. Ex. 1008, ¶¶[0018]-[0024], [0027]-[0030], Abstract, Fig. 1. Ikeda’s microphone apparatus includes “two omnidirectional microphone units M1 and M2.” Ex. 1008, ¶[0018], Fig. 1.

43. As shown in Fig. 1, Ikeda teaches a virtual microphone Vm (main signal) that is formed by combining a delayed physical microphone signal M2 with a physical microphone signal M1. Ex. 1008, ¶¶[0002]-[0004], [0012], [0018]-[0022], [0043-44], Figs. 2-3. Ikeda also teaches another virtual microphone Vr (reference signal) that is formed by combining a delayed M1 with M2. Ex. 1008, ¶¶[0012], [0018]-[0020], [0023]-[0024], Figs. 4-5.



Ex. 1008, Fig. 1 (annotated).

44. Ikeda explains that Vm, the **main virtual microphone**, “has directional characteristics for picking up a target sound and noise from the front surface side,” like the top branch of the GJBF. On the other hand, Vr, the **reference virtual microphone**, “has directional characteristics for picking up noise from the back surface side without picking up the target sound and noise from the front surface side,” like the bottom branch of the GJBF. Ex. 1008, ¶¶[0021], [0023], ¶¶[0012], [0026], Fig. 10, Abstract.

45. Ikeda’s microphone array converts main signal Vm and reference signal Vr to digital signals Dm and Dr, respectively, and inputs them into a digital signal processor (DSP), which acts as an adaptive **noise cancellation system**. As shown in Fig. 1, the DSP includes an adaptive filter 7 and a digital adder 8. Ex. 1008, ¶¶[0012], [0027]-[0029]. Signal Dr is input to adaptive filter 7, and digital adder 8 calculates the difference between the output of adaptive filter 7 and signal Dm. Ex. 1008, ¶[0027].

46. Ikeda explains that the response characteristics of adaptive filter 7 “are determined so as to cancel the respective noise components contained in” signal Dm and signal Dr. Ex. 1008, ¶[0028]. Specifically, Ikeda states:

As a result of the response characteristics of the adaptive filter 7 being determined successively in this way, the noise component contained in the main signal Vm is cancelled by the noise component contained in the reference signal Vr due to subtraction by digital adder 8. The output

of the digital adder 8 is converted to an analog signal by a D/A converter 9. A signal primarily consisting of the target sound component can be acquired from the D/A converter 9.

Ex. 1008, ¶[0029], ¶¶[0012], Abstract.

E. Near-field Compensation Techniques Were Well Known.

47. Delaying signals based on geometry to adjust for differences in arrival times was well known. Adaptive beamforming relies on the spatial geometry of the array and the signal's direction of arrival. Ex. 1005, 87. When the source of the target signal is close to the array, instead of treating the signal's wavefront as a plane, it may be useful to account for the spherical geometry of the wavefront—i.e., to use a *near-field* design. Near-field compensation techniques were well known as of June 13, 2007 and used basic calculations to determine the delays between signal arrival at the various microphones in the array. Ex. 1005, 87. Brandstein discloses the standard near-field design:

- Nearfield design, including attenuation of the desired signal [14], [22]

$$\mathbf{d}^T = [a_0 \exp(-j\omega\tau_0), a_1 \exp(-j\omega\tau_1), \dots, a_{N-1} \exp(-j\omega\tau_{N-1})], \quad (2.28)$$

$$a_i = \frac{\|\mathbf{q} - \mathbf{p}_{ref}\|}{\|\mathbf{q} - \mathbf{p}_i\|}, \quad (2.29)$$

$$\tau_i = \frac{\|\mathbf{q} - \mathbf{p}_{ref}\| - \|\mathbf{q} - \mathbf{p}_i\|}{c}, \quad (2.30)$$

where $\|\mathbf{q} - \mathbf{p}_{ref}\|$ and $\|\mathbf{q} - \mathbf{p}_i\|$ denote the distance between the vector location of the source $\mathbf{q}$ and a reference sensor $\mathbf{p}_{ref}$, or the sensor $\mathbf{p}_i$, respectively.

Ex. 1005, 25. The time it takes for a signal to travel from a source to a microphone is the distance traveled divided by the speed of sound, c . The τ_i terms above represent the time difference for a signal to propagate from the source to microphone i compared to a reference microphone. *Id.* The complex exponential shown above includes the τ_i time difference terms. By applying that exponential to the signal received at the reference microphone, the received signal is delayed to reproduce (in terms of phase) the signal as received at microphone i . Applying the a_i term to the reference signal adjusts the amplitude of the signal to match the amplitude of the signal as it is received at microphone i . The signal x_i on microphone i can be expressed in terms of the signal x_{ref} on the reference microphone as follows (where $j = \sqrt{-1}$):

$$x_i(\omega) = a_i(e^{j\omega\tau_i})x_{ref}(\omega).$$

48. The math relevant to near-field compensation is further explained in the January 2002 edition of the journal DIGITAL SIGNAL PROCESSING, in an article by McCowan et al. titled *Near-Field Adaptive Beamformer for Robust Speech Recognition* (“McCowan”). Ex. 1009, 89-90. As I explain below, McCowan expressly motivates and teaches incorporating the near-field design into the GSC. *Id.*, 90-96.

F. GSCs Incorporated Voice Activity Detection

49. Voice activity detection was known in the art and incorporated into GJBFs. As an example, Yang discloses a noise suppression system that includes a beam forming unit and a noise suppression unit. Ex. 1010, ¶[0026]. Yang explains that the beam forming unit “may be implemented with a ‘Griffiths-Jim’ beam former.” Ex. 1010, ¶[0038].

50. Yang also teaches an adaptive noise reduction system 230b as shown in Fig. 4 (annotated below). Ex. 1010, ¶¶ [0058]-[0062]. Yang discloses:

The mostly noise signal $x(t)$ is provided to an adaptive filter 450, which filters the noise with a particular transfer function $h(t)$. The filtered noise $p(t)$ is then provided to summer 434 and subtracted from the filtered speech plus noise signal to provide an intermediate signal $d(t)$ having predominantly speech and some amount of noise.

Ex. 1010, ¶ [0059]. The adaptive noise suppression unit includes a **voice activity detector** 440 and an adaptive filter 450. The adaptive filter 450 “is adapted during periods of non-speech activity.” Ex. 1010, ¶¶ [0060]-[0061]. Yang discloses that a voice activity detector 440 “detects the presence of speech activity ... and provides

a control signal that enables the adaptation of the coefficients of the [adaptive filter 450] when no speech activity is detected.” Ex. 1010, ¶¶ [0060]-[0061].

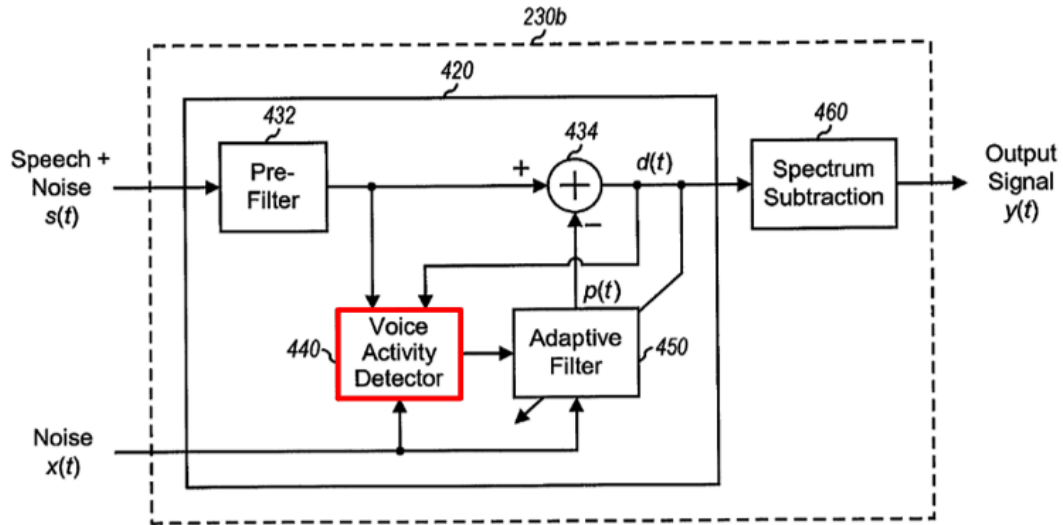


FIG. 4

Ex. 1009, Fig. 4 (annotated).

VII. THE '080 PATENT

A. Specification

51. The '080 patent purports to distinguish itself from prior art “conventional arrays and algorithms” that “seek to reduce noise by nulling out noise sources,” by using a virtual microphone with a null in the direction of speech “to remove the speech of the user.” Ex. 1001, 3:49-56. However, in my opinion, this is exactly what a GSC does. Specifically, a GSC uses a virtual microphone with a null in the direction of the target signal to obtain a noise-only reference.

52. The '080 patent describes a “dual omnidirectional microphone array (DOMA)” for adaptive noise suppression. Ex. 1001, 3:49-50, 4:52-53, Abstract. Two omnidirectional microphones are used to “form two distinct virtual directional microphones which are configured to have very similar noise responses and very dissimilar speech responses.” Ex. 1001, 3:49-56, 4:48-51, 7:39-51. The two virtual microphones are “paired with an adaptive filter algorithm” to reduce noise. Ex. 1001, 3:57-62.

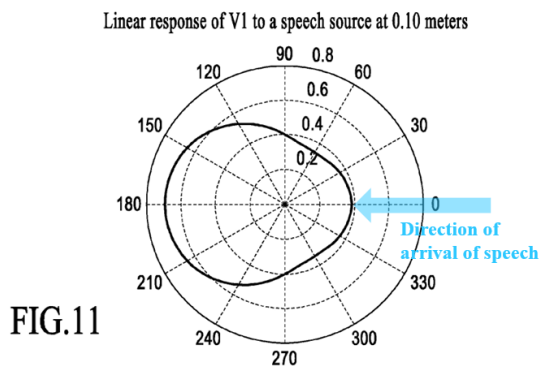
53. In one embodiment, two virtual directional microphones V_1 and V_2 are formed based on two physical omnidirectional microphone outputs O_1 and O_2 according to certain parameters expressed in the '080 patent as:

$$\begin{aligned}\beta &= \frac{d_1}{d_2} \\ \gamma &= \frac{d_2 - d_1}{c} \cdot f_s \text{ (samples)} \\ d_1 &= \sqrt{d_s^2 - 2d_{s1}d_0\cos(\theta) + d_0^2} \\ d_2 &= \sqrt{d_s^2 + 2d_{s1}d_0\cos(\theta) + d_0^2}\end{aligned}$$

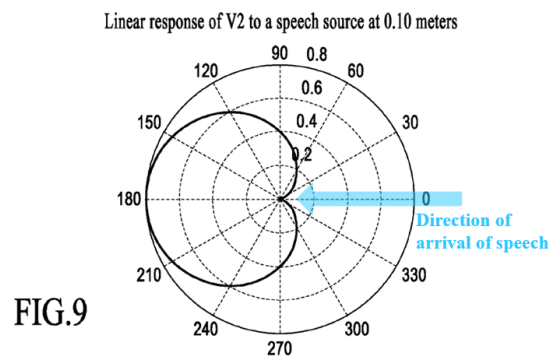
Ex. 1001, 9:55-62. The '080 patent explains that “[t]he distances d_1 and d_2 are the distance from O_1 and O_2 to the speech source (see Fig. 2), respectively, and γ is their difference divided by c , the speed of sound, and multiplied by the sampling frequency f_s .” Ex. 1001, 9:64-67. Thus, γ is simply the difference in propagation

time to the microphones taken from the standard near-field design equations (denoted τ_i in Section VI.E., above), expressed in number of samples instead of other units of time, like seconds. Similarly, β is simply the amplitude-adjustment factor from the standard near-field design (denoted a_i in Section VI.E., above).

54. According to the '080 patent, using these metrics for V_1 and V_2 produces desirable directivity patterns with respect to a speech signal expected to arrive at an angle of zero degrees from 10 centimeters away, as illustrated in Figs. 9 and 11:



Ex. 1001, Fig. 11



Ex. 1001, Fig. 9

As shown, V_1 (Fig. 11) exhibits a response to the speech signal, while V_2 (Fig. 9) has a null in the expected direction of speech “and will therefore exhibit minimal response to the speech.” Ex. 1001, 10:18-33, 11:20-24. In contrast, the '080 patent explains that V_1 and V_2 will supposedly have “very similar” responses to noise originating a meter away from the array. Ex. 1001, 11:25-29.

55. The '080 patent describes the linear response of virtual microphone V_1 as “flat for all frequencies within approximately ± 30 degrees of the axis of the array,” as shown in Fig. 13 (reproduced below). Ex. 1001, 11:30-35, 11:43-61.

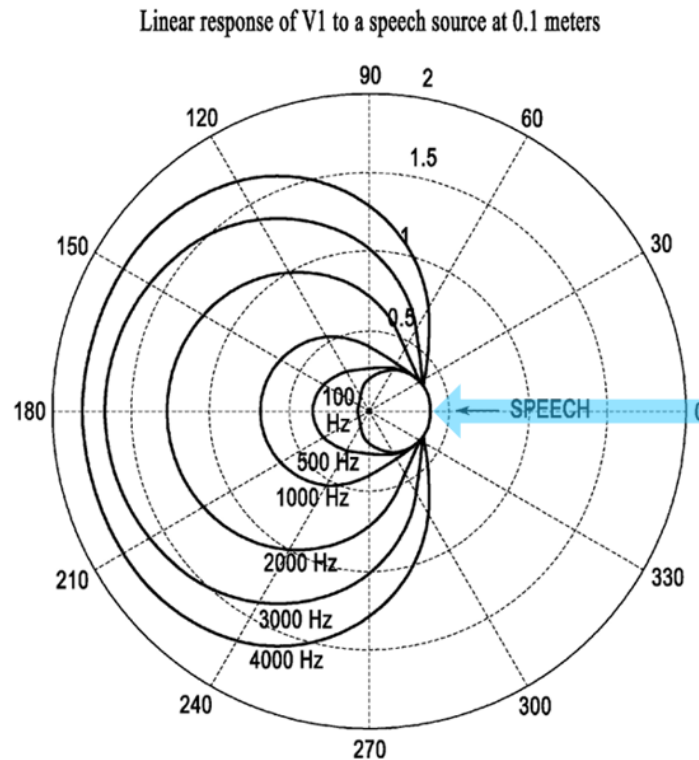


FIG.13

Ex. 1001, Fig. 13 (annotated).

56. The virtual microphone signals are input to an adaptive noise removal system, which can use “any adaptive filter or noise removal algorithm.” Ex. 1001, 3:57-62, 6:67-7:3, 8:50-9:3, 9:22-38, 4:52-7:16. In an embodiment, the adaptive noise removal system combines the signals from virtual microphones V_1 and V_2 “by filtering and summing in the time domain.” Ex. 1001, 7:4-16, 8:50-9:3, 9:22-38. The

'080 patent states that “[t]he adaptive filter generally uses the signal received from a first microphone ... to remove noise from the speech received from at least one other microphone ..., which relies on a slowly varying linear transfer function between the two microphones for sources of noise.” Ex. 1001, 7:7-12. The system then generates an output signal “in which the noise content is attenuated with respect to the speech content.” Ex. 1001, 7:12-16.

B. Prosecution History

57. I understand that the '080 patent was filed on July 22, 2013, as U.S. Patent Application No. 13/948,160 (“the '160 application”). Ex. 1001, Cover. I understand that the '080 patent claims priority to U.S. Patent Application No. 12/139,355, filed on June 13, 2008, and four provisional applications, the earliest of which was filed on June 13, 2007. Ex. 1001, Cover.

58. During prosecution of the '160 application, the Examiner rejected the initial claim as anticipated by U.S. Patent No. 8,068,619 and based on non-statutory double patenting. Ex. 1002, 57, 98-106. In response, the applicant canceled the claim and added twenty new claims. Ex. 1002, 113-123. The Examiner then rejected the new claims based on non-statutory double patenting. Ex. 1002, 130-136. The applicant filed a terminal disclaimer to overcome the double patenting rejection. Ex.

1002, 142-153, 160-168, 222-234, 248. The Examiner then allowed the claims, stating that “a dual virtual microphone array as claimed including having a linear response to both speech and noise could not be anticipated nor, in combination, be rendered obvious over the prior art of record.” Ex. 1002, 249-256.

C. The Earliest Priority Date of the '080 Patent Claims Is June 13, 2007.

59. The earliest application to which the '080 patent claims priority is U.S. provisional application 60/934,551. I understand that the earliest possible priority date for all challenged claims is the '551 application's filing date, June 13, 2007.

D. The References Are Prior Art.

60. I understand that all references relied on in the grounds (§4, above) are prior art under pre-AIA §102(b) because they published more than one year before June 13, 2007:

- (1) Brandstein published no later than 2001 (Ex. 1005);
- (2) Ikeda published no later than 1999 (Ex. 1007; Ex. 1008 (English Translation));
- (3) McCowan published no later than 2002 (Ex. 1009);
- (4) Yang published no later than 2002 (Ex. 1010).

E. The Asserted References Are Analogous Art.

61. I understand that prior art is analogous if (1) it is from the same field of endeavor, regardless of the problem addressed, or (2) it is reasonably pertinent to the particular problem with which the inventor is involved. The references I rely on for the Grounds are analogous art because, in my opinion, they are from the same field as the '080 patent, e.g., microphone array signal processing and noise reduction. Ex. 1001, 1:21-24, 3:49-50. Additionally, in my opinion, those references are also pertinent to a problem the inventor was focused on, e.g., noise suppression. Ex. 1001, 1:21-24, 3:49-50.

VIII. GROUNDS OF UNPATENTABILITY

62. For the reasons I explain in greater detail below, it is my opinion that as of the priority date of the '080 patent on its face (June 13, 2007), a POSITA would have considered the claims of the '080 patent unpatentable as obvious over the prior art discussed in this Declaration and listed in the Grounds above in ¶4.

A. Ground 1: Brandstein in View of Ikeda

63. For the reasons I explain below, it is my opinion that claims 1-3, 5-9, 11-14, and 16-20 would have been obvious in view of Brandstein and Ikeda.

1. **Motivation to Combine**

a) **Motivation to Combine Teachings of Brandstein's Embodiments**

64. Brandstein teaches two embodiments of the GJBF: a basic GJBF (depicted in Figs. 5.1-5.3) and a proposed GJBF (or GJBF-CNA, depicted in Figs. 5.8-5.10). While both embodiments are similar (with identical top branches including the same FBF), the bottom branch of the proposed GJBF is more sophisticated, in that it includes delay modules and adaptive filters in the BM, norm-constrained adaptive filters in the MC, and an adaptation-mode controller (AMC) to control adaptation of the BM and MC. (*Compare* Fig. 5.1 *with* Fig. 5.8; *see also* Ex. 1005, 88-90, 97-99.) While Brandstein does not expressly disclose that the basic GJBF's directivity pattern for the top branch that includes the FBF (*the first linear response*) is *substantially similar across a plurality of frequencies*, in my opinion, this would have been obvious to the POSITA based on the teachings of Brandstein related to the proposed GJBF and the knowledge of the POSITA.

65. Specifically, Brandstein's GJBF-CNA (Fig. 5.8) meets this limitation because the corresponding response depicted in Fig. 5.10 depicts a "constant" sensitivity for different frequencies. Ex. 1005, 99-100, Fig. 5.10. A POSITA would have been motivated to modify Brandstein's basic GJBF with Brandstein's GJBF-

CNA structure, so that the *first linear response to speech* is constant across a *plurality of frequencies*.

66. It is my opinion that a POSITA would have been motivated to apply Brandstein's teachings related to the GJBF-CNA to the simple GJBF. Brandstein teaches that target-signal cancellation due to steering-vector errors may occur in the basic GJBF, which potentially leads to undesirable "leakage" of the target in the BM and target signal cancellation at the output. Ex. 1005, 90-91, 97. To reduce this undesirable effect, Brandstein incorporates adaptive filters into the bottom branch of the GJBF-CNA, coupled with an AMC to detect presence or absence of the target signal. *Id.*, 87-90, Fig. 5.8. As a result, the modified GJBF can (1) track the direction of the target signal, and (2) adapt the BM coefficients only when the target signal is absent, so that any target leakage has a minimal effect on performance of the beamformer. Ex. 1005, 90-91, 97. In my opinion, a POSITA would have been motivated to apply these modifications, to reduce target cancellation.

67. Moreover, because both of Brandstein's embodiments share the same goals and have the same overall beamforming structure (*e.g.*, subtract a noise-only signal from a target plus noise signal using spatial channels), the POSITA would have understood that the overall shapes and characteristics of the response patterns

of the basic GJBF's top and bottom branches (including the FBF and BM) are unchanged when the basic GJBF is modified.

68. For example, comparing Figs. 5.1 and 5.8, the POSITA would have appreciated that the FBF is the same between the simple and proposed GJBFs (Ex. 1005, 89, 98), such that the response of the top branch (containing both speech and noise) is the same in both.

69. The POSITA would have known that the entire purpose of modifications to the bottom branch of the GJBF is to make it more robust to steering vector errors toward the direction of the target without changing the response of the BM elsewhere. As Brandstein depicts in Figs. 5.9 and 5.10, the proposed GJBF also has the benefit of having a wider region that is responsive to the target *speech* signal, thus achieving a “robustness against 20° target-direction error and high interference-reduction performance.” Ex. 1005, 100. The POSITA would have understood that this wider region is achieved due to the proposed GJBF's BM, which incorporates adaptive filters to “provide[] a wider null for the target with sharper edges.” Ex. 1005, 97. A POSITA would have understood that the other modifications to the bottom branch of the GJBF-CNA are to provide for better tracking performance, which has no impact on the response of the BM or overall characteristics and shape of the response pattern for the bottom branch.

70. For these reasons, it is my opinion that a POSITA would have understood that the overall characteristics and shape of the response patterns would remain unchanged, and therefore would have been motivated to apply Brandstein's teachings regarding the directivity patterns for the basic GJBF, FBF, and BM to Brandstein's modified GJBF.

b) Motivation to Combine Brandstein and Ikeda

71. Although Brandstein generally discloses noise removal techniques (Ex. 1005, 88-90), it does not expressly provide details about how the noise removal techniques are implemented. Ikeda provides that detail and implements its adaptive noise removal technique on signals that are similar to the signals from the top and bottom branches of the GSC (first and second virtual microphone signals, respectively). In my opinion, it would have been obvious to implement the DSP of Ikeda in a GSC, such as the one described in Brandstein. First, a POSITA would have recognized that the output of Brandstein's GSC FBF is similar to Ikeda's main signal V_m because both contain speech and noise. Thus, a POSITA would have been motivated to couple Brandstein's FBF output with Ikeda's V_m (which, I will refer to as a first virtual microphone).

72. For the second virtual microphone, I have considered two alternative mappings for the combination of Brandstein with Ikeda, which I explain below.

73. In the first mapping, a POSITA would have been motivated to use the output of Brandstein's MC's summer (together with the BM, the second virtual microphone) as the reference signal V_r for feeding into Ikeda's DSP. I have annotated this mapping below:

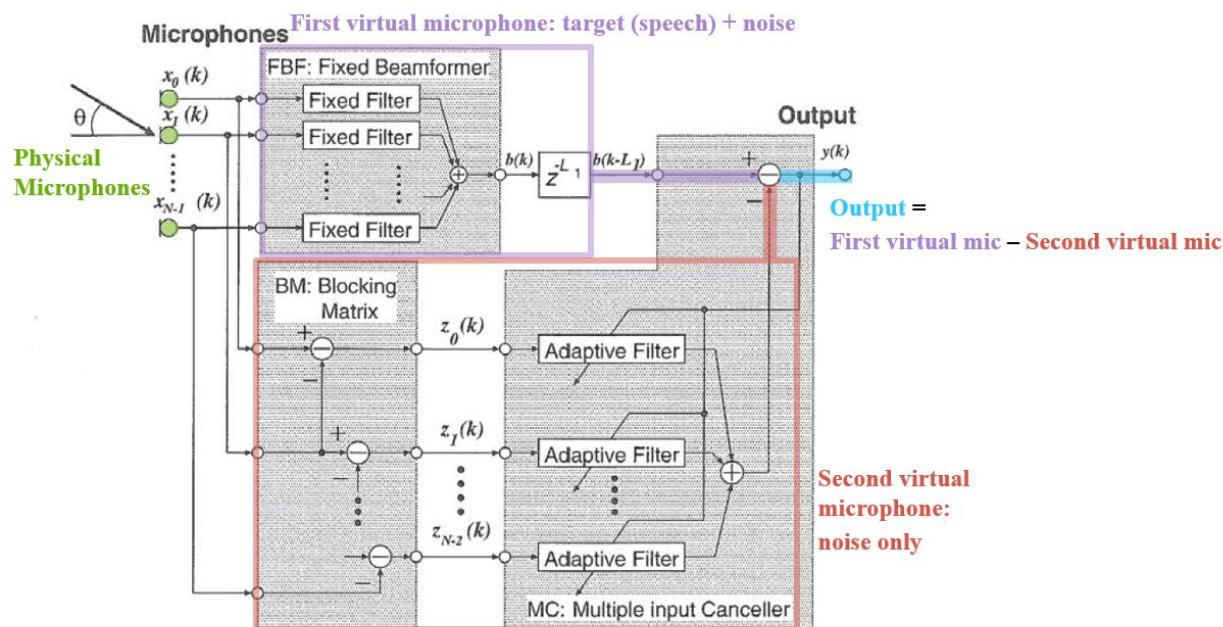


Fig. 5.1. Griffiths-Jim beamformer. It comprises a fixed beamformer (FBB), a multiple-input canceler (MC), and a blocking matrix (BM).

Ex. 1005, Fig. 5.1 (annotated, first mapping)

74. A POSITA would have recognized that adding the DSP of Ikeda to Brandstein's GSC in this way would have improved the noise-cancellation performance of the GSC. Specifically, the POSITA would have known that including additional noise-cancellation stages in Brandstein's GSC would have been beneficial as it would have improved system robustness and stability, while also reducing residual noise and distortion from the GSC beamformer. Specifically, a POSITA

would have readily understood that including additional noise-cancellation stages in Brandstein's modified GJBF would have been beneficial as it would have improved system robustness and stability, while also reducing residual single-channel noise and distortion remaining from the beamformer, which mainly reduces spatial noise across multiple channels.

75. In the second mapping, a POSITA would have been motivated to use the output of Brandstein's BM (the second virtual microphone) as the reference signal V_r for feeding into Ikeda's DSP. Brandstein's MC has a similar structure as Ikeda's DSP because both include adaptive filters that receive feedback from a noise-reduced output signal, and include a difference module that subtracts the filtered signal (a noise signal) from a speech-and-noise signal. As shown below, the BM's output passes through the MC's adaptive filter, much like how Ikeda's reference signal V_r passes through the DSP's adaptive filter 7 (*compare* Ex. 1005, Fig. 5.1's MC *with* Ex. 1008, Fig. 1's DSP):

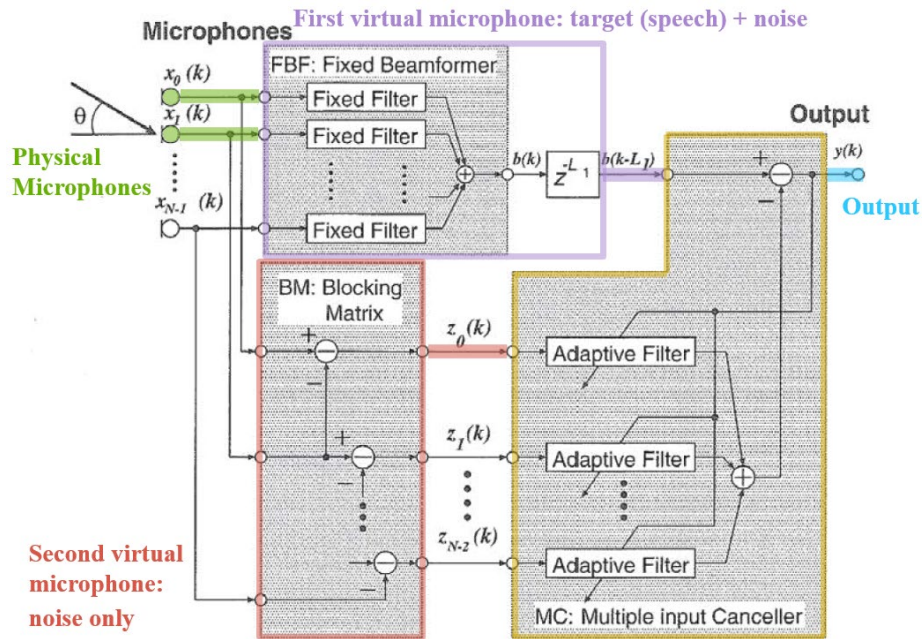
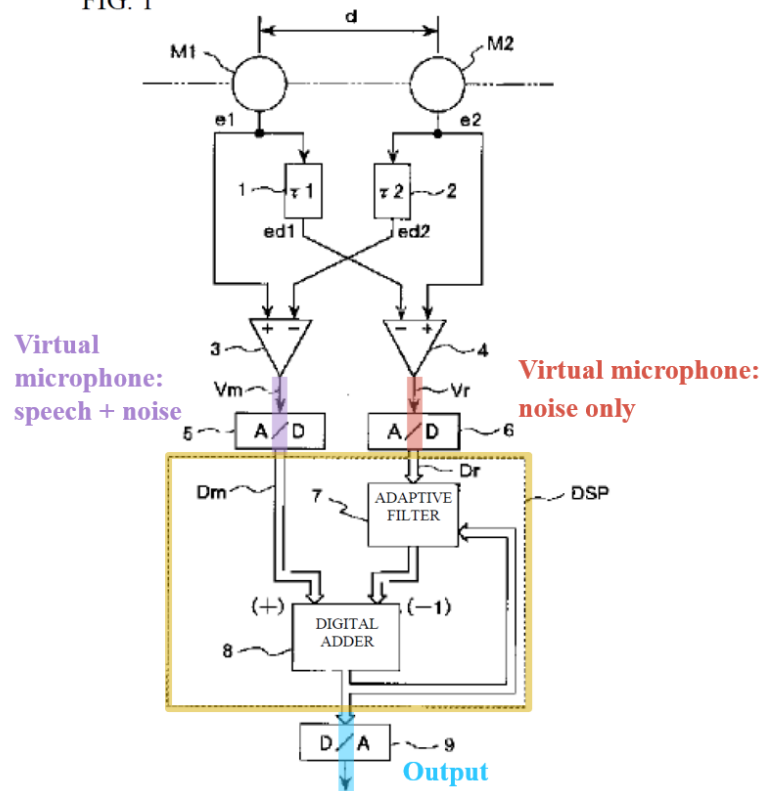


Fig. 5.1. Griffiths-Jim beamformer. It comprises a fixed beamformer (FBF), a multiple-input canceler (MC), and a blocking matrix (BM).

Ex. 1005, Fig. 5.1 (annotated, second mapping)

FIG. 1



EX1008, Fig. 1 (annotated).

76. In Brandstein, the output of the MC's summer is subtracted from the FBF output to produce overall output $y(k)$, which is provided as feedback to the MC's adaptive filter. Likewise, Ikeda's DSP subtracts the output of the adaptive filter 7 from the main signal V_m (which, like the FBF signal of the GSC, contains speech and noise) to produce an overall output that is fed back into the adaptive filter 7. Given these structural similarities, a POSITA would have been motivated to apply the specific teachings of Ikeda's DSP to Brandstein's MC.

77. In my opinion, for the reasons I discuss above, a POSITA would have known that implementing Brandstein's GSC to include Ikeda's DSP would have further advanced Brandstein's goal of removing undesirable signals such as noise and reverberation.

2. Reasonable Expectation of Success

a) Reasonable Expectation of Success in Combining Teachings of Brandstein's Embodiments

78. In my opinion, a POSITA would have reasonably expected to successfully apply Brandstein's teachings regarding the frequency insensitivity of the proposed GJBF to the simple GJBF. In my opinion, a POSITA would have recognized that the proposed GJBF operates on the same principles as the simple GJBF (*e.g.*, subtracting the noise-only output of a second virtual microphone from the speech-and-noise output of a first virtual microphone, to produce an isolated speech signal as the output).

79. In my opinion, the POSITA would have appreciated that, for purposes of isolating the speech signal at the output, the top and bottom branches of both the simple and proposed GJBFs would have maintained the same relevant characteristics, such as those depicted in the Figs. 5.2 and 5.3 directivity plots, including the directivity plots of the final output, FBF, BM, top and bottom branches. Specifically, as shown in Figs. 5.1 and 5.8, the POSITA would have understood that the FBF is

the same between the simple and proposed GJBFs (Ex. 1005, 89, 98), such that the response of the top branch (containing both speech and noise) is the same in both.

80. Moreover, despite the structure of bottom branches being different between the simple and proposed GJBFs, the POSITA would have understood that the responses of the bottom branches would have maintained a single null in the speech response toward the direction of speech, as Brandstein depicts for the BM in Fig. 5.3.

81. Brandstein explicitly states that the BM in the GJBF-CNA “provides a wider null for the target with sharper edges.” Ex. 1005, 97. Besides widening the null, the POSITA would have understood that the GJBF-CNA bottom branch does not otherwise modify the overall shape and characteristic of the response pattern from the BM response shown in Fig. 5.3. This is because the added adaptive filters in the GJBF-CNA bottom branch provide improved tracking to movements of the speech source, but do not change the overall goal of the GJBF to cancel the noise. To cancel the noise, the noise-only signal of the bottom branch must null out the speech, so that the overall shape and characteristics of the BM apply to the bottom branch of the proposed GJBF.

82. Thus, because the entire purpose of the modifications to the bottom branch of the GJBF is to make it more robust to steering vector errors toward the

direction of the target without changing the response of the BM elsewhere, and so a POSITA would understand that the BM's response would maintain relevant characteristics, such as having a single null in the speech response toward the direction of speech.

83. For these reasons, it is my opinion that a POSITA would have had a reasonable expectation of success in applying the teachings of Brandstein's GJBF-CNA to Brandstein's basic GJBF.

**b) Reasonable Expectation of Success in Combining
Brandstein and Ikeda**

84. It is my opinion that a POSITA would have expected to be able to successfully implement Ikeda's DSP in Brandstein's GSC. A POSITA would have recognized that Ikeda's DSP would be compatible with Brandstein's GSC. Ikeda explains that the main virtual microphone V_m , like the top branch of the GJBF, "has directional characteristics for picking up a target sound and noise from the front surface side," while the reference virtual microphone V_r , like the bottom branch of the GJBF, "has directional characteristics for picking up noise from the back surface side without picking up the target sound and noise from the front surface side." Ex. 1008, ¶¶[0021], [0023], ¶¶[0012], [0026], Fig. 10, Abstract.

85. As discussed above, one input for Ikeda's DSP is the main microphone signal V_m , which contains both the target sound and noise. Ex. 1008, ¶¶[0021].

Declaration of Dr. Michael Johnson in Support of
Petition for *Inter Partes* Review of USP No. 10,779,080

Specifically, Ikeda teaches that “[t]he main microphone has directional characteristics on the front surface side, and therefore had directional characteristics for picking up a target sound and noise from the front surface side.” Ex. 1008, ¶¶[0021]. Like Ikeda’s main microphone signal, the top branch of the GJBF also contains both the target sound (speech) and noise. The other input for Ikeda’s DSP is the reference microphone which primarily contains noise. Ex. 1008, ¶¶[0023] Ikeda states that “[t]he reference microphone has directional characteristics on the back surface side, and therefore has directional characteristics for picking up noise from the back surface side without picking up the target sound and noise from the front surface side.” Like Ikeda’s reference microphone signal, the bottom branch of the GJBF contains primarily noise signals. Therefore, a POSITA would have known that the virtual microphone signals of Brandstein are akin to the virtual microphone signals received by the DSP of Ikeda. As the POSITA would have recognized, processing of noise-only or speech-and-noise signals for further noise reduction or noise cancellation from Ikeda would have been readily applied to Brandstein. For these reasons, in my opinion, a POSITA would have had a reasonable expectation of success in applying Ikeda’s DSP of Ikeda to Brandstein’s GJBF.

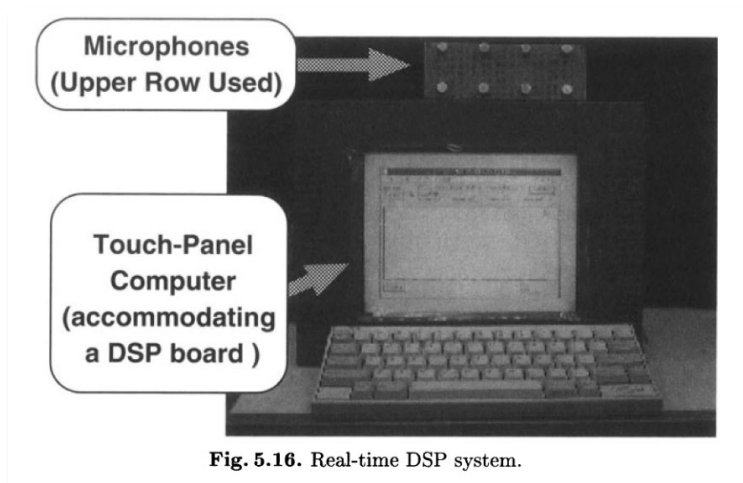
3. **Claim 7**

86. In order to simplify my analysis, I start with independent claim 7, and then address independent claims 14 and 1.

a) **[7pre] “A system, comprising:”**

87. To the extent the preamble is limiting, it is my opinion that Brandstein discloses this limitation.

88. Brandstein teaches a system with a microphone array and signal processing that implements the GSC. Ex. 1005, 89, Figs. 5.1, 5.8, 5.16. For example, Brandstein discloses the use of beamforming in fields such as hands-free telephones and hearing aids. Ex. 1005, 87. As another example of how Fig. 5.1’s basic GSC may be implemented, Brandstein reports the results of experiments performed with a real-world implementation of a GSC on a “Real-time DSP system,” which is illustrated in Fig. 5.16:



Ex. 1005, Fig. 5.16 (annotations in original)

Ex. 1005, 105. Brandstein describes a “system compris[ing] a microphone array and a compact touch-panel personal computer which includes a floating point DSP, the ADSP-21062.” Ex. 1005, 104 (internal citation omitted). Thus, it is my opinion that Brandstein discloses a system. In my opinion, in the alternative, a POSITA would have known that Brandstein’s basic GSC in Fig. 5.1 could have been implemented in a system for use in telephones or hearing aids (as Brandstein teaches), and therefore it would have been obvious.

- b) **[7a] “a first virtual microphone formed from a first combination of a first microphone signal and a second microphone signal, wherein the first microphone signal is generated by a first physical microphone and the second microphone signal is generated by a second physical microphone;”**

89. In my opinion, Brandstein discloses this limitation.

90. Fig. 5.1 shows the output $b(k-L_1)$ of the basic GJBF’s top branch (including the FBF and a delay module, together corresponding to the first virtual microphone), which is the delayed summation of filtered **first and second physical microphones** generating **first and second microphone signals** (e.g., x_0 and x_1), respectively. These signals form a **first virtual microphone** along the top branch of the GSC. Because the filtered signals are summed, the first virtual microphone is

formed by combining the first and second microphone signals. The first virtual microphone is formed from a first combination of the first and second microphone signals:

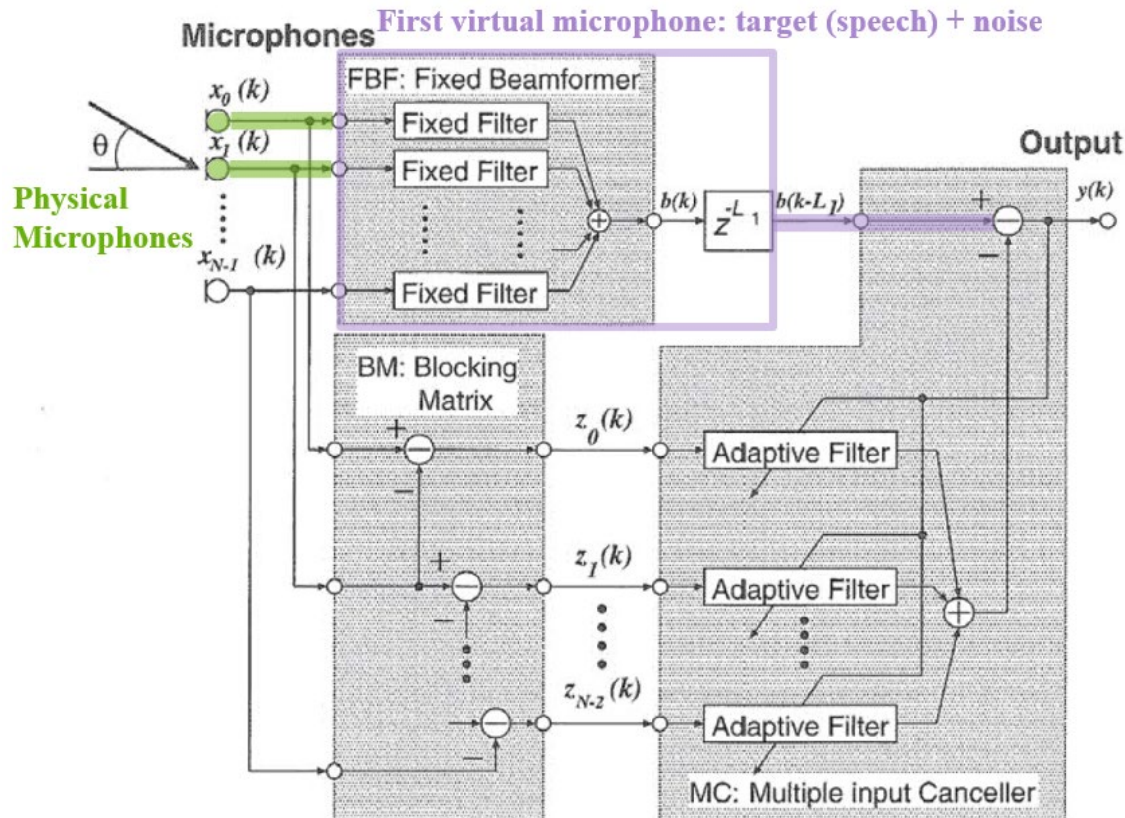


Fig. 5.1. Griffiths-Jim beamformer. It comprises a fixed beamformer (FBF), a multiple-input canceller (MC), and a blocking matrix (BM).

Ex. 1005, Fig. 5.1 (annotated)

Ex. 1005, 89. Brandstein further discloses that the physical microphones may be omnidirectional microphones. Ex. 1005, 101.

- c) [7b] “a second virtual microphone formed from a second combination of the first microphone signal and

the second microphone signal, wherein the second combination is different from the first combination,”

91. In my opinion, Brandstein discloses this limitation.

92. As explained above, for the *second virtual microphone*, I have considered two alternative mappings as explained below.

93. In the first mapping, *the second virtual microphone* signal is the output of the bottom branch (from the MC's summer), before being subtracted from the top branch's output. As shown in Fig. 5.1, a *second virtual microphone* along the bottom branch of the GSC is *formed from* a *second combination* of the *first and second microphone signals* (e.g., x_0 and x_1) that *is different from the first combination*:

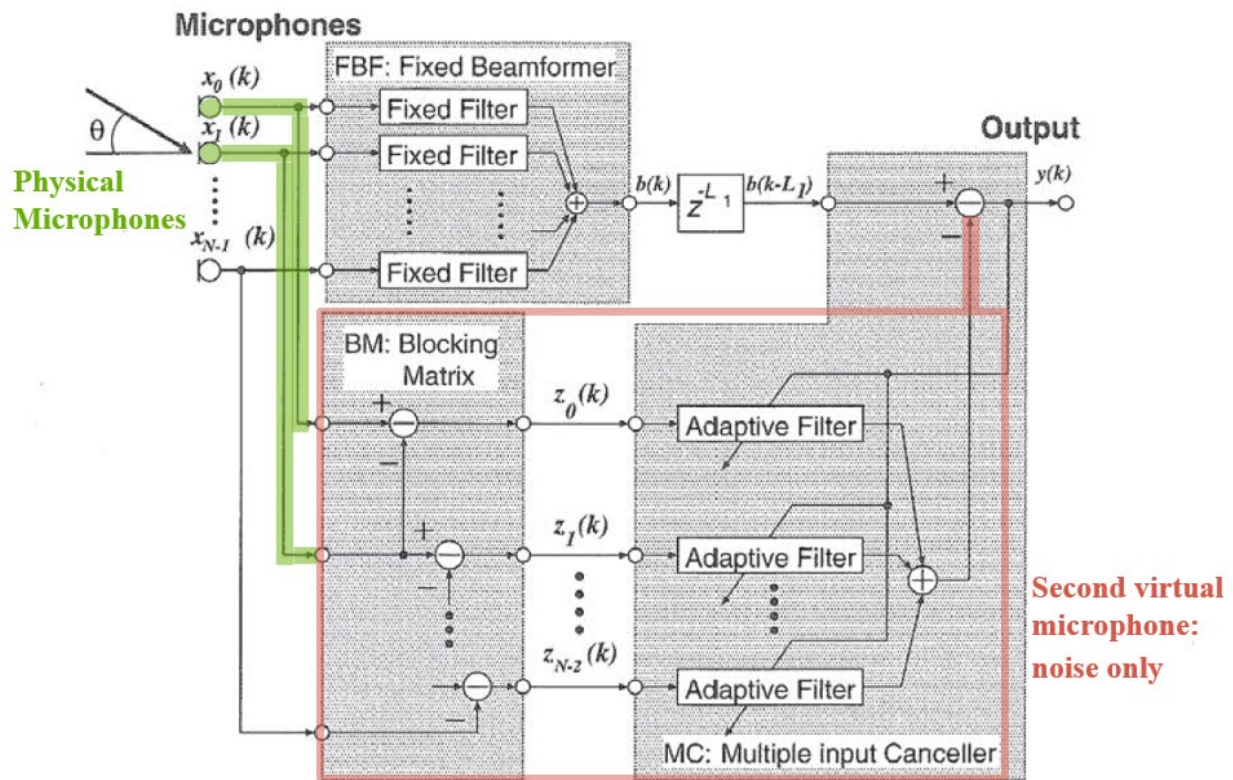


Fig. 5.1. Griffiths-Jim beamformer. It comprises a fixed beamformer (FBF), a multiple-input canceler (MC), and a blocking matrix (BM).

First Mapping, Ex. 1005, Fig. 5.1 (annotated)

94. In the second mapping, the *second virtual microphone* signal is the output from the BM. Here, the *second virtual microphone* is the BM, and is formed from a *second combination* of the *first and second microphone signals* (x_0 and x_1) that is *different from the first combination*:

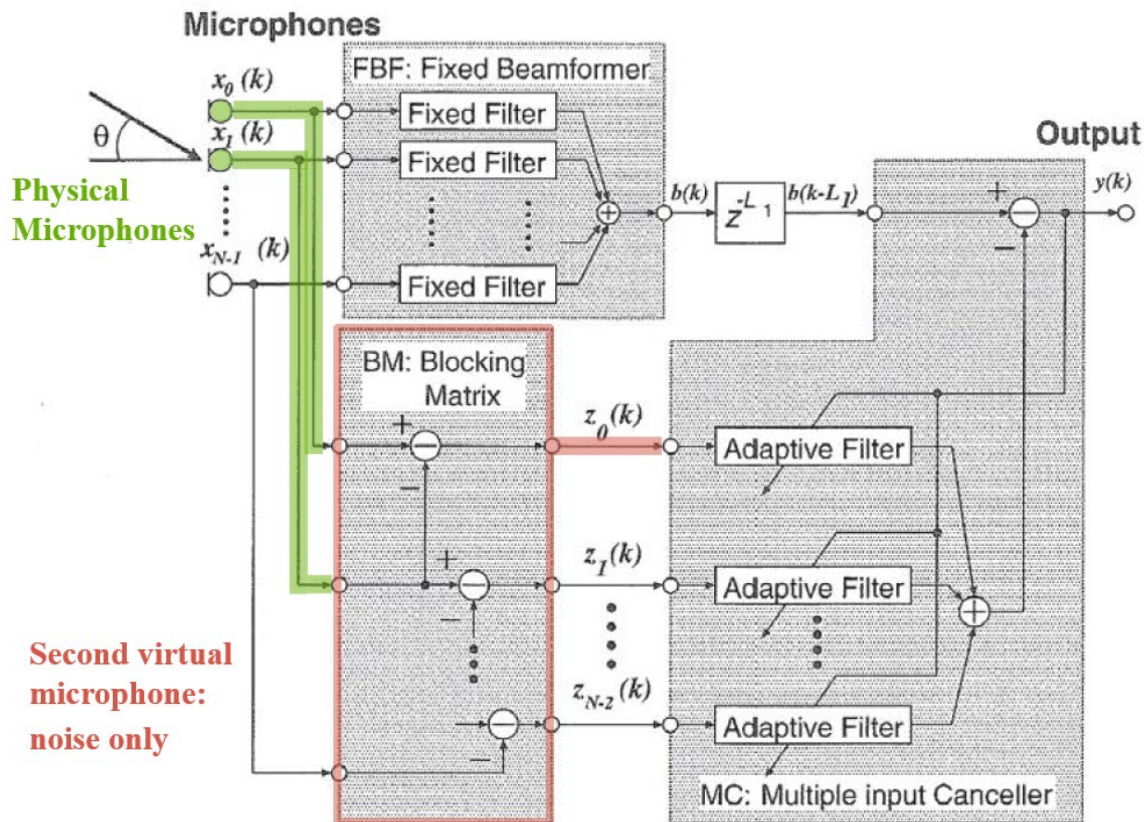


Fig. 5.1. Griffiths-Jim beamformer. It comprises a fixed beamformer (FBF), a multiple-input canceller (MC), and a blocking matrix (BM).

Second Mapping, Ex. 1005, Fig. 5.1 (annotated)

When there are only two physical microphones in the GSC, there is only one output, $z_0(k)$, from the BM.

- d) [7c] “wherein the first virtual microphone has a first linear response to speech and first linear response to noise,”

95. In my opinion, Brandstein discloses this limitation.

96. Brandstein discloses that the *first and second virtual microphones* (top and bottom branches of the GSC, respectively) output separate signals. As discussed

in Section VIII.A.3.b, the top branch of the GSC includes an FBF and corresponds to the *first virtual microphone*. The target signal is speech, which is shown at 0 degrees, while the interference signal, which is noise, is shown at approximately 45 degrees.

97. The '080 patent defines “noise” as “unwanted environmental acoustic noise.” Ex. 1001, 4:29-30. A POSITA would understand that Brandstein refers to environmental acoustic noise as “interference” because Brandstein states that interference can be, for example, air-conditioner noise or white noise. Ex. 1005, 101, 103, Fig. 5.13.

98. Brandstein’s Fig. 5.3 shows an example directivity pattern of an FBF and a BM:

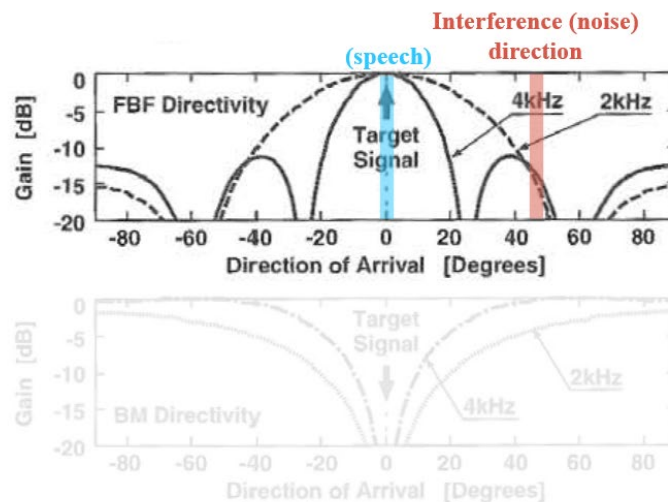


Fig. 5.3. Directivity pattern of a fixed beamformer (FBF) and a blocking matrix (BM).

Ex. 1005, Fig. 5.3 (annotated)

99. As shown in Fig. 5.3, the FBF of the top branch (*first virtual microphone*) has *a first linear response to speech*, which is in the direction of a “*target signal*,” i.e. *speech*. The top branch of the GSC also has a *first linear response to interference*, or *noise*. As I explain above in Section VI.C, Brandstein subtracts the output of the GSC’s bottom branch from the output of the top branch. Fig. 5.2 shows the directivity pattern for the resulting output of the example GJBF:

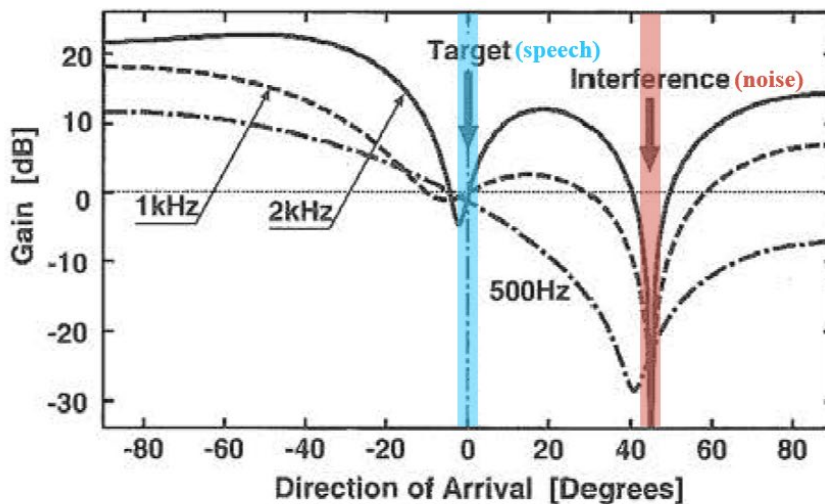


Fig. 5.2. Example directivity pattern of the Griffiths-Jim beamformer.

Ex. 1005, Fig. 5.2 (annotated)

The horizontal axis of the graph shows direction of arrival measured in degrees relative to the microphone array. The target signal, which, is speech, is shown at 0 degrees, while the interference signal, which is noise, is shown at approximately 45 degrees. According to Brandstein, Fig. 5.2 shows that “in the direction of the

interference, a deep null is formed.” Ex. 1005, 90. This “deep null” in the direction of the interference, or noise, is formed by subtracting the linear response of the bottom branch in the direction of noise from the linear response of the top branch (*first virtual microphone*) in the direction of noise. Therefore, the *first virtual microphone* of the GSC has a *first response to noise*.

100. The '080 patent does not explain what a “linear response” is. To the extent the response is required to be linear, a POSITA would have recognized that the *responses* to speech and noise of Brandstein’s first virtual microphone’s are *linear* because the responses are represented by a linear system. For example, Brandstein’s Fig. 5.1 depicts the FBF as applying “fixed filter[s]” that are summed together, which are all *linear* components. Ex. 1005, 95, Figs. 5.1, 5.4.

- e) **[7d] “the first linear response to speech being substantially similar across a plurality of frequencies for a speech source located within a predetermined angle relative to an axis of the microphone array and devoid of a null,”**

101. In my opinion, Brandstein discloses or renders obvious this limitation.

102. As I explained in Section VIII.A.1, a POSITA would have been motivated to apply the teachings of Brandstein’s proposed GJBF to Brandstein’s simple GJBF. For the simple GJBF, Brandstein depicts the overall shapes of the directivity patterns for a limited set of frequencies for the FBF (of top branch) and BM (of

bottom branch, with sharp null) (*see* Fig. 5.3) as well as for final output (*e.g.*, the difference between top and bottom branches (*see* Fig. 5.2). This proposed modified structure, which Brandstein calls the “GJBF-CNA” is shown in Fig. 5.8.

103. In my opinion, a POSITA would have recognized that the GJBF-CNA operates on the same principles as the GSC; that is, it subtracts the noise-only output of the second virtual microphone from the speech-and-noise output of the first virtual microphone. A POSITA would have further known that the directivity pattern of the final output, FBF and BM would have the same features as depicted in the Figs. 5.2 and 5.3 directivity plots. As shown in Fig. 5.8, the FBF is unchanged in the GJBF-CNA. Ex. 1005, 98. Therefore, the response of the FBF in the GJBF-CNA would be identical to the FBF in the basic GSC, and the FBF output would still contain speech-and-noise.

104. Fig. 5.10 of Brandstein is a plot of “the sensitivities to the frequency component of the target signal” (*first linear response to speech across a plurality of frequencies*) at 500, 1000, 2000, 3000, and 3500 Hz for the modified GSC of Fig. 5.8:

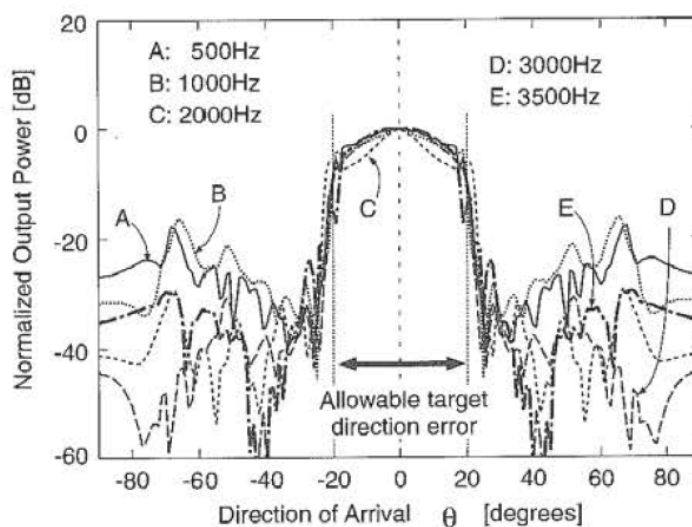


Fig. 5.10. Sensitivities after convergence as a function of DOA at different frequencies.

105. Brandstein explains that the “[f]requency dependency of the GJBF-CNA is small” and that the sensitivity in the region around the target direction (i.e. direction of speech) “is constant.” Ex. 1005, 99-100. In my opinion, the POSITA would have understood Fig. 5.10 depicts the output $y(k)$ of Fig. 5.8, as Brandstein describes the plot in relation to “the GJBF-CNA.” Ex. 1005, 100. Moreover, because the output of Brandstein’s bottom branch in Fig. 5.8 results from “a wider null for the target [speech] with sharper edges,” the POSITA would have further understood that the Fig. 5.10 plot in the target *speech* direction would have also been the top branch’s response in the same *speech* direction, at different *frequencies*. For these reasons, the top branch’s response to *speech* of the GJBF-CNA (*first response*) is constant (which meets *substantially similar*) across a plurality frequencies.

106. Fig. 5.10, which is a plot of the sensitivity to the frequency component of the target signal (*first linear response to speech across a plurality of frequencies*), shows that the sensitivity is constant within the “high sensitivity regions” across different frequencies for a range of angles $-20^\circ < \theta < 20^\circ$. Here, θ is *an angle relative to an axis of the microphone array* as shown in Fig. 5.1. Ex. 1005, 100. In my opinion, a POSITA would have understood that the “maximum allowable target-direction error” refers to a pre-set threshold, which means it is *a predetermined angle* defining a range of angles *within which a speech source is located*.

107. There is also no *null* in the target (speech) direction in the sensitivity plot of Fig. 5.10 because there is no zero or minima in the spatial response (i.e. directivity pattern) in that direction. (Ex. 1001, 4:31-32 (defining “null” as “a zero or minima in the spatial response of a physical or virtual directional microphone.”)) Therefore, Brandstein discloses a *first linear response to speech* that is *devoid of a null*.

108. Alternatively, a POSITA would have understood that the top branch’s directivity pattern (*first linear response to speech*) is *devoid of a null* in all directions for certain frequencies. A POSITA would have seen from the top directivity plot in Fig. 5.3 that the largest maximum in the directivity is centered around the direction of speech (DOA of 0°) and would have understood that the width of this maximum

depends on frequency. Specifically, a POSITA would have seen that the width of the maximum approximately doubles from approximately 50° to 100° when the frequency is approximately halved from 4 kHz to 2 kHz. Therefore, a POSITA would have recognized that for a frequency of below about 1000 Hz, the main lobe's width would span an entire range of directions plotted in Fig. 5.3, such that the *response is devoid of a null* in any direction covered by the plot. Moreover, even for the directions not depicted in the top directivity plot in Fig. 5.3, it is my opinion that a POSITA would have understood, based on these same observations, that the directivity pattern (*first linear response to speech*) of the FBF is *devoid of a null* in all directions.

- f) **[7e] “wherein the second virtual microphone has a second linear response to speech that has a single null oriented in a direction toward a source of the speech and a second linear response to noise,”**

109. In my opinion, Brandstein discloses this limitation for either mapping (*i.e.*, whether the *second virtual microphone* is the output of the BM alone or the entire bottom branch, up to and including the MC's summer).

110. Furthermore, it is my opinion that Brandstein meets this limitation under both the first and second mapping (*i.e.*, whether the *second virtual microphone* is the output of the BM alone or with the MC). As Brandstein depicts in Fig. 5.3,

the BM of the bottom branch (*second virtual microphone*) has *a second linear response to speech*, which is in the direction of a “target signal,” i.e. *speech*:

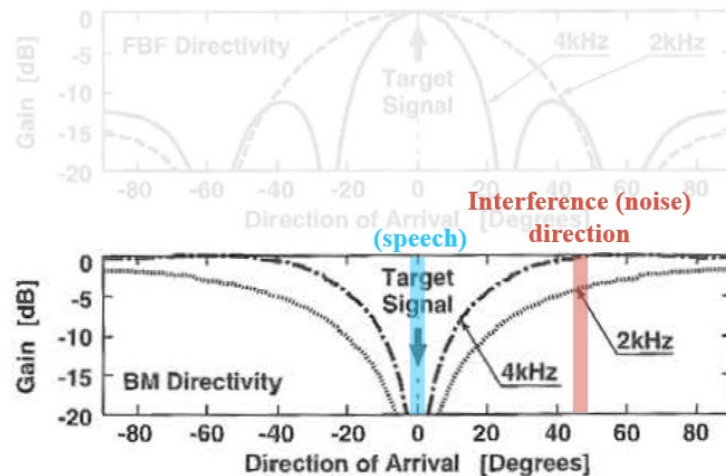


Fig. 5.3. Directivity pattern of a fixed beamformer (FBF) and a blocking matrix (BM).

Ex. 1005, Fig. 5.3 (annotated)

111. The general characteristics and overall shape of the directivity pattern of the BM are the same as that of the *second virtual microphone* because the purpose of both is to isolate the *noise* for subtraction at the output. For the same reasons as discussed above with respect to [7c], the bottom branch of the GSC (*second virtual microphone*) also has *a second response to noise* at the approximately 45-degree direction of interference. Furthermore, as shown in Fig. 5.3, the bottom branch of the GSC has *a single null oriented in the direction toward the target signal (a source of speech)*.

112. Additionally, if the *response to speech* is not limited to the direction of the target *speech* and includes a wider range of directions, a POSITA would have understood the response of the bottom branch (*second linear response to speech*) has a *single null oriented in a direction toward a source of the speech* across different directions. First, the bottom plot of Figure 5.3 depicts on a single null for the directions ranging from -90 to 90 degrees.

113. Additionally, although Brandstein's Figure 5.3 does not expressly show the response for a full 360 degrees, a POSITA would have known that the response of the BM has a *single null oriented in a direction toward a source of the speech* for certain speech directions. For example, this is the case if the speech direction is parallel to the axis of the microphone array. The POSITA's understanding is reflected in, for example, Figs. 3 and 4a of U.S. Patent No. 7,386,135 ("Fan," Ex. 1022.

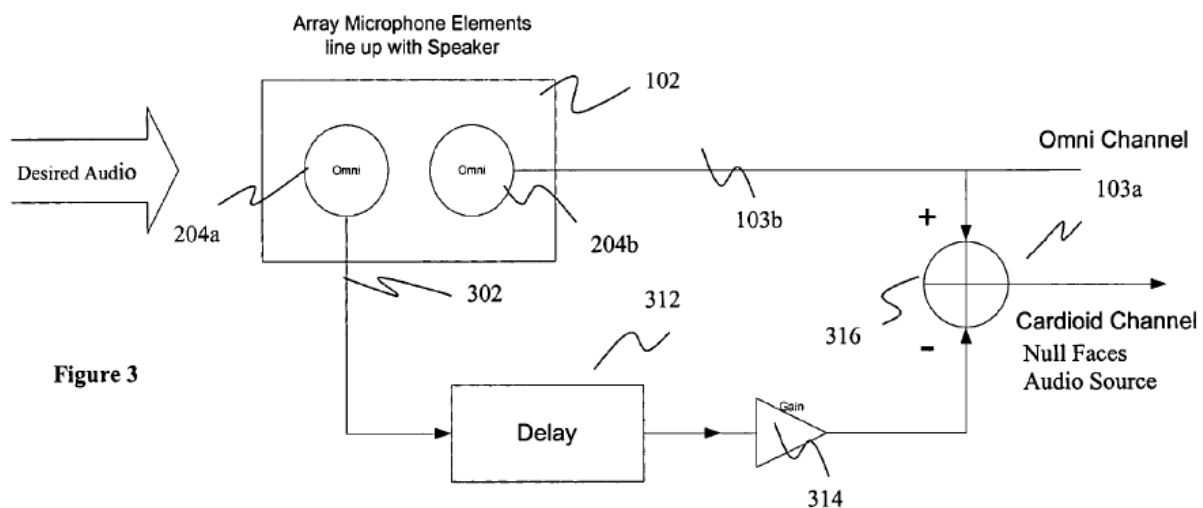
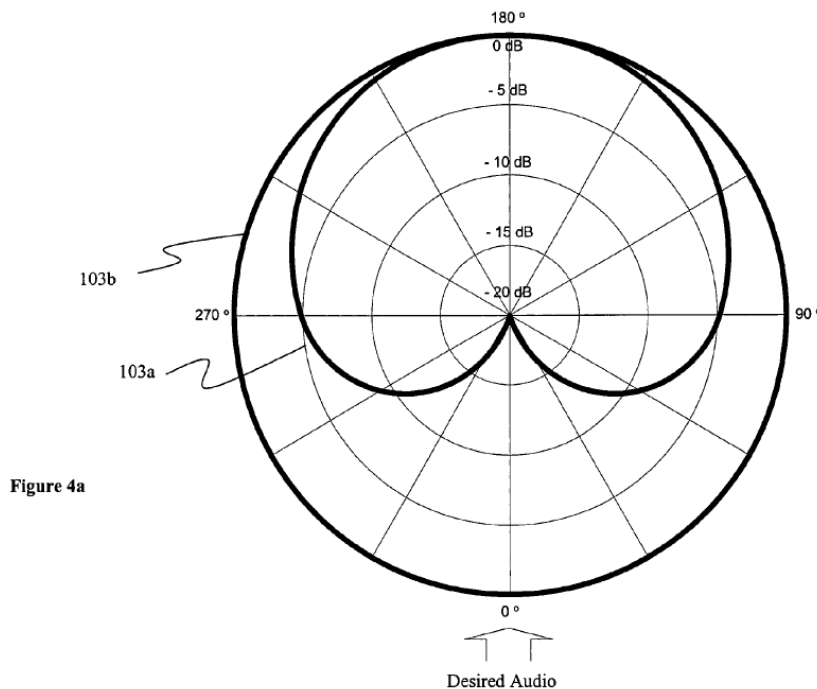


Fig. 3 depicts a delay-and-subtract structure like that of the GJBF's BM. The signal from the omnidirectional physical microphone 204a is delayed and then subtracted from the signal from the other omnidirectional physical microphone 204b.

114. Fig. 4a “illustrates the acoustic beams generated by the various acoustic devices of FIG. 2a-2c, and FIG.2g (when supplemented with the circuitry of FIG.3) in the form of a polar sensitivity plots.” Fig. 2g depicts the 2-omindirectional microphone array 102. Fig. 4a, reproduced below, shows that the response of Fig. 3's delay-and-subtract structure has *a single null oriented in a direction toward the “desired audio” (source of the speech)*. Ex. 1022, 2:16-23 (emphasis added), Fig. 4a.



115. The '080 patent does not explain what a “linear response” is. To the extent the response is required to be linear, a POSITA would understand that the second virtual microphone’s response to the speech and response to the noise are *linear responses* because the bottom branch is represented by difference modules, adaptive filters, and a summer, which are all *linear* components.

116. Although the bottom directivity pattern of Fig. 5.3 is for that of the BM, a POSITA would have understood that the directivity pattern of the bottom branch after filtering would have resulted in a directivity pattern with the same overall characteristics. In particular, the noise-only signal provided to the final difference module would have also had a null in the speech direction and higher responses elsewhere, as it has a “wider null... with sharper edges.” Ex. 1005, 97. Thus, just like the output of the BM, the bottom branch’s output (*second linear response*) would have also had *a single null oriented in a direction toward a source of the speech as well as a second linear response to noise*. Therefore, the teachings of Brandstein

with respect to the *second linear response to speech* and noise are applicable under both the first and second mappings.³

- g) [7f] “wherein the second linear response to noise is substantially similar to the first linear response to noise,” and [7h] “the second linear response to speech is substantially dissimilar to the first linear response to speech,”

117. In my opinion, Brandstein discloses or renders obvious these limitations.

118. In my opinion, a POSITA would have known that, for both mappings, Brandstein’s virtual microphones have very different (meeting the *substantially dissimilar* limitation) *responses* to the target signal because the top branch (*first virtual microphone*) is designed to form a beam that captures the target signal, while the

³ The overall characteristics of the directivity patterns for the BM output (first mapping for *second virtual microphone*) and the bottom branch’s output (second mapping for *second virtual microphone*) are the same. Thus, a mapping for the *second virtual microphone* or the *second linear response to speech* or *noise* to the BM output would also apply to the bottom branch’s output, and vice versa. For this reason, I do not repeat an explanation for both mappings for every claim limitation, and for such claim limitations, my explanation applies to both mappings.

second virtual microphone (in both mappings) includes a BM designed to block the target signal. Ex. 1005, 88. Brandstein discloses that the target signal is *speech*. For example, Brandstein states that, for a particular embodiment, the target source was an English male speech signal. Ex. 1005, 101; *see also* Ex. 1005, 103, Figs. 5.13, 5.14. Indeed, just before noting that the GSC was the most widely known adaptive beamformer, Brandstein states: “Applications of beamforming include microphone arrays for speech enhancement. The goal of speech enhancement is to remove undesirable signals such as noise and reverberation.” Ex. 1005, 87.

119. A POSITA would have further known that the two virtual microphones in the GSC also have essentially the same (meeting the *substantially similar* limitation) *responses to noise* because the final subtraction output the noise cancels out. The different responses to the target signal and similar responses to noise are evident from Fig. 5.2, which shows the directivity pattern for the final output of an example GJBF, and Fig. 5.3, which shows the directivity patterns of the FBF and BM:

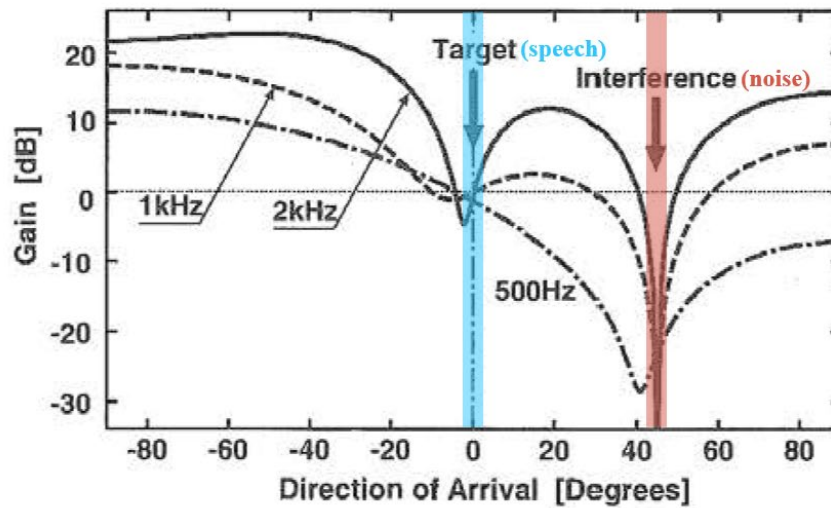


Fig. 5.2. Example directivity pattern of the Griffiths-Jim beamformer.

Ex. 1005, Fig. 5.2 (annotated)

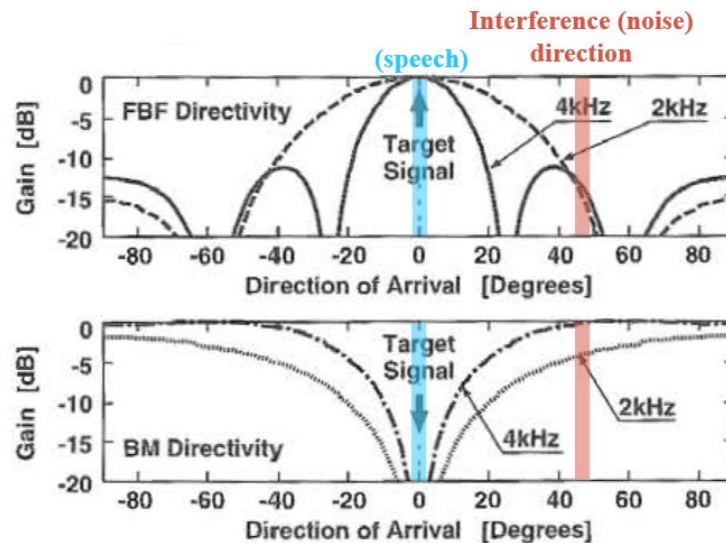


Fig. 5.3. Directivity pattern of a fixed beamformer (FBF) and a blocking matrix (BM).

Ex. 1005, Fig. 5.3 (annotated)

120. Brandstein states that in the direction of the target signal (highlighted in blue), the GJBF achieves “almost constant gains close to 0 dB.” Ex. 1005, 90. A POSITA would have known that a gain of zero dB means that subtracting the bottom branch’s noise-only output from the first virtual microphone’s speech-and-noise output would produce the target signal in the direction of the target, thereby showing that the two virtual microphones have very different responses to the target signal. Therefore, in my opinion this shows that *the second linear response to speech is substantially dissimilar to the first linear response to speech*. This is also evident in Fig. 5.3, which shows that the FBF’s response in the direction of speech has a maximum making it very different from the BM’s response in the direction of speech having a null.

121. In contrast, Brandstein states that “in the direction of the interference, a deep null is formed,” as highlighted in red above. Ex. 1005, 90. A POSITA would have known that the deep null reflects that subtracting the two virtual microphones’ outputs cancels the interference, and confirms that the virtual microphones’ responses to the interference are essentially the same. Therefore, in my opinion, this shows that *the second linear response to noise is substantially similar to the first linear response to noise*.

122. As I explain above, a POSITA would have understood that the teachings of Brandstein with respect to the second linear response to speech and noise are applicable under both first and second mappings because the directivity pattern of the bottom branch post-filter would have had the same overall characteristics as that of the bottom plot of Fig. 5.3.

h) [7g] **“one or both of the first linear response to noise and the second linear response to noise being nonzero in a direction toward a source of noise, and”**

123. In my opinion, Brandstein discloses or renders obvious this limitation.

124. Brandstein’s Fig. 5.3 shows that, for either mapping, the only null in the directivity pattern of the BM of the bottom branch of the GJBF (*second virtual microphone*) is in the direction of speech. In other words, the response of the BM or bottom branch (*second linear response*) in all other directions is nonzero, including in a direction *toward the source of noise*. Therefore, at least *the second linear response to noise is nonzero in a direction toward the source of noise*.

i) [7i] **“wherein the speech is human speech; and”**

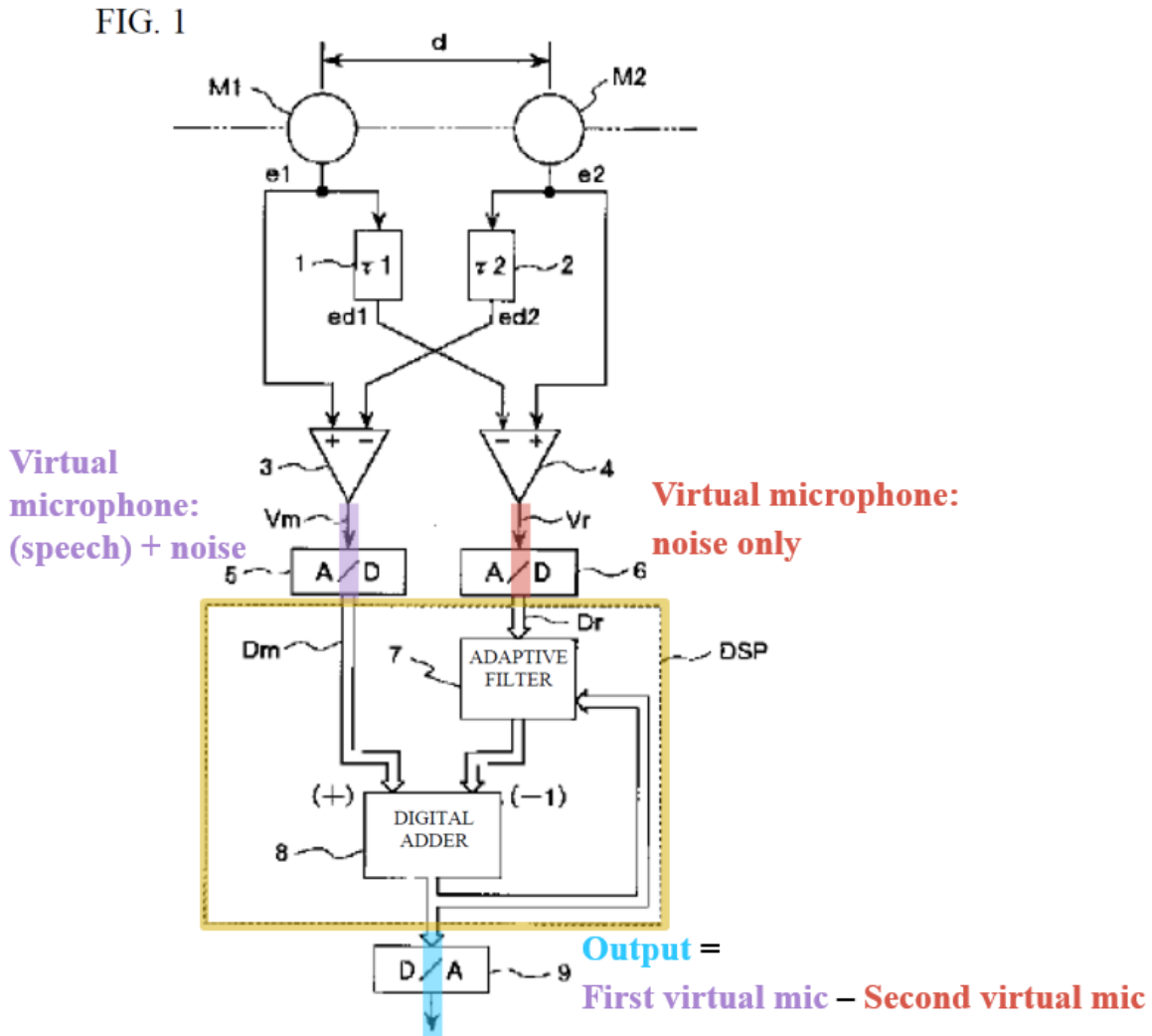
125. In my opinion, Brandstein discloses this limitation.

126. Brandstein discloses that, simultaneously, the target speech signal can be *human speech*, like “an English male speech signal.” *E.g.*, Ex. 1005, 101.

- j) [7j] **“an adaptive noise removal application coupled to the first and second virtual microphones and generating denoised output signals”**

127. In my opinion, Brandstein in combination with Ikeda discloses or renders obvious this limitation.

128. Ikeda discloses an adaptive noise removal system that includes an adaptive filter 7 and a digital adder 8 as shown in Fig. 1 (annotated below). Ex. 1008, ¶¶[0027]-[0029]. In my opinion, Ikeda’s **adaptive noise removal system** is an *application* that can be implemented using a DSP. Ex. 1008, ¶¶[0016]-[0017], [0030].



Ex. 1008, Fig. 1 (annotated).

129. Ikeda's *adaptive noise removal application* receives main signal Vm (analogous to Brandstein's *first virtual microphone*) and reference signal Vr (analogous to *second virtual microphone*),⁴ and is therefore coupled to the first and second virtual microphones via A/D converters 5 and 6. Ex. 1008, ¶¶[0027]-[0029], Fig. 1; Ex. 1001, 32:25-32.

130. Ikeda's *adaptive noise removal application* generates denoised output signals. Ex. 1008, ¶¶[0028]-[0029]. Ikeda discloses that "the noise component contained in the main signal Vm is cancelled by the noise component contained in the reference signal Vr due to subtraction by digital adder 8," thereby generating denoised output signals. Ex. 1008, ¶¶[0028]-[0029].

131. Under both first and second mappings, it would have been obvious to couple the *first virtual microphone* and the *second virtual microphone* Ikeda's DSP (*adaptive noise removal application*).

⁴ As discussed above in Section VIII.A.1.b, the second virtual microphone signal may be the signal output from the MC (first mapping) or the signal output from the BM (second mapping). The teachings of Ikeda are equally applicable to either mapping.

132. For the first mapping, it is my opinion that a POSITA would have found it obvious to couple Ikeda's DSP to the output of the top branch and the MC's summer where the subtractor is below:

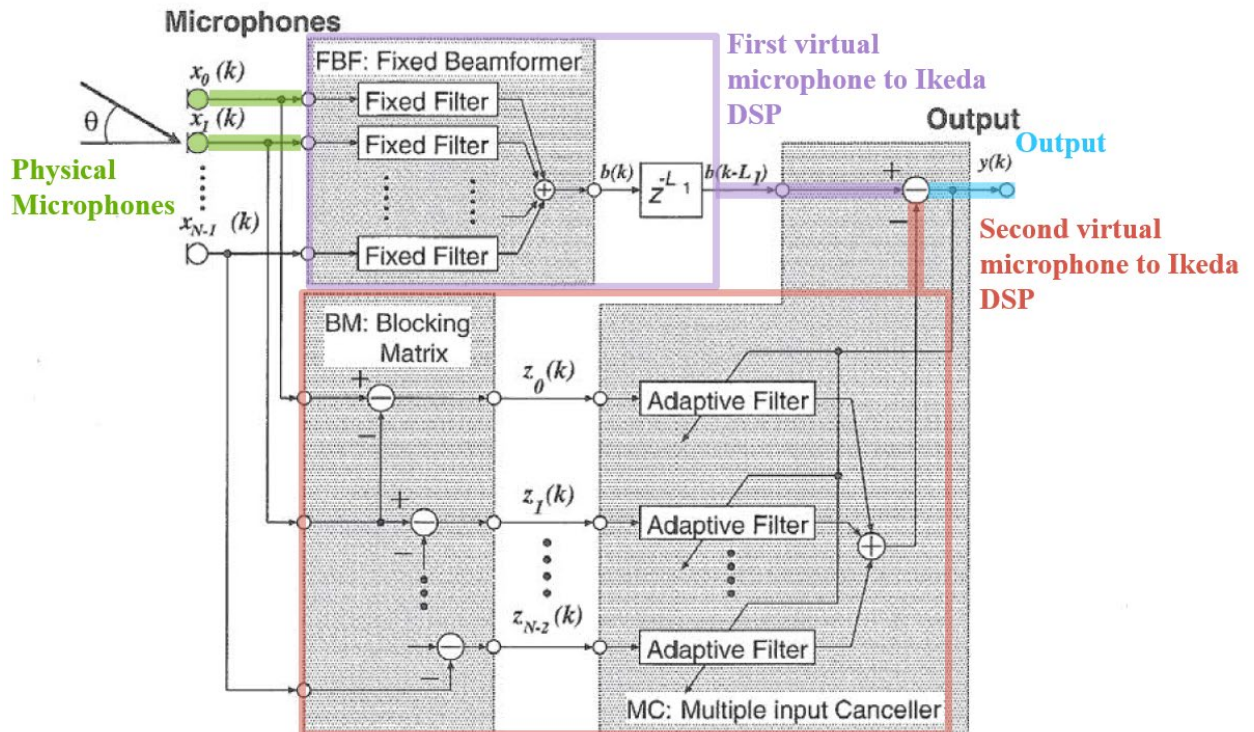


Fig. 5.1. Griffiths-Jim beamformer. It comprises a fixed beamformer (FBF), a multiple-input canceller (MC), and a blocking matrix (BM).

Ex. 1005, Fig. 5.1 (annotated)

133. For the second mapping, it is my opinion that a POSITA would have found it obvious couple Ikeda's DSP to the output of the top branch and the output of the BM, as shown below:

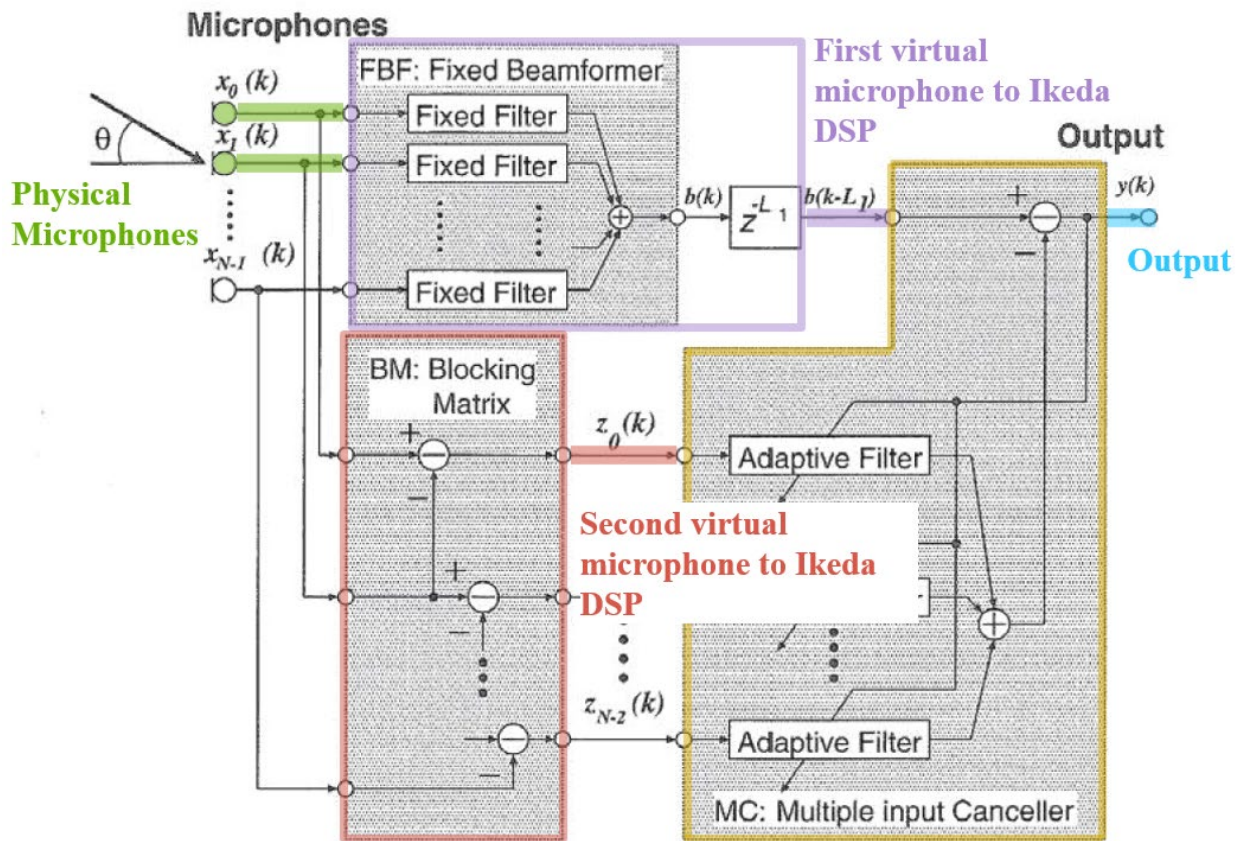


Fig. 5.1. Griffiths-Jim beamformer. It comprises a fixed beamformer (FBF), a multiple-input canceler (MC), and a blocking matrix (BM).

Ex. 1005, Fig. 5.1 (annotated)

As I explain above in VIII.A.1.b, Ikeda's DSP operates in a very similar way as Brandstein's MC.

134. In both mappings, the output of the top branch (*the first virtual microphone*) is input into Ikeda's DSP (*coupled to the adaptive noise removal application*). Ikeda's DSP then subtracts the output of the adaptive filter 7 from the main signal V_m (analogous to the top branch's output and containing speech-and-noise) to produce an output that is fed back into the adaptive filter 7.

- k) **[7k] “by forming a plurality of combinations of signals output from the first virtual microphone and the second virtual microphone,”**

135. In my opinion, Brandstein in combination with Ikeda discloses or renders obvious this limitation.

136. Under both first and second mappings, Ikeda’s adaptive noise removal application forms a *plurality of combinations of signals output from the first virtual microphone and the second virtual microphone* because main signal V_m and reference signal V_r are continuously converted to digital signals and input to the adaptive noise removal system implemented on a DSP so that, over time, they form multiple combinations of signals output from the first and second virtual microphones. Ex. 1008, ¶¶[0027]- [0029]. The ’080 patent explains that “[w]ords using the singular or plural number also include the plural or singular number respectively.” Ex. 1001, 32:61-63. This understanding is also consistent with the ’080 patent’s description that an adaptive noise removal system “combin[es] the two microphone signals (e.g., Mic1, Mic2) by filtering and summing in the time domain.” Ex. 1001, 7:4-16, 8:50-9:3, 9:22-38.

- l) **[7l] “by filtering and summing the plurality of combinations of signals in the time domain, and”**

137. In my opinion, Brandstein in combination with Ikeda discloses or renders obvious this limitation.

138. Under both first and second mappings, Ikeda's adaptive noise removal system filters and sums the plurality of combinations of signals in the time domain because Ikeda discloses that digital versions of reference signal V_r are "supplied to an adaptive filter" (*filtering*) and the output of adaptive filter 7 and digital versions of main signal V_m are differenced at digital adder 8 (*summing*). Ex. 1008, ¶¶[0027]-[0029]. This is consistent with the '080 patent's description that an adaptive noise removal system "combin[es] the two microphone signals (e.g., Mic1, Mic2) by filtering and summing in the time domain." Ex. 1001, 7:4-16, 8:50-9:3, 9:22-38. In my opinion, a POSITA would have understood that adaptive filter 7 filters in the time domain, and digital adder 8 adds in the time domain for at least two reasons. First, Ikeda's system does not include components for converting signals from the time domain to the frequency domain or from the frequency domain to the time domain, and second, a signal picked up by a microphone is in the time domain if it is not processed by a component and converted to the frequency domain. Ex. 1008, ¶¶[0027]-[0029]; Ex. 1010, [0042]-[0043], [0058]-[0062]. Additionally, a POSITA would have found it obvious to implement Ikeda's system, including adaptive filter 7 and digital adder 8 to process signals, in the time domain.

m) **[7m] “by a varying linear transfer function between the plurality of combinations of signals,”**

139. In my opinion, Brandstein in combination with Ikeda discloses or renders obvious this limitation.

140. Under both first and second mappings, Ikeda’s adaptive noise removal system discloses a *varying transfer function between the plurality of combinations of signals* because Ikeda discloses that adaptive filter 7’s characteristics “are varied in real time” by “a tap coefficient updating operation for updating the tap coefficient based on an error signal obtained as the difference between the filter output and a signal from the target system to return the filter output,” which is successively repeated to rewrite the tap coefficient. Ex. 1008, ¶¶[0027]-[0029]. This “error signal is defined as the output of the digital adder 8,” which sums the plurality of combinations of signals. Ex. 1008, ¶[0028]. Ikeda discloses performing this tap coefficient updating operation when “the microphone units M1 and M2 do not pick up the target sound,” which is the time period in which the output of digital adder 8 should be zero (because there is no speech). Ex. 1008, ¶[0028]. By updating the tap coefficient over time to minimize this error signal, the adaptive filter calculates and applies a *varying transfer function between the plurality of combinations of signals* that results in the output of adaptive filter 7 equaling the digital signal from main signal Vm at a time when there is no speech. Ex. 1008, ¶[0028]. This is consistent with the ’080

patent's description that "[t]he adaptive filter generally uses the signal received from a first microphone ... to remove noise from the speech received from at least one other microphone ..., which relies on a slowly varying linear transfer function between the two microphones for sources of noise." Ex. 1001, 7:4-16, 5:19-44, 6:26-29, 8:50-9:3, 9:22-38.)

141. Ikeda also discloses that this *transfer function* is *linear* because it discloses that adaptive filter 7 is "a transversal filter" and "perform[s] an output operation for obtaining filter output by multiplying the input signal Dr by a tap coefficient and adding the value in multiple tap units." Ex. 1008, ¶¶[0027]-[0029]. A POSITA would have understood that this discloses a *linear* filter because similar filters have been described in the prior art as linear. Ex. 1013, 16-17, Fig. 2.2 ("Adaptive linear combiner in the form of single-input adaptive transversal filter"); Ex. 1014, 366-367. Alternatively, in my opinion, it would have been obvious to a POSITA to use a *linear* filter to implement adaptive filter 7 because the use *linear* filters for adaptive noise cancellation was well-known, and it would have been an obvious design choice. Ex. 1015, 4:28-5:15, Fig. 2.

- n) [7n] **"wherein the denoised output signals include less acoustic noise than acoustic signals received at the first and second physical microphones."**

142. In my opinion, Brandstein and Ikeda each disclose this limitation.

143. Both Brandstein and Ikeda are directed to removing or cancelling acoustic noise received from physical microphones in a microphone array. Brandstein states that applications of the beamforming techniques it describes include speech enhancement for “remov[ing] undesirable signals such as noise and reverberation.” Ex. 1005, 87. As shown in Fig. 5.2 and as I explain above in [7f], the denoised output signal *includes less acoustic noise* because the deep null in the direction of the interference (noise), the interference signal is decreased compared to the acoustic signal received by the physical microphones of the GJBF array. Ex. 1005, 89-90. Specifically, Brandstein states that “target signal extraction and interference suppression are simultaneously achieved.” Ex. 1005, 89-90. Therefore Brandstein discloses a denoised output signals that includes less acoustic noise than acoustic signals received at the first and second physical microphones

144. Ikeda also discloses that the output signals of the adaptive noise removal system (“*the denoised output signals*”) *include less acoustic noise than acoustic signals received at the physical microphones*) because, unlike the acoustic signals received at the physical microphones, the denoised output signals have had “the noise component contained in the main signal V_m ” canceled and “primarily consist[]

of the target sound component.” Ex. 1008, ¶¶[0001], [0011]-[0012], [0029], Abstract. Therefore Ikeda discloses a denoised output signals that includes less acoustic noise than acoustic signals received at the first and second physical microphones

4. **Claim 8: “The system of claim 7 and further comprising: a microphone array, the first and second physical microphones positioned m [sic] the microphone array.”**

145. In my opinion, Brandstein discloses this limitation.

146. As discussed above in [7a], Brandstein’s Fig. 5.1 shows a *microphone array* that includes *first and second physical microphones* that generate first and second microphone signals (e.g., x_0 and x_1). Ex. 1005, 89, Fig. 5.1.

5. **Claim 9: “The system of claim 7, wherein the single null is a region of the second linear response to speech having a measured response level that is lower than the measured response level of any other region of the second linear response to speech.”**

147. In my opinion, Brandstein discloses this limitation.

148. As the bottom diagram in Fig. 5.3 illustrates, the BM or bottom branch of the GSC (*second virtual microphone*) has a null in the direction of the speech signal:

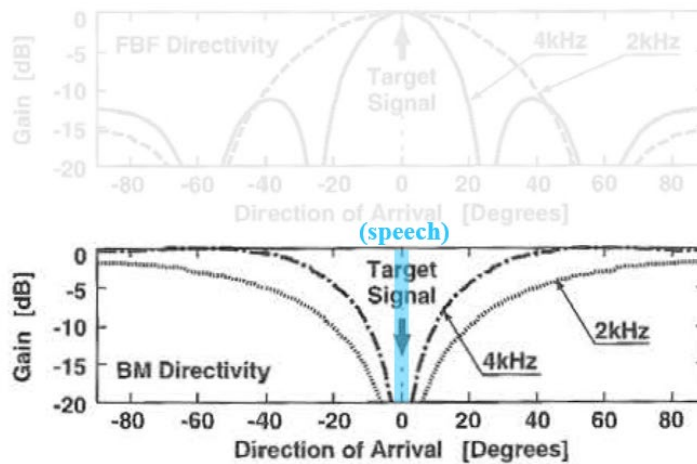


Fig. 5.3. Directivity pattern of a fixed beamformer (FBF) and a blocking matrix (BM).

Ex. 1005, Fig. 5.3 (annotated)

Fig. 5.3 shows the null is a true minimum for the second virtual microphone. Therefore, the null comprises a region of the *second* virtual microphone's *linear response to speech having a measured response level that is lower than a measured response level of any other region of the linear response to speech.*

149. This is true for both the first and second mappings because although the bottom directivity pattern of Fig. 5.3 is for that of the BM, a POSITA would have understood that the directivity pattern of the bottom branch after filtering through the MC would have resulted in a directivity pattern with the same overall characteristics.

6. **Claim 11: “The system of claim 7 and further comprising: a communications channel coupled with the processing component and including one or more of a wireless channel, a wired channel, and a hybrid wireless/wired channel.”**

150. In my opinion, Brandstein in combination with Ikeda renders this limitation obvious.⁵

151. Ikeda discloses a processing component in the form of the DSP. Ex. 1008, ¶¶[0016]-[0017] [0030] (“the adaptive filter 7, the digital adder 8, and the like can be configured a digital signal processor DSP”).

152. Brandstein discloses that applications of beamforming include microphone arrays for speech enhancement, such as in “hands free telephones.” Ex. 1005,

⁵ I understand that a claim is indefinite when it contains words or phrases whose meaning is unclear. I further understand that the lack of clarity could arise where a claim limitation is not properly introduced. For example, a claim referring “the [limitation]” may lack clarity where the claim contains no earlier recitation of that limitation and where it would be unclear as to what element the limitation was making reference. In my opinion, claim 11 lacks antecedent basis because claim 7, from which it depends, does not recite any “processing component.” For the purposes of my analysis, I assume that claim 11 requires “a processing component.”

87. In my opinion, it would have been obvious to a POSITA that there is a *communications channel* between such microphone arrays in a GJBF and the “hands-free telephone.” The POSITA would have readily appreciated that a *communications channel* relates to the flow of information (signal) from one point to another, and Ikeda’s DSP is along a *communication channel*, as it receives inputs from both *virtual microphones* and provides an output corresponding to the difference between the speech plus noise signal and the adaptively filtered noise signal. Therefore, a POSITA would have understood that Ikeda’s DSP represents part of the *communications channel*, which *is coupled with the processing component* in the Brandstein-Ikeda system.

153. It also would have been obvious to a POSITA that the *communications channel includes one or more of a wireless channel, a wired channel, and a hybrid wireless/wired channel* because all channels are wireless, wired, or both. Moreover, as Ikeda’s DSP is wired and Brandstein’s “hands free telephone” includes a wireless component, the POSITA would have known to implement the *communication channel* to be *wireless, wired, or hybrid wireless/wired*. Ex. 1016, 2:58-64, Fig. 2; Ex. 1017, 4:1-5:25, Figs. 3-4; Ex. 1018, 3:20-25, 3:50-64, Claim 1, Fig. 4b; Ex. 1019, ¶¶[0029]-[0030]; Ex. 1020, ¶[0138].

7. **Claim 12: “The system of claim 11 and further comprising: a communication device wirelessly coupled with the wireless channel of the communications channel.”**

154. In my opinion, Brandstein in combination with Ikeda renders this limitation obvious.

155. It is my opinion that a POSITA would have understood that the “hands-free telephone” disclosed in Brandstein is a *communication device*. Ex. 1005, 87. As discussed above for claim 11, the communications channel is between the “microphone array” and the “hands-free telephone.” A POSITA would have understood that a hands-free telephone operates wirelessly and that any hands-free device (*communication device*) is *wirelessly coupled with the wireless channel of the communications channel*.

8. **Claim 13: “The system of claim 7, wherein the second microphone signal is multiplied by a ratio, wherein the ratio is a ratio of a third distance to a fourth distance, the third distance being between the first physical microphone and the speech source and the fourth distance being between the second physical microphone and the speech source.”**

156. Claim 13 requires the second microphone signal to be multiplied by a ratio of the distance between the speech source and the first physical microphone to the distance between the speech source and the second physical microphone. In my opinion, Brandstein discloses or renders obvious this limitation.

157. It is my opinion that claim 13 would have been obvious to a POSITA based on Brandstein alone. A POSITA would have known that some applications involve near-field target signals, that the GSC requires time-alignment, that near-field time-alignment involves applying delays, and that time-aligning signals for near-field beamforming can be done by taking into account a standard near-field propagation model. The motivation and knowledge to use near-field time-alignment to implement the delays for the GSC (as described in Brandstein's Chapter 5) would have been immediately apparent to a POSITA.

158. As Brandstein teaches, the GSC's BM has a delay-and-sum (or delay-and-subtract) structure because the microphone signals are delayed so that their target-signal components are aligned in time, which is what allows the subtraction operations in the BM to block the target signal. Ex. 1005, 88-89. Specifically, Brandstein states that "[d]elay-and-sum beamformers apply time shifts to the array signals to compensate for the propagation delays in the arrival of the source signal at each microphone. These signals are time-aligned and summed together to form a single output signal." Ex. 1005, 159. When the Griffiths-Jim paper published in 1982, delay-and-sum beamforming was already well known in the art, and POSITAs under-

stood why Griffiths-Jim disclosed delay elements to time-align signals before feeding them to the BM. By 2007, it had been known for decades that time-alignment could be applied to the lower branch of the GSC.

159. In addition, in Chapter 5, Brandstein explains that the GSC's BM has a delay-and-subtract (delay-and-sum with negative coefficients) structure: "The BM in Fig. 5.1 has a simple delay-and-sum structure" (Ex. 1005, 90); and the "simplest structure for the BM is a delay-and-subtract beamformer which was described in the previous section" Ex. 1005, 88. A POSITA would have understood that the "previous section" refers to Chapter 2, which is the last chapter of Brandstein before Chapter 5 to discuss delay-and-sum/subtract structures.

160. In section 2.4, Brandstein discloses "all designs presented in section 2.3 can be implemented by using the GSC-structure." Ex. 1005, 35. For example, Brandstein teaches that a "multi-channel *time-aligned* input signal ... is projected onto the noise-only subspace^[footnote] by a blocking matrix **B**," with the footnote stating: "Which means that the desired signal is spatially filtered out (blocked)." Ex. 1005, 34 & n.5 (emphasis added). A "well-known example" of **B** "is the original Griffith[s]-Jim matrix which subtracts two adjacent channels." Ex. 1005, 34. Chapter 2 further states that there is a "close relationship between superdirective arrays and adaptive beamformers and their well-known implementation as a generalized

sidelobe canceler” and that “nearfield assumptions ... can be easily included.” Ex. 1005, 36-37.

161. Furthermore, a POSITA would have known of the differences between far-field and near-field sound and how to time-align signals from near-field sources. For example, Brandstein Chapter 2 discloses that time-aligning signals for near-field beamforming takes into account a standard near-field propagation model:

- Nearfield design, including attenuation of the desired signal [14], [22]

$$\mathbf{d}^T = [a_0 \exp(-j\omega\tau_0), a_1 \exp(-j\omega\tau_1), \dots, a_{N-1} \exp(-j\omega\tau_{N-1})], \quad (2.28)$$

$$a_i = \frac{\|\mathbf{q} - \mathbf{p}_{ref}\|}{\|\mathbf{q} - \mathbf{p}_i\|}, \quad (2.29)$$

$$\tau_i = \frac{\|\mathbf{q} - \mathbf{p}_{ref}\| - \|\mathbf{q} - \mathbf{p}_i\|}{c}, \quad (2.30)$$

where $\|\mathbf{q} - \mathbf{p}_{ref}\|$ and $\|\mathbf{q} - \mathbf{p}_i\|$ denote the distance between the vector location of the source $\mathbf{q}$ and a reference sensor $\mathbf{p}_{ref}$, or the sensor $\mathbf{p}_i$, respectively.

Ex. 1005, 25. The microphone attenuation factor a_i in Brandstein equation (2.29) is a ratio of the distance between the speech source and a reference (*first*) microphone to the distance between the speech source and the i th (*e.g., second*) microphone. This attenuation factor is *multiplied* to the i th (*e.g., second*) microphone signal to achieve equal amplitudes across all channels.

162. Brandstein's Chapter 2 therefore would have provided a POSITA with express motivation to use Brandstein's disclosure of near-field time-alignment in section 2.3 to implement the delay for the GSC. Furthermore, in Chapter 13, Brandstein explains that delay-and-sum beamforming was "first developed on the basis of the far-field assumption, but may also be extended to near-field beamforming." Ex. 1005, 290 (internal citation omitted). In my opinion, a POSITA would have understood that Brandstein teaches applying the complex-exponential term specified in Brandstein equation 2.28 to delay a signal in proportion to τ_i to time-align the signals by disclosing: (1) the delay-and-sum or delay-and-subtract structure for the FBF and BM, respectively; (2) the purpose of the delay in delay-and-sum beamforming is to time-align the sensor signals; (3) delay-and-sum beamforming can be extended to near-field designs; and (4) the standard near-field design specifies a delay whose magnitude is equal to the difference in propagation time of the target signal to one microphone relative to another (reference) microphone. Combining the teachings of Brandstein Chapters 2 and 5 to arrive at claim 13 is a simple addition of one known element (near-field time alignment) to another known element (the GSC) to obtain predictable results (*e.g.*, effective target-signal blocking in the lower branch). The combination also applies a known technique of time-alignment for near-field sources to the well-known GSC that is ready for improvement, with the combination yielding

predictable results. Thus, a POSITA would have had a reasonable expectation of success, and this limitation would have been obvious in view of Brandstein's Chapters 2 and 5 and the background knowledge of a POSITA. As discussed above, the microphone attenuation factor a_i is expressed by Brandstein equation (2.29):

- Nearfield design, including attenuation of the desired signal [14], [22]

$$\mathbf{d}^T = [a_0 \exp(-j\omega\tau_0), a_1 \exp(-j\omega\tau_1), \dots, a_{N-1} \exp(-j\omega\tau_{N-1})], \quad (2.28)$$

$$a_i = \frac{\|\mathbf{q} - \mathbf{p}_{ref}\|}{\|\mathbf{q} - \mathbf{p}_i\|}, \quad (2.29)$$

$$\tau_i = \frac{\|\mathbf{q} - \mathbf{p}_{ref}\| - \|\mathbf{q} - \mathbf{p}_i\|}{c}, \quad (2.30)$$

where $\|\mathbf{q} - \mathbf{p}_{ref}\|$ and $\|\mathbf{q} - \mathbf{p}_i\|$ denote the distance between the vector location of the source $\mathbf{q}$ and a reference sensor $\mathbf{p}_{ref}$, or the sensor $\mathbf{p}_i$, respectively.

Ex. 1005, 25. The distance between the speech source and a reference physical microphone (*first physical microphone*) can be considered a *third distance*. The distance between the speech source and the i th physical microphone (*e.g., second physical microphone*) can be considered a *fourth distance*. The microphone attenuation factor a_i is the ratio of the third distance to the fourth distance. This attenuation factor is *multiplied* to the i th (*e.g., second*) *microphone signal* to achieve equal amplitudes across all channels.

9. **Claim 14**

a) **[14pre] “A system, comprising:”**

163. To the extent the preamble is limiting, in my opinion, Brandstein discloses this limitation for the same reasons discussed above regarding [7pre].

b) **[14a] “a first virtual microphone comprising a first combination of a first microphone signal and a second microphone signal,”**

164. In my opinion, Brandstein discloses this limitation for the same reasons discussed above regarding [7a].

c) **[14b] “the first virtual microphone having a first linear response to speech and a first linear response to noise,”**

165. In my opinion, Brandstein discloses this limitation for the same reasons discussed above regarding [7c].

d) **[14c] “the first linear response to speech being substantially similar across a plurality of frequencies for a speech source located within a predetermined angle relative to an axis of a microphone array,”**

166. In my opinion, Brandstein discloses this limitation for the same reasons discussed above regarding [7d].

e) **[14d] “wherein the first microphone signal is output from a first physical microphone and the second**

microphone signal is output from a second physical microphone;”

167. In my opinion, Brandstein discloses this limitation for the same reasons discussed above regarding [7a].

- f) **[14e] “a second virtual microphone comprising a second combination of the first microphone signal and the second microphone signal,”**

168. In my opinion, Brandstein discloses this limitation for the same reasons discussed above regarding [7b].

- g) **[14f] “the second virtual microphone having a second linear response to speech and a second linear response to noise,”**

169. In my opinion, Brandstein discloses this limitation for the same reasons discussed above regarding [7e].

- h) **[14g] “the second linear response to noise being substantially similar to the first linear response to noise,”**

170. In my opinion, Brandstein discloses this limitation for the same reasons discussed above regarding [7f].

- i) **[14h] “one or both of the first linear response to noise and the second linear response to noise being non- zero in a direction toward a source of noise, and”**

171. In my opinion, Brandstein discloses this limitation for the same reasons discussed above regarding [7g].

- j) **[14i] “the second linear response to speech being substantially dissimilar to the first linear response to speech,”**

172. In my opinion, Brandstein discloses this limitation for the same reasons discussed above regarding [7h].

- k) **[14j] “wherein the second combination is different from the first combination,”**

173. In my opinion, Brandstein discloses this limitation for the same reasons discussed above regarding [7b].

- l) **[14k] “wherein the first virtual microphone and the second virtual microphone are distinct virtual directional microphones; and”**

174. In my opinion, Brandstein discloses this limitation for the same reasons discussed above regarding [7b].

175. Having different combinations forming the *first* and *second virtual microphones* means that the *first* and *second virtual microphones are distinct*. In the GJBF, the signals output from the top branch (*first virtual microphone*) are *distinct* from the signals output from the bottom branch or BM (*second virtual microphone* in both first and second mappings) because the first and second virtual directional microphones are formed with different combinations of signals from the microphone array (i.e. different combinations of $x_0(k)$ and $x_1(k)$). Ex. 1005, 89, Fig 5.1. Both

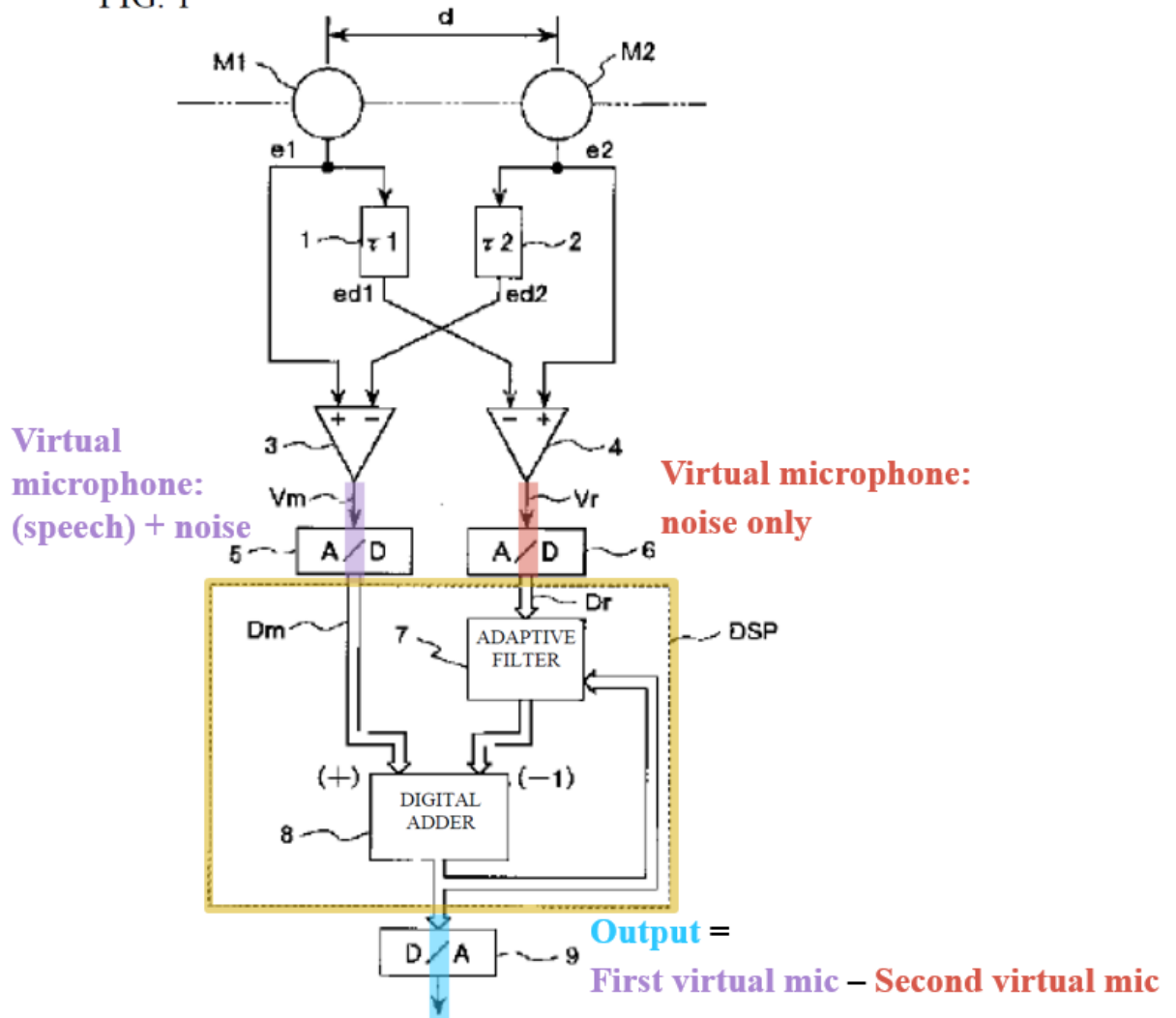
virtual microphones are *directional* because the directivity patterns of the FBF and BM are not constant in all directions, as shown in Fig. 5.3. Ex. 1005, 90, Fig 5.3.

- m) [14l] “a processing component coupled to the first and second virtual microphones, the processing component including an adaptive noise removal application receiving acoustic signals from the first virtual microphone and the second virtual microphone, filtering and summing the acoustic signals in the time domain, applying a varying linear transfer function between the acoustic signals, and generating an output signal,”

176. In my opinion, Ikeda discloses this limitation for the same reasons discussed above regarding [7j]-[7m].

177. Regarding the *processing component*, Ikeda discloses that a DSP as shown in Fig. 1 (annotated below) is a *processing component* including an adaptive noise removal application. (x. 1008, ¶¶ [0016]-[0017], [0027]-[0030]. The DSP is *coupled* to the main and reference microphones and corresponding main signal V_m and reference signal V_r via A/D converters 5 and 6. Ex. 1008, ¶ [0027], Fig. 1; Ex. 1001, 32:25-32.

FIG. 1



Ex. 1008, Fig. 1 (annotated).

- n) [14m]“wherein the output signal is a denoised acoustic signal.”

178. In my opinion, Brandstein and Ikeda disclose this limitation for the same reasons discussed above regarding [7j]-[7n].

10. **Claim 16: “The system of claim 14 and further comprising: a communications channel coupled with the processing component and including one or more of a wireless channel, a wired channel, and a hybrid wireless/wired channel.”**

179. In my opinion, Brandstein and Ikeda render this limitation obvious for the same reasons discussed above regarding claim 11.

11. **Claim 17: “The system of claim 16 and further comprising: a communication device wirelessly coupled with the wireless channel of the communications channel.”**

180. In my opinion, Brandstein and Ikeda render this limitation obvious for the same reasons discussed above regarding claim 12.

12. **Claim 18: “The system of claim 14, wherein the acoustic signals from the first virtual microphone, the second virtual microphone or both are delayed.”**

181. In my opinion, Brandstein discloses this limitation.

182. Brandstein discloses that the signal from the FBF is delayed. Brandstein discloses that FBF output $b(k)$ is *delayed* through the delay block labeled Z^{-L_1} to form the “delayed output signal $b(k-L_1)$ ” (Ex. 1005, 88), which is highlighted below in yellow:

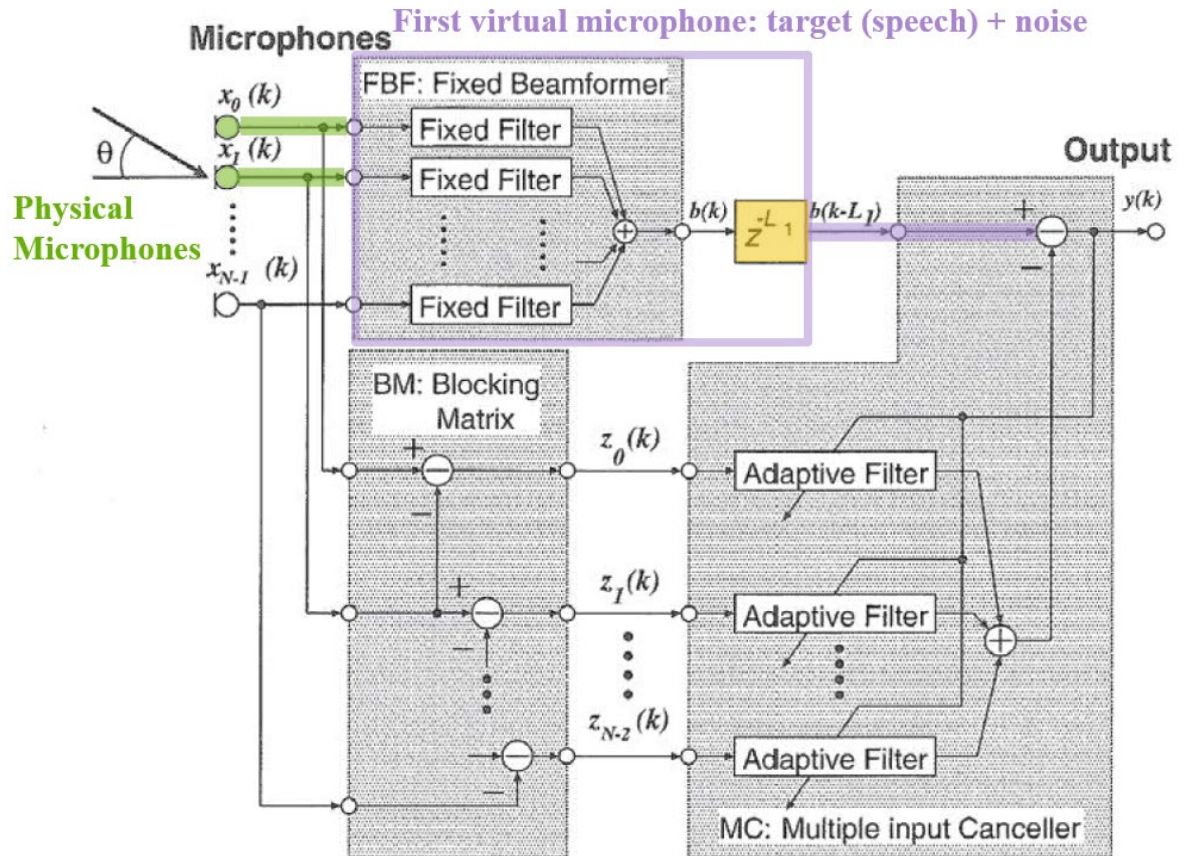


Fig. 5.1. Griffiths-Jim beamformer. It comprises a fixed beamformer (FBF), a multiple-input canceller (MC), and a blocking matrix (BM).

Ex. 1005, Fig. 5.1 (annotated).

Therefore, at least the *acoustic signals* from the FBF (*first virtual microphone*) are *delayed*.

13. **Claim 19: “The system of claim 18, wherein the delay is raised to a power that is proportional to a time difference between arrival of the speech at the first virtual microphone and arrival of the speech at the second virtual microphone.”**

183. In my opinion, Brandstein renders obvious this limitation. The '080 Patent defines:

$$\gamma = \frac{d_2 - d_1}{c} f_s \text{ (samples),}$$

where “[t]he distances d_1 and d_2 are the distance from O_1 and O_2 to the speech source . . . , and γ is their difference divided by c , the speed of sound, and multiplied by the sampling frequency f_s .” Ex. 1001, 9:64-67. The '080 Patent later refers to γ as “the time difference between arrival of speech at V_1 compared to V_2 .” Ex 1001, 13:66-67. Thus, according to the '080 Patent, the *time difference* between arrival of speech at the virtual microphones is calculated from the difference between distances from the physical microphones to the speech source.

184. As discussed above for claim 13, Brandstein renders obvious extending the standard GSC to a near-field design, including by time-aligning the microphone signals fed to the BM by applying delay term $e^{-j\omega\tau_i}$, whose exponent (*power*) is *proportional* to τ_i . Brandstein defines τ_i as:

$$\tau_i = \frac{\|q - p_{ref}\| - \|q - p_i\|}{c}, \quad (2.30)$$

where $\|q - p_{ref}\|$ and $\|q - p_i\|$ denote the distance between the vector location of the source q and a reference sensor p_{ref} , or the sensor p_i , respectively.

Ex. 1005, 25.

185. Brandstein's τ_i results from dividing the *difference* of the distance between the source (i.e. *speech*) q and a reference sensor (*second microphone*)⁶ and the distance between the source and the i th sensor (*first microphone*) by the speed of sound. Brandstein's τ_i is thus *a time difference between arrival of the speech at*

⁶ I labeled the reference sensor as the first microphone and the i th sensor as the second microphone for claim 13 because that labeling met the limitations of claim 13. Likewise the labeling of the reference sensor as the second microphone and the i th sensor as the first microphone for this claim meets the limitations of this claim, which does not depend in any way from claim 13. Furthermore, a POSITA would have recognized that either microphone in a GSC could be designated as the reference microphone, and therefore be designated the “first” or the “second” microphone without affecting the analysis. For these reasons, I switch the label for the first and second physical microphones for this claim.

the first virtual microphone and arrival of the speech at the second virtual microphone.

14. **Claim 20: “The system of claim 19, wherein the power is proportional to a sampling frequency multiplied by a quantity equal to a third distance subtracted from a fourth distance, the third distance being between a first physical microphone and the speech source, the fourth distance being between a second physical microphone and the speech source, and the first and second physical microphones are positioned in the microphone array.”**

186. In my opinion, Brandstein discloses this limitation.

187. Brandstein discloses “the first and second physical microphones are positioned in the microphone array” for the same reasons discussed in above regarding claim 8.

188. As I explain above, Brandstein’s τ_i obtained from subtracting the distance between the source (i.e. *speech*) q and the i th sensor (*first physical microphone*), which I call the *third distance* here, from the distance between the source and a reference sensor (*second physical microphone*), which I call the *fourth distance* here, and then dividing by the speed of sound, c . Therefore, τ_i is *proportional to a quantity equal to a third distance subtracted from a fourth distance*.

189. Claim 20 further recites an exponent “proportional” to “a sampling frequency multiplied by” a quantity such as Brandstein’s τ_i . But τ_i is proportional to

the sampling frequency multiplied by τ_i , so Brandstein's delay term $e^{-j\omega\tau_i}$ already satisfies claim 20.

190. Alternatively, although Brandstein discloses the delay term in the form above with time expressed in seconds, a POSITA would have known, as a matter of routine practice, how to rewrite the formula to express time in terms of number of samples, as is common for digital signal processing, which involves discrete sampling. To convert from seconds to samples, a POSITA would simply multiply the delay expressed in seconds by the sampling frequency f_s expressed in samples per second. For example, after that trivial operation, Brandstein's delay term $e^{-j\omega\tau_i}$ becomes $e^{-j\omega\tau_i f_s}$. Using z-transform notation, this is commonly written as $z^{-\tau_i f_s}$. Applying that exponential term delays a signal by $\tau_i f_s$ sample.

191. In my opinion, a POSITA would have been motivated to rewrite Brandstein's near-field delay in terms of samples. Brandstein itself motivates rewriting this delay element in terms of samples for use with the GSC because Fig. 5.1 is shown in terms of samples. Ex. 1005, 88-89 & n.1. Brandstein explains the effect of z^{-L_1} block in Fig. 5.1 as an " L_1 -sample delay." Brandstein also discloses implementing the GSC with a DSP. Ex. 1005, 104 & Fig. 5.16. A POSITA would have been motivated to implement the GSC digitally because digital signal processing was a known technique, and, among other advantages over analog systems, digital

circuits can be reproduced more easily in large quantities at low cost, digital systems are unaffected by changes of component values in analog systems caused by factors such as age, temperature, and humidity, digital signal processing allows for flexibility through digitally programmable systems, digital systems are generally more accurate, digital signals can be easily stored on media using semiconductor chips, and digital signals can be transmitted over long distances. Converting Brandstein's delay formula from seconds to samples for digital signal processing would have involved the simple and well-known operation of multiplying by the sampling frequency to convert time units from, e.g., seconds to samples.

15. **Claim 1**

a) **[1pre] "A system, comprising:"**

192. To the extent the preamble is limiting, in my opinion, Brandstein discloses this limitation for the same reasons discussed above regarding [7pre].

b) **[1a] "a microphone array including a first physical microphone outputting a first microphone signal and a second physical microphone outputting a second microphone signal;"**

193. In my opinion, Brandstein discloses this limitation for the same reasons discussed above regarding [7a] and [7d].

c) **[1b] "a processing component coupled to the microphone array and generating a virtual**

**microphone array including a first virtual microphone
and a second virtual microphone,”**

194. In my opinion, Brandstein discloses *a virtual microphone array including a first virtual microphone and a second virtual microphone* for the same reasons discussed above regarding [7a] and [7b].

195. It is my opinion that it would have been obvious to implement the top branch and the bottom branch of the GJBF digitally using a processor (“*processing component*”), which was a well-known technique and the standard way to accomplish such operations. A POSITA would have understood that such a processor is *coupled* to the GJBF microphone array to process the microphone signals $x_0(k)$ and $x_1(k)$.

- d) **[1c] “the first virtual microphone including a first combination of the first microphone signal and the second microphone signal, the second virtual microphone including a second combination of the first microphone signal and the second microphone signal, wherein the second combination is different from the first combination,”**

196. In my opinion, Brandstein discloses this limitation for the same reasons discussed above regarding [7a] and [7b].

- e) **[1d] “wherein the first virtual microphone and the second virtual microphone have substantially similar**

**responses to noise and substantially dissimilar
responses to speech; and”**

197. In my opinion, Brandstein discloses this limitation for the same reasons discussed above regarding [7c], [7e], [7f], and [7h].

- f) **[1e] “an adaptive noise removal application coupled to the processing component and generating denoised output signals by forming a plurality of combinations of signals output from the first virtual microphone and the second virtual microphone, by filtering and summing the plurality of combinations of signals in the time domain, and by a varying linear transfer function between the plurality of combinations of signals,”**

198. In my opinion, Ikeda discloses this limitation for the same reasons discussed above regarding [7j]-[7m].

199. It is my opinion that a POSITA would have understood Ikeda’s *adaptive noise removal application is coupled to the processing component* to generate the main and reference microphones. Ex. 1008, ¶¶ [0027]-[0029], Fig. 1.

- g) **[1f] “wherein the denoised output signals include less acoustic noise than acoustic signals received at the microphone array.”**

200. In my opinion, Brandstein and Ikeda disclose this limitation for the same reasons discussed above regarding [7j]-[7m] and [7n].

16. **Claim 2: “The system of claim 1, wherein the acoustic noise comprises noise content and the acoustic signals comprise speech content.”**

201. In my opinion, Brandstein discloses this limitation for the same reasons discussed above regarding [7i].

17. **Claim 3: “The system of claim 2, wherein the speech content comprises human speech.”**

202. In my opinion, Brandstein discloses this limitation for the same reasons discussed above regarding [7i].

18. **Claim 5: “The system of claim 1 and further comprising: a communications channel coupled with the processing component and including one or more of a wireless channel, a wired channel, and a hybrid wireless/wired channel.”**

203. In my opinion, Brandstein and Ikeda render this limitation obvious for the same reasons discussed above regarding claim 11.

19. **Claim 6: “The system of claim 5 and further comprising: a communication device wirelessly coupled with the wireless channel of the communications channel.”**

204. In my opinion, Brandstein and Ikeda render this limitation obvious for the same reasons discussed above regarding claim 12.

B. Ground 2: Brandstein in combination with Ikeda and McCowan

205. For the reasons I explain below, it is my opinion that claims 13, 19, and 20 would have been obvious in view of Brandstein, Ikeda, and McCowan.

1. Motivation to Combine

206. Applying near-field time-alignment for the lower branch of the GSC would have been part of a POSITA's background knowledge because McCowan discloses this:

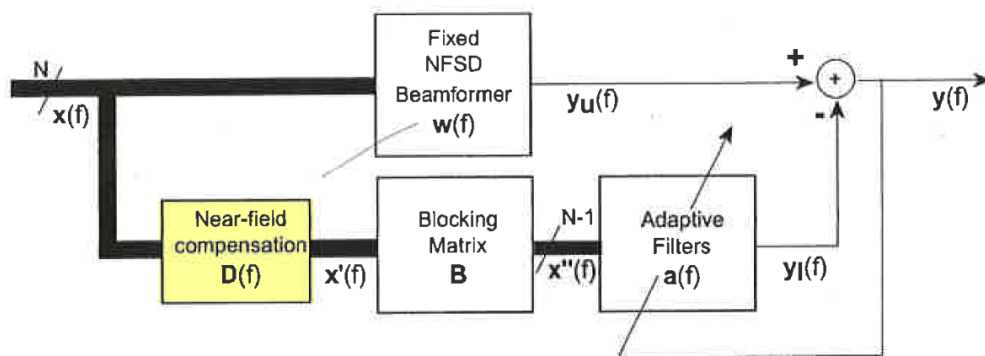


FIG. 3. Near-field adaptive beamformer.

Ex. 1009, 91, Fig. 3.

207. McCowan discloses an adaptive noise cancellation system for near-field speech sources. Ex. 1009, 87-89, 97. Fig. 3 shows that the input $x(f)$ includes signals from multiple microphones is processed via two signal paths. Ex. 1009, 90, 93 (equation 15), 95-96. The BM “block[s] the desired signal from the adaptive noise estimate” by subtracting one microphone signal from another. Ex. 1009, 93,

95 (equations 18, 19). As McCowan explains, “[t]o ensure complete blocking, the desired signal must both be time aligned and have equal amplitudes across all channels.” Ex. 1009, 93.

208. McCowan teaches that for a near-field sound source, phase misalignment across microphones results from “[t]he difference in propagation time to [the *i*th] microphone with respect to the reference microphone (*i*=1),” defined as:

$$\tau_i = \frac{d_i - d_1}{c}, \quad (3)$$

where d_i is the distance between the source and the *i*th microphone and c is the speed of sound. Ex. 1009, 89-90. Additionally, amplitude differences across microphones are “negligible for far-field sources, but can be significant in the near-field case.” Ex. 1009, 90. The attenuation factor of wave amplitude on the *i*th microphone, with respect to the amplitude on the reference microphone (*i*=1), is defined as:

$$\alpha_i = \frac{d_1}{d_i}. \quad (4)$$

Ex. 1009, 90.

209. McCowan explains that if $x_1(f)$ is defined as the desired source at a reference microphone, then the signal on the i th microphone is given by the equation:

$$x_i(f) = \alpha_i x_1(f) e^{-j2\pi f \tau_i} \quad (5)$$

Ex. 1009 90.

210. Fig. 3 of McCowan shows that “[o]nce this near-field compensation has been performed, a standard GSC blocking matrix can be employed to block the desired signal from the adaptive path.” Ex. 1009, 93. McCowan also states that “[t]he blocking matrix and adaptive noise canceling filters are taken from the standard GSC technique [citation]” Ex. 1009, 93 (citing Griffiths-Jim).

211. In view of this background knowledge, it is my opinion that claim 13 would have been obvious to a POSITA based on the teachings of Brandstein and McCowan. A POSITA would have known that some applications involve near-field target signals, that the GSC’s BM requires time-alignment, that near-field time-alignment involves applying delays, and that McCowan discloses such near-field time-alignments followed by a standard GSC BM. The motivation and knowledge to use near-field time-alignment to implement the delays for the delay-and-subtract BM in the original Griffiths-Jim GSC (as described in Brandstein’s Chapter 5) would have been immediately apparent to a POSITA.

212. Alternatively, to the extent near-field time-alignment would require modifying the GSC described in Brandstein's Chapter 5, such modification represents nothing more than the simple addition of one known element (near-field time alignment) to another known element (the GSC) to obtain the predictable results of effective target-signal blocking in the lower branch (to permit effective noise cancellation in the final output without canceling the target signal).

1. **Expectation of success**

213. In my opinion, the combination of McCowan and Brandstein applies a known technique (time-alignment for near-field sources) to a well-known device (GSC) that is ready for improvement, with the combination yielding predictable results. A POSITA would have had a reasonable expectation of success because, as McCowan demonstrates, near-field time-alignment with the GSC was trivial and shown to work. Claim 13 would therefore have been obvious in view of Brandstein's Chapter 5 and the background knowledge of a POSITA.

2. **Claim 13: "The system of claim 7, wherein the second microphone signal is multiplied by a ratio, wherein the ratio is a ratio of a third distance to a fourth distance, the third distance being between the first physical microphone and the speech source and the fourth distance being between the second physical microphone and the speech source."**

214. In my opinion, Brandstein in combination with McCowan renders this limitation obvious.

215. As discussed above, the microphone attenuation factor α_i in McCowan equation (4):

$$\alpha_i = \frac{d_1}{d_i}, \quad (4)$$

is a ratio of d_1 (the distance between the speech source and a reference (*first*) microphone) to d_i (the distance between the speech source and the i th (*e.g., second*) microphone). Ex. 1009, 89. This attenuation factor is *multiplied* to the i th (*e.g., second*) microphone signal to achieve equal amplitudes across all channels.

3. **Claim 19: “The system of claim 18, wherein the delay is raised to a power that is proportional to a time difference between arrival of the speech at the first virtual microphone and arrival of the speech at the second virtual microphone.”**

216. In my opinion, Brandstein in combination with McCowan discloses this limitation.

217. McCowan defines τ_i as:

$$\tau_i = \frac{d_i - d_1}{c}, \quad (3)$$

Ex. 1009, 89. In the above equations, c is the speed of sound. For McCowan, d_1 is the distance from the *speech source* to the reference (i.e. *first*) microphone and d_i is

the distance from the speech source to the i th (e.g., *second*) microphone. The difference between d_i and d_1 is divided by the speed of sound to obtain *a time difference between arrival of the speech at the first virtual microphone and arrival of the speech at the second virtual microphone, τ_i .*

4. **Claim 20: “The system of claim 19, wherein the power is proportional to a sampling frequency multiplied by a quantity equal to a third distance subtracted from a fourth distance, the third distance being between a first physical microphone and the speech source, the fourth distance being between a second physical microphone and the speech source, and the first and second physical microphones are positioned in the microphone array.”**

218. In my opinion, Brandstein in combination with McCowan discloses this limitation.

219. Claim 20 further recites an exponent “proportional” to “a sampling frequency multiplied by” a quantity such as McCowan’s τ_i . But τ_i is proportional to the sampling frequency multiplied by τ_i , so McCowan’s delay term $e^{-j\omega\tau_i}$ satisfies claim 20 for the same reasons as discussed above in Section VIII.A.14. (*Id.*)

C. Ground 3: Brandstein in combination with Ikeda and Yang

220. For the reasons I explain below, it is my opinion that claims 4, 10, and 15 would have been obvious in view of Brandstein, Ikeda, and Yang.

1. Motivation to Combine

221. Neither Brandstein nor Ikeda expressly teach using a voice activity detector. In my opinion, it would have been obvious to incorporate a voice activity detector into the Brandstein-Ikeda combination. A POSITA would have recognized the benefit of using a VAD to identify when there is speech and when there is no speech so that the system knows when to perform adaptation. A POSITA would have known that better adaptation and filtering to the noise characteristics for noise cancellation can be achieved when noise is the only thing being received. Therefore, it would have been helpful to identify when speech stops and starts, which a VAD can do. In my opinion, this modification would have been nothing more than combining prior art elements according to known methods to yield predictable results. Indeed, Yang states that “Voice activity detector 240 detects the presence of speech at the microphones and provides the Act control signal indicating periods of speech activity. The detection of speech activity can be performed in various manners known in the art.” Ex. 1010 [¶0031]

2. Reasonable expectation of Success

222. A POSITA would have had a reasonable expectation of success to add the voice activity detector of Yang to the Brandstein-Ikeda combination. Yang teaches that the voice activity detector can be implemented with a GJBF. Yang discloses a noise suppression system that includes a beam forming unit and a noise

suppression unit. Ex. 1010, ¶[0026]. Yang teaches that the beam forming unit “may be implemented with a ‘Griffiths-Jim’ beam former.” Ex. 1010, ¶[0038].

223. Yang also teaches an adaptive noise reduction system as shown in Fig. 4 (annotated below), which takes as inputs a speech-plus-noise signal and a mostly noise only signal. Ex. 1010, ¶¶ [0058]-[0062]. Thus, a POSITA would have recognized that Yang’s adaptive noise reduction system takes as inputs, the output of the top and bottom branches of the GJBF. Therefore, Yang itself suggests that a voice activity detector can be combined with a GJBF, and it is my opinion that using a voice activity detector to determine the timing when microphones do not pick up human speech would have been within the knowledge and skill of a POSITA.

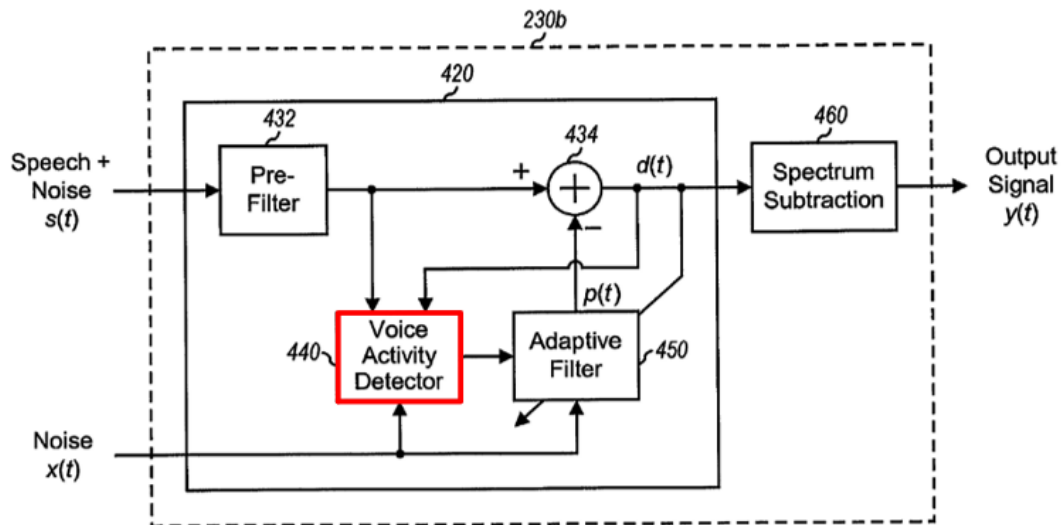


FIG. 4

Ex. 1009, Fig. 4 (annotated).

3. **Claim 10: “The system of claim 7 and further comprising: a voice activity detector (VAD) coupled with the processing component and operative to generate voice activity signals.”**

224. In my opinion, Brandstein in combination with Ikeda and Yang renders obvious this limitation.

225. Ikeda discloses a *processing component*. Ex. 1008, ¶¶ [0016]-[0017], [0030] (“a digital signal processor DSP”).

226. Ikeda discloses that adaptive filter 7 “adaptively varies its characteristics with respect to changes in the input signal over time,” where “[t]he training sequence timing [for determining the filter response characteristics] is defined as a timing at which the microphone units M1 and M2 do not pick up the target sound.” Ex. 1008, ¶ [0028].

227. In my opinion, it would have been obvious to determine the “timing at which the microphone units M1 and M2 do not pick up the target sound,” such as human speech, using a voice activity detector as taught by Yang. Ex. 1008, ¶¶ [0028], [0036]; Ex. 1010, ¶¶ [0058]-[0062], Fig. 4. Yang teaches: “Voice activity detector 440 detects the presence of speech activity ... and provides a control signal that enables the adaptation of the coefficients of the [adaptive filter 450] when no speech activity is detected.” Ex. 1010, ¶¶ [0060]-[0061].

228. Therefore, Ikeda in view of Yang teaches a *voice activity detector* is *coupled with the processing component* (where adaptive filter 7 is configured) and operative to generate voice activity signals. Ex. 1008, ¶¶ [0016]-[0017], [0030]; Ex. 1010, ¶¶ [0058]-[0062], Fig. 4.

4. **Claim 15: “The system of claim 14 and further comprising: a voice activity detector (VAD) coupled with the processing component and operative to generate voice activity signals.”**

229. In my opinion, Brandstein in combination with Ikeda and Yang renders obvious this limitation for the same reasons discussed above regarding claim 10.

5. **Claim 4: “The system of claim 1 and further comprising: a voice activity detector (VAD) coupled with the processing component and operative to generate voice activity signals.”**

230. In my opinion, Brandstein in combination with Ikeda and Yang renders obvious this limitation for the same reasons discussed above regarding claim 10.

IX. CONCLUSION

231. For the foregoing reasons, it is my opinion that claims 1-20 of the ‘080 patent would have been obvious to a person of ordinary skill in the art at the time of the alleged invention in view of the prior art discussed above.

232. I reserve the right to supplement my opinions in the future to address or respond to any issues that the Patent Owner may raise, as well as new information including, but not limited to, any claim constructions advanced by the Patent Owner

Declaration of Dr. Michael Johnson in Support of
Petition for *Inter Partes* Review of USP No. 10,779,080

or adopted by the Board in the Institution Decision, and to respond to any alleged secondary considerations as they become available to me.

233. I declare that all statements made herein are of my own knowledge and are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code.

Declaration of Dr. Michael Johnson in Support of
Petition for *Inter Partes* Review of USP No. 10,779,080

Executed on December 20, 2023.



Michael Johnson, Ph.D.