

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

AMAZON.COM, INC.,
Petitioner,

v.

JAWBONE INNOVATIONS, LLC,
Patent Owner.

IPR2023-00251
U.S. Patent No. 11,122,357

DECLARATION OF RICHARD M. STERN, Ph.D.
IN SUPPORT OF PETITION FOR *INTER PARTES* REVIEW
OF CLAIMS 1-20 OF U.S. PATENT NO. 11,122,357

TABLE OF CONTENTS

I.	BACKGROUND -----	1
A.	Experience and Qualifications -----	1
B.	Materials Considered-----	2
II.	APPLICABLE LEGAL STANDARDS -----	4
A.	Claim Construction -----	4
B.	Obviousness -----	6
III.	PERSON OF ORDINARY SKILL IN THE ART -----	9
IV.	TECHNOLOGY BACKGROUND AND OVERVIEW OF PRIOR ART -----	11
A.	[1982] Griffiths-Jim: The GSC / Griffiths-Jim Beamformer -----	11
B.	[1982-c.2001] Use of GSC with Microphone Arrays for Speech Applications-----	15
C.	[2001] Brandstein-----	16
D.	[2001] Gannot -----	21
E.	[2001-2002] Brandstein and McCowan Discuss Near- Field Design -----	24
F.	[2003] Burnett '848 -----	30
V.	THE '357 PATENT -----	33
VI.	CLAIMS 1-20 OF THE '357 PATENT WOULD HAVE BEEN OBVIOUS -----	37

TABLE OF AUTHORITIES
(Cont'd)

A.	Claims 1-20 Would Have Been Obvious in View of Brandstein and Gannot-----	37
1.	Claim 1 -----	37
a.	Preamble-----	37
b.	1[a]: First Virtual Microphone Comprising a Combination of Signals from First and Second Physical Microphones -----	38
c.	1[b][1]: Second Virtual Microphone-----	41
d.	1[b][2]: Substantially Similar Responses to Noise and Substantially Dissimilar Responses to Speech -----	42
e.	1[c][1]: Signal Processor Operative to Combine Microphone Signals by Filtering and Summing in the Time Domain -----	45
f.	1[c][2]: Applying a Varying Linear Transfer Function -----	48
g.	1[c][3]: Generating an Output Signal with Reduced Noise -----	50
h.	Further Comments on Claim 1 in View of Brandstein and Gannot-----	51
2.	Claim 2 -----	52
3.	Claim 3 -----	53
4.	Claim 4 -----	54
5.	Claim 5 -----	56

TABLE OF AUTHORITIES

(Cont'd)

6.	Claim 6 -----	64
7.	Claim 7 -----	67
8.	Claim 8 -----	67
9.	Claim 9 -----	68
10.	Claim 10-----	69
11.	Claim 11 -----	70
12.	Claim 12-----	72
13.	Claim 13-----	73
14.	Claim 14-----	74
15.	Claim 15-----	76
a.	Preamble-----	76
b.	15[a]: First Virtual Microphone -----	76
c.	15[b][2]: Second Virtual Microphone -----	76
d.	15[b][2]: Substantially Similar Responses to Noise and Substantially Dissimilar Responses to Speech -----	77
e.	15[c]: Virtual Microphone Array with a Single Null -----	77
f.	15[d][1]: Signal Processor-----	78
g.	15[d][2]: Applying a Varying Linear Transfer Function -----	79

TABLE OF AUTHORITIES
(Cont'd)

h.	15[d][3]: Generating an Output Signal with Reduced Noise -----	79
i.	Further Comments on Claim 15 in View of Brandstein and Gannot -----	79
16.	Claim 16-----	80
17.	Claim 17-----	80
18.	Claim 18-----	82
19.	Claim 19-----	82
20.	Claim 20-----	82
B.	Claims 4, 13, 14, 18, and 19 Would Have Been Obvious in View of Brandstein, Gannot, and Griffiths-Jim -----	83
C.	Claims 4-6, 13, 14, 18, and 19 Would Have Been Obvious in View of Brandstein, Gannot, and McCowan -----	86
VII.	CONCLUSION -----	88

1. I am making this declaration at the request of Petitioner Amazon.com, Inc. (“Amazon”). I have been retained by Amazon as a technical expert in this matter.

2. I am being compensated for my work on this case. My compensation does not depend on the content of this Declaration or the outcome of these proceedings. I do not own any stock in Amazon and, to my knowledge, I have no financial interest in Amazon.

I. BACKGROUND

A. Experience and Qualifications

3. I am a Professor at Carnegie Mellon University (CMU) in the Department of Electrical and Computer Engineering, the Department of Computer Science, and the Language Technologies Institute. I have been on the faculty of Carnegie Mellon since 1977.

4. I received the S.B. degree from the Massachusetts Institute of Technology (MIT) in 1970, the M.S. from the University of California, Berkeley, in 1972, and the Ph.D. from MIT in 1977, all in electrical engineering.

5. I am a fellow of the Institute of Electrical and Electronics Engineers (IEEE), the Acoustical Society of America, and the International Speech Communication Association (ISCA). I was the ISCA 2008-2009 Distinguished Lecturer, a recipient of the Allen Newell Award for Research Excellence in 1992,

and I served as the General Chair of Interspeech 2006. Interspeech is the world's largest technical conference focused on speech processing and application.

6. Much of my current research is in spoken language systems, where I am particularly concerned with the development of techniques with which automatic speech recognition can be made more robust with respect to changes in environment and acoustical ambience.

7. I have actively worked on the theory and application of systems using microphone arrays over a period of decades (*e.g.*, Stern et al., 2008; Stern and Menon, 2020), and my research group has developed several array-based algorithms to improve speech recognition accuracy in difficult acoustical environments (*e.g.*, Seltzer et al., 2004; Stern et al., 2007; Kim et al., 2009; Moghimi and Stern, 2019). My relevant publications, including those cited above, are available on Carnegie Mellon's web site at <http://www.cs.cmu.edu/afs/cs/user/robust/www/papers.html>.

8. I understand a copy of my current curriculum vitae, which lists my publications for the last ten years, is being submitted as Exhibit 1012.

B. Materials Considered

9. In preparing this declaration, I have considered the following materials:

Exhibit No.	Description
1001	U.S. Patent No. 11,122,357 ("the '357 patent")

Exhibit No.	Description
1003	Excerpts of MICROPHONE ARRAYS: SIGNAL PROCESSING TECHNIQUES AND APPLICATIONS (Michael Brandstein & Darren Ward eds., Springer-Verlag 2001) (“Brandstein”)
1004	Sharon Gannot et al., <i>Signal Enhancement Using Beamforming and Nonstationarity with Applications to Speech</i> , vol. 49, no. 8 IEEE TRANSACTIONS ON SIGNAL PROCESSING, 1614 (Aug. 2001) (“Gannot”)
1005	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”)
1006	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”)
1007	U.S. Patent No. 5,651,071 (“Lindemann”)
1008	U.S. Patent No. 5,627,799 (“Hoshuyama”)
1009	U.S. Patent Publication No. 2003/0128848 (“Burnett ’848”)
1010	Excerpts from the ’357 patent’s file history
1011	U.S. Provisional Patent Application No. 61/045,377
1012	Curriculum Vitae of Richard M. Stern, Ph.D.

10. I have also relied on my education, training, and experience, and my knowledge of pertinent literature in the field of the ’357 patent.

II. APPLICABLE LEGAL STANDARDS

11. I have been asked to provide my opinion as to whether the claims of the '357 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.

12. I am an electrical engineer by training and profession. The opinions I am expressing in this report involve the application of my training and technical knowledge and experience to the evaluation of certain prior art with respect to the '357 patent.

13. Although I have been involved as a technical expert in patent matters before, I am not an expert in patent law. Therefore, attorneys from Knobbe, Martens, Olson & Bear have provided me with guidance as to the applicable patent law in this matter. The paragraphs below express my understanding of how I must apply current principles related to patent validity to my analysis.

A. Claim Construction

14. It is my understanding that in determining whether a patent claim is anticipated by or obvious in view of the prior art, the Patent and Trademark Office construes the claim by giving the claim terms their plain and ordinary meaning, as they would have been understood by a person of ordinary skill in the art at the time of the invention in view of the intrinsic record (patent claims, patent specification, and file history). For the purposes of this review, and to the extent necessary, I have

interpreted each claim term in accordance with its plain and ordinary meaning as it would have been understood by a person of ordinary skill in the art at the time of the invention, in view of the intrinsic record. I have been instructed to assume that the time of the invention is June 13, 2007, which I understand to be the filing date of the earliest provisional application to which the '357 patent claims priority.

15. I understand that a patent and its prosecution history are considered “intrinsic evidence” and are the most important sources for interpreting claim language in a patent. I also understand that in reading the claim, I must not import limitations from the specification into the claim terms; in other words, I must not narrow the scope of the claim terms by implicitly adding disclosed limitations that have no express basis in the claims. The prosecution history of related patents and applications can also be relevant.

16. I understand that sources extrinsic to a patent and its prosecution history (such as dictionary definitions and technical publications) may also be used to help interpret the claim language, but that such extrinsic sources cannot be used to contradict the unambiguous meaning of the claim language that is evident from the intrinsic evidence.

17. Unless expressly stated herein, I have applied the plain and ordinary meaning of the claim terms, which I understand is the meaning that a person of

ordinary skill in the art would have given to terms in June 2007 based on a review of the intrinsic evidence.

B. Obviousness

18. It is my understanding that a claim is “obvious” if the claimed subject matter as a whole would have been obvious to a person of ordinary skill in the art at the time of the alleged invention. I understand that an obviousness analysis involves a number of considerations. I understand that the following factors must be evaluated to determine whether a claim would have been obvious: (i) the scope and content of the prior art; (ii) the differences, if any, between each claim of the ’357 patent and the prior art; (iii) the level of ordinary skill in the art in June 2007; and (iv) additional considerations, if any, that indicate that the invention was obvious or not obvious. I understand that these “additional considerations” are often referred to as “secondary considerations” or “objective indicia” of non-obviousness or obviousness.

19. I also understand that the frame of reference when evaluating obviousness is what a hypothetical person of ordinary skill in the pertinent art would have known in June 2007. I understand that the hypothetical person of ordinary skill is presumed to have knowledge of all pertinent prior art references.

20. I understand that a prior art reference may be a pertinent prior art reference (or “analogous art”) if it is in the same field of endeavor as the patent or if

it is pertinent to the problem that the inventors were trying to solve. I understand that a reference is reasonably pertinent if it logically would have commended itself to an inventor's attention in considering the problem at hand. If a reference relates to the same problem as the claimed invention, I understand that supports use of the reference as prior art in an obviousness analysis. Here, all of the references relied on in my obviousness analysis below are in the same field of endeavor as the '357 patent, e.g., microphone array signal processing and noise reduction. The references are also pertinent to a particular problem the inventor was focused on, e.g., noise suppression.

21. It is my understanding that the law recognizes several rationales for combining references or modifying a reference to show obviousness of claimed subject matter. Some of these rationales include:

- combining prior art elements according to known methods to yield predictable results;
- simple substitution of one known element for another to obtain predictable results;
- a predictable use of prior art elements according to their established functions;
- using known techniques to improve similar devices (methods, or products) in the same way;

- applying a known technique to a known device (method, or product) ready for improvement to yield predictable results;
- choosing from a finite number of identified, predictable solutions, with a reasonable expectation of success (in which case a claim would have been obvious to try);
- known work in one field of endeavor may prompt variations of it for use in either the same field or a different one based on design incentives or other market forces if the variations would have been predictable to one of ordinary skill in the art; and
- some teaching, suggestion, or motivation in the prior art that would have led one of ordinary skill to modify the prior art reference or to combine prior art reference teachings to arrive at the claimed invention.

22. I understand that “secondary considerations” must be considered as part of the obviousness analysis when present. I further understand that the secondary considerations may include: (1) a long-felt but unmet need in the prior art that was satisfied by the claimed invention; (2) the failure of others; (3) skepticism by experts; (4) commercial success of a product covered by the patent; (5) unexpected results achieved by the claimed invention; (6) industry praise of the claimed invention; (7) deliberate copying of the invention; and (8) teaching away by others. I also understand that evidence of the independent and nearly simultaneous “invention” of

the claimed subject matter by others is a secondary consideration supporting an obviousness determination and may support a conclusion that a claimed invention was within the knowledge of a person of ordinary skill as of June 2007. I am not aware of any evidence of secondary considerations that would suggest that the claims of the '357 patent would have been nonobvious in June 2007.

23. I understand that when assessing obviousness, using hindsight is impermissible; that is, what is known today or what was learned from the teachings of the patent should not be considered. The patent should not be used as a road map for selecting and combining items of prior art. Rather, obviousness must be considered from the perspective of a person of ordinary skill at the time the alleged invention was made—which in this case, I have been asked to assume is June 2007.

24. I also understand that an obviousness analysis must consider the invention as a whole, as opposed to just a part or element of the invention. I understand this “as a whole” assessment to require showing that one of ordinary skill in the art at the time of invention, confronted by the same problems as the inventor and with no knowledge of the claimed invention, would have selected the elements from the prior art and combined them in the claimed manner.

III. PERSON OF ORDINARY SKILL IN THE ART

25. It is my understanding that when interpreting the claims of the '357 patent and evaluating whether a claim would have been obvious, I must do so based

on the perspective of a person of ordinary skill in the art at the time of invention. I understand that the earliest claimed priority date of the '357 patent is June 13, 2007, and I have been instructed to assume for the purposes of my opinions that this is the invention date for the '357 patent.

26. I understand that in determining the level of ordinary skill in the art, several factors are considered. Those factors may include: (i) the type of problems encountered in the art; (ii) prior art solutions to those problems; (iii) the rapidity with which innovations are made; (iv) the sophistication of the technology; and (v) the educational level of active workers in the field. A person of ordinary skill in the art must have the capability of understanding the scientific and engineering principles applicable to the pertinent art.

27. Based on my review of the specification and claims of the '357 patent, it is my opinion that a person of ordinary skill in the art 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. Work experience could substitute for formal education, and additional formal education could substitute for work experience.

28. My conclusions below that the claims of the '357 patent would have been obvious would remain the same even if the invention date, field of endeavor, or level of ordinary skill were slightly different.

29. I meet the above definition of a person of ordinary skill in the art, and did so as of June 2007. Also, I have worked with persons of ordinary skill in the art through my professional and academic experiences, and I have an understanding of their skill level around June 2007.

IV. TECHNOLOGY BACKGROUND AND OVERVIEW OF PRIOR ART

A. [1982] Griffiths-Jim: The GSC / Griffiths-Jim Beamformer

30. One noise-reduction technique known for decades before 2007 was the Generalized Sidelobe Canceler, or GSC. The GSC is sometimes also called the Griffiths-Jim beamformer.

31. The GSC was described in a paper by Lloyd Griffiths and Charles Jim titled *An Alternative Approach to Linearly Constrained Adaptive Beamforming*, included in the January 1982 volume of IEEE Transactions on Antennas and Propagation. I will refer to this paper as “**Griffiths-Jim**” and I understand it is Exhibit 1005. I have been teaching the Griffiths-Jim algorithm for decades (including well before 2006) in my CMU course 18-792 Advanced Digital Signal

Processing, working from the description in the standard text *Adaptive Signal Processing* by Widrow and Stearns, which was published in 1985.

32. Griffiths-Jim described “a simple time-varying beamformer which can be used to combine the outputs of an array of sensors.” (Ex. 1005, 27.) “The purpose of the beamformer is to minimize the effects of noise and interference at the array output” while capturing the target signal. (*Id.*, 27.) Griffiths-Jim called the beamformer a “generalized sidelobe canceling” structure or model. (*Id.*, 29.)

33. Figure 4 of Griffiths-Jim shows the GSC (highlighting and annotations added):

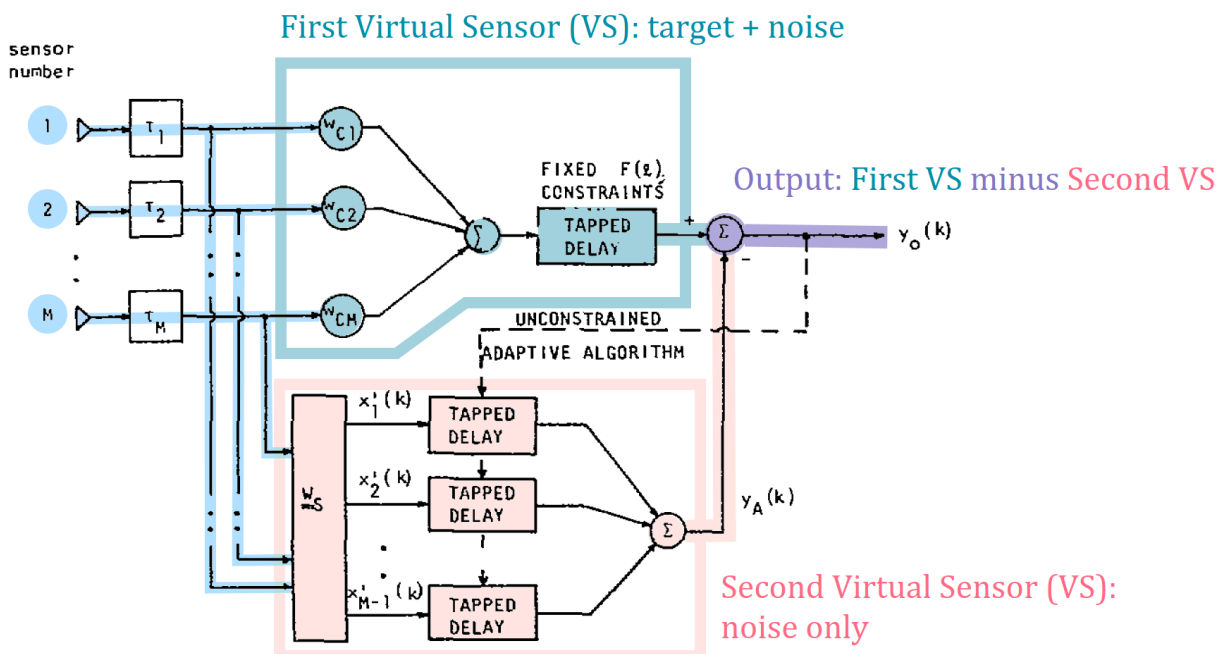


Fig. 4. Generalized sidelobe canceling form of linearly constrained adaptive array processing algorithm.

(*Id.*, 30 (annotated).)

34. As shown in Figure 4, the signal processor had two main substructures: the top branch was a “conventional beamformer” designed to capture the target signal plus noise, and the bottom branch was the “sidelobe canceling path” that would be used to capture only noise so that the noise could be subtracted or canceled out.

35. I will refer to the top branch as the **first virtual sensor**. The first virtual sensor was formed by combining the outputs of the **array sensors**. The sensor outputs were combined by multiplying the sensor output signals by “weights” $w_{c1}, \dots, w_{cM}$ (sometimes also called “gains”) and summing the weighted sensor signals to produce a “nonadaptive-beamformed signal $y_c(k)$.” (*Id.*, 29.) (The variable k in Griffiths-Jim’s equations is time expressed in samples. (*E.g., id.*, 27 (“We denote the sampled output of the m th time-delayed sensor by $x_m(k)$.”), 28 (“We let $\mathbf{A}_l$ and $\mathbf{X}(k - l)$ represent the filter coefficient and signal vectors at the l th delay point”).) Griffiths-Jim’s discussion is thus in the time domain, which refers to analyzing signals as a function of time.) The output of the first virtual sensor (top branch) is $y_c'(k)$, which “is obtained by filtering $y_c(k)$ and the FIR operator containing . . . constraint values $f(l)$ ” (*id.*, 30):

$$y_c'(k) = \sum_{l=-K}^K f(l)y_c(k-l). \quad (25)$$

(*Id.*, 30.) This output $y_c'(k)$ contains the target signal plus noise.

36. “The lower path in Fig. 4 is the sidelobe canceling path” (*id.*, 30), which I will refer to as the **second virtual sensor**. The second virtual sensor is formed from a combination of the outputs of the array sensors that differs from the combination used for the first virtual sensor.

37. The lower path includes a blocking matrix designed to “block the desired signal $s(k)$ from the lower path.” (*Id.*, 30.) In Griffiths-Jim, the blocking matrix is denoted by $\overline{W}_s$. (*Id.*, 30.)

38. In the lower path, the array sensor outputs were combined by applying the blocking matrix and adaptive filter coefficients and summing the result:

$$y_A(k) = \sum_{l=-K}^K [A_l'(k)]^T \mathbf{X}'(k-l), \quad (28)$$

where

$$\mathbf{X}'(k) = \overline{W}_s \mathbf{X}(k). \quad (26)$$

(*Id.*, 30.)

39. The second virtual sensor’s output “ $y_A(k)$ contains no desired signal terms” and instead “contains only noise and interference terms.” (*Id.*, 30.) In other words, the second virtual sensor is designed to capture noise only.

40. The overall output of the GSC, $y(k)$, was produced by subtracting the noise-only output of the second virtual sensor from the target-plus-noise output of the first virtual sensor:

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

(*Id.*, 30.) The result was a signal with reduced noise. The response of the processor to the desired signal $s(k)$ ¹ is that produced only by $y_c'(k)$,” i.e., only by the first virtual sensor (top branch). (*Id.*, 30.) The “output due to the presence of only the desired signal satisfies the constraint defined by (9)” (*id.*, 30), which Griffiths-Jim explains is commonly made to be a constraint of “zero distortion” (*id.*, 28).

B. [1982-c.2001] Use of GSC with Microphone Arrays for Speech Applications

41. After Griffiths and Jim’s GSC article, over the next two decades, the GSC was used in many signal-processing applications, including with microphone arrays to reduce noise in speech applications.

42. I understand Exhibit 1007 is U.S. Patent No. 5,651,071 (“**Lindemann**”). Filed in 1993, Lindemann cites the Griffiths-Jim article and explains that using a Griffiths-Jim beamformer “to improve signal-to-noise ratio for hearing aids” was known. (Ex. 1007, 1:40-46, 12:12-14.¹)

¹ Patent citations are in the form of *column:line*.

43. I understand Exhibit 1008 is U.S. Patent No. 5,627,799 (“**Hoshuyama**”). Filed in 1995, Hoshuyama “relates generally to interference cancelers, and more particularly to a generalized sidelobe canceler, or adaptive beamformer for an array of sensors such as microphones and the like.” (Ex. 1008, 1:8-11.) Describing the prior art at the time, Hoshuyama states:

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 W. Jim, the IEEE Transactions on Antenna and Propagation, Vol. AP-30, No. 1, January 1982, pages 27-34.

(*Id.*, 1:17-26.)

C. [2001] Brandstein

44. In 2001, Springer-Verlag published the reference book MICROPHONE ARRAYS: SIGNAL PROCESSING TECHNIQUES AND APPLICATIONS (“**Brandstein**”), edited by Michael Brandstein of Harvard University and Darren Ward of Imperial College. In 2001, Springer-Verlag was already a well-established publishing house and a leading publisher of technical monographs and serial publications consulted by academics, scientists, engineers, and other practitioners across a wide spectrum of technical disciplines, including audio-signal processing. I purchased a copy of

Brandstein before the end of 2002, and Exhibit 1003 is a scanned copy of excerpts from my copy of Brandstein.

45. Following its publication, Brandstein quickly became a standard reference for those working in the field of audio signal processing.

46. Chapter 5 of Brandstein concerns robust adaptive beamforming, and at the beginning of that chapter 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. 1003 (Brandstein), 87.) Brandstein further notes: “Among various adaptive beamformers, the Griffiths-Jim beamformer (GJBF), or the generalized sidelobe canceler, is most widely known.” (*Id.*, 88 (internal citation omitted).)

[continued on next page]

47. As Brandstein states, “Figure 5.1 depicts the structure of the GJBF.”

(*Id.*, 88.) Figure 5.1 appears below with highlighting and annotations:

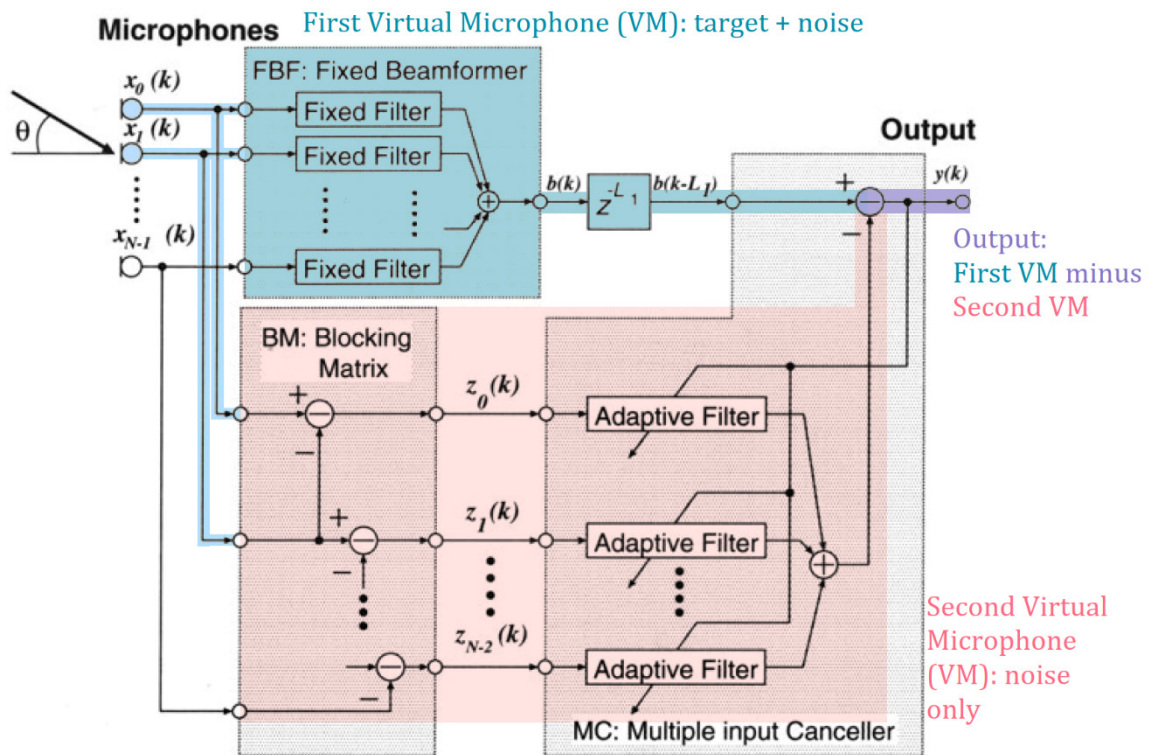


Fig. 5.1. Griffiths-Jim beamformer. It comprises a fixed beamformer (FBF), a multiple-input canceller (MC), and a blocking matrix (BM).

(*Id.*, 89, Fig. 5.1 (annotated).)

48. As illustrated in the top branch of Figure 5.1, the signals of at least **two physical microphones**, $x_0(k)$ and $x_1(k)$, are combined by filtering and summing the signals to form a fixed beamformer that is a **first virtual microphone**. The first virtual microphone captures the target speech signal plus noise. As in the case of Griffiths-Jim, the variable k is time expressed in samples. Brandstein’s discussion is thus in the time domain.

49. In the bottom branch of Figure 5.1, the signals of the **two physical microphones** are combined by filtering and summing the signals in a different way to form a **second virtual microphone**. The second virtual microphone includes a blocking matrix (BM). “[T]he BM forms a null in the look direction so that the target signal is suppressed and all other signals are passed through.” (*Id.*, 88.) “The BM was named after its function, which is to block the target signal.” (*Id.*) As a result, the second virtual microphone captures noise only.

50. The **output** of the GSC (or GJBF) is the output of the top branch (first virtual microphone) minus the output of the bottom branch (second virtual microphone). The result is that, “in the subtracter output $y(k)$, the target signal is enhanced and undesirable signals such as ambient noise and interferences are suppressed.” (*Id.*, 88-89.)

51. The two virtual microphones have very different responses to the target speech signal: the top branch is designed to capture the target signal, while the bottom branch is designed to block it. In contrast, they have similar responses to noise so that in the final subtraction output the noise is removed. Figure 5.2 illustrates this. Figure 5.2 shows the directivity pattern for the final output of an example Griffiths-Jim beamformer:

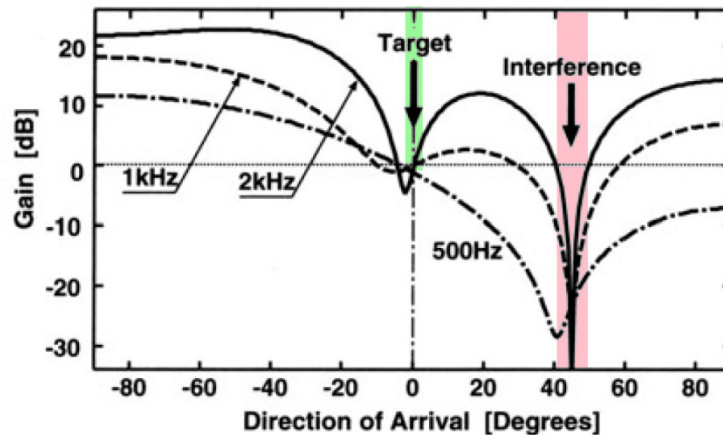


Fig. 5.2. Example directivity pattern of the Griffiths-Jim beamformer.

(*Id.*, 89, Fig. 5.2 (highlighting added).) The horizontal axis of the graph shows direction of arrival measured in degrees relative to the microphone array. The target signal is shown at 0 degrees, while interference (noise) is shown at approximately 45 degrees. The vertical axis of the graph shows gain in decibels (dB). Zero gain in decibels corresponds to no change in sound pressure or signal power, a positive gain in decibels corresponds to an increase in signal power, and a negative gain in decibels corresponds to a decrease in signal power.

52. As highlighted in green in Figure 5.2: “In the direction of the target signal, almost constant gains close to 0 dB are obtained over a wide range of frequencies.” (*Id.*, 90.) Subtracting the bottom branch’s noise-only output from the top branch’s target-plus-noise output will therefore produce the target signal in the direction of the target, as discussed above for Griffiths-Jim. (§40 above.)

53. “On the contrary,” as highlighted in red in Figure 5.2, “in the direction of the interference, a deep null is formed.” (Ex. 1003 (Brandstein), 90.) This reflects that subtracting the two virtual microphones’ outputs cancels the noise.

D. [2001] Gannot

54. The August 2001 edition of IEEE Transactions on Signal Processing (vol. 49, no. 8) included an article titled *Signal Enhancement Using Beamforming and Nonstationarity with Applications to Speech* (“**Gannot**”) by Sharon Gannot, David Burshtein, and Ehud Weinstein. I was a subscriber to the IEEE Transactions on Signal Processing and would have received my copy of that article at that time. I understand Gannot is Exhibit 1004.

55. Gannot addresses a sensor array “where arbitrary transfer functions (TFs) relate the source signal and the sensors.” (*Id.*, 1614 (Abstract).) As an audio signal travels from its source to a microphone, the signal may change, such that the signal received at the microphone is not exactly the same as the signal when it originated from the source. The acoustical path from the source to the microphone can be thought of as a system that brings about this change, and the operation of the acoustical path on the signal can be represented mathematically by a *transfer function*.

56. Gannot explains that the generalized sidelobe canceler (GSC) relies on certain assumptions about the acoustical paths’ transfer functions, such as that the

signals received at the sensors “are simple delayed versions of the source signal.” (*Id.*, 1614 (Abstract).) When these assumptions hold, the GSC attains “good interference suppression.” (*Id.*, 1614 (Abstract).) But the original Griffiths-Jim GSC may suppress interference poorly “in complicated acoustic environments, where arbitrary TFs [transfer functions] may be encountered.” (*Id.*, 1614 (Abstract).) Gannot thus proposes a GSC solution adapted to handle arbitrary transfer functions. (*Id.*, 1614 (Abstract).)

57. In Gannot, Z_m corresponds to sensor (microphone) m , and A_m is the transfer function of the acoustical path from the desired speech source to microphone m .

58. Gannot’s Figure 3 shows the proposed GSC structure:

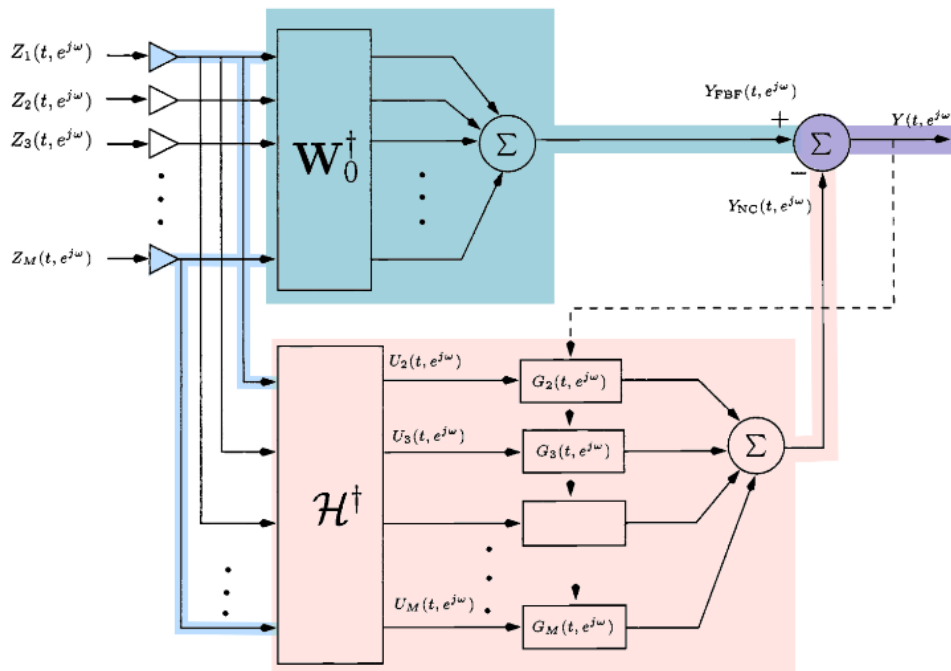


Fig. 3. Linearly constrained adaptive beamformer.

(*Id.*, Fig. 3 (highlighting added).)

59. Figure 4 summarizes Gannot's proposed algorithm:

- 1) TF-s ratios: $\mathbf{H}(e^{j\omega}) = \frac{\mathbf{A}(e^{j\omega})}{A_1(e^{j\omega})}$
- 2) Fixed beamformer:

$$Y_{\text{FBF}}(t, e^{j\omega}) = \mathbf{W}_0^\dagger(e^{j\omega}) \mathbf{Z}(t, e^{j\omega})$$
- 3) Noise reference signals:

$$\mathbf{U}(t, e^{j\omega}) = \mathcal{H}^\dagger(e^{j\omega}) \mathbf{Z}(t, e^{j\omega})$$
- 4) Output signal:

$$Y(t, e^{j\omega}) = Y_{\text{FBF}}(t, e^{j\omega}) - \mathbf{G}^\dagger(t, e^{j\omega}) \mathbf{U}(t, e^{j\omega})$$
- 5) Filters update, for $m = 1, \dots, M - 1$:

$$\tilde{G}_m(t + 1, e^{j\omega}) = G_m(t, e^{j\omega}) + \mu \frac{U_m(t, e^{j\omega}) Y^*(t, e^{j\omega})}{P_{\text{est}}(t, e^{j\omega})}$$

$$G_m(t + 1, e^{j\omega}) \stackrel{\text{FIR}}{\leftarrow} \tilde{G}_m(t + 1, e^{j\omega})$$

where,

$$P_{\text{est}}(t, e^{j\omega}) = \rho P_{\text{est}}(t - 1, e^{j\omega}) + (1 - \rho) \sum_m |Z_m(t, e^{j\omega})|^2$$
- 6) keep only non-aliased samples.
 (note: $\mathbf{W}_0(e^{j\omega})$ is defined in (16).
 $\mathcal{H}(e^{j\omega})$ is defined in (19)).

Fig. 4. Suggested algorithm.

(*Id.*, Fig. 4.)

60. In Gannot's GSC, the blocking matrix is denoted $\mathcal{H}^\dagger$ and it is used to create noise reference signals U_m that apply the linear transfer functions A_m of the acoustical paths:

$$U_m(e^{j\omega}) = Z_m(t, e^{j\omega}) - \frac{A_m(e^{j\omega})}{A_1(e^{j\omega})} Z_1(t, e^{j\omega})$$

$$m = 2, \dots, M. \quad (20)$$

(*Id.*, 1618-20.)

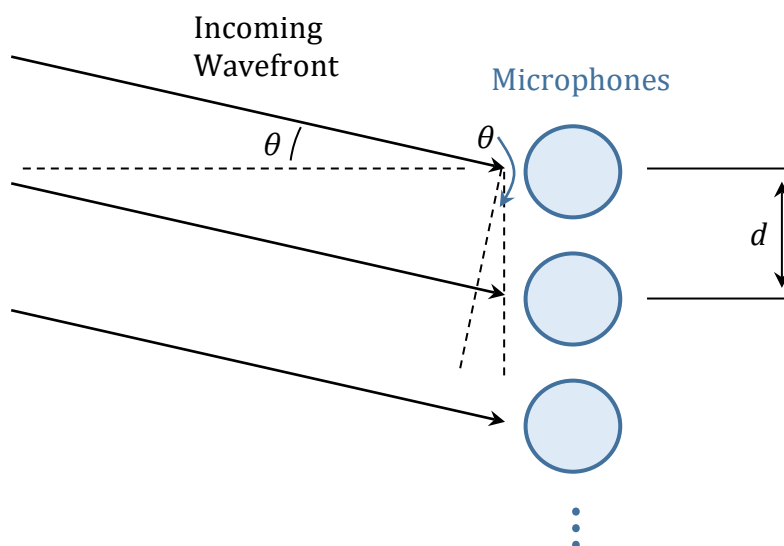
61. Gannot's more-general GSC thus accounts for source-to-microphone acoustical paths with arbitrary transfer functions. In particular, by using the ratio of the transfer function of the acoustical path to microphone m to the transfer function of the acoustical path to microphone 1, Gannot's algorithm develops a transfer function for an acoustical path between the two microphones. For each m greater than 1, the transfer function is incorporated into the blocking matrix, which is a critical component of the GSC and prevents the target signal from appearing in the secondary channel. By applying the blocking matrix to the microphone signals, Gannot applies the transfer functions. Gannot's noise reduction thus relies on the transfer functions between the various microphones.

E. [2001-2002] Brandstein and McCowan Discuss Near-Field Design

62. Adaptive beamforming typically relies on the spatial geometry of the array and the signal's direction of arrival. (Ex. 1003 (Brandstein), 87.)

63. Sound waves from a point source typically propagate through air as spherical waves. The amplitude of the wave is inversely proportional to the distance traveled, meaning the amplitude decays as the wave travels. When a source is sufficiently far away from a microphone, by the time the signal reaches the microphone, the radius of the sphere is large enough that the wavefront arriving at the microphone can be accurately approximated by a plane. Additionally, if the

distance between microphones in an array is sufficiently small compared to their mutual distance to the point source, the amplitudes of the signals received by each microphone in the array can be approximated as being identical. Such a scenario is referred to as a “far-field” configuration. A far-field configuration for a particular microphone-array application might thus assume that the signal as captured at one microphone of the array does not differ from the signal captured at another microphone of the array in amplitude, and instead differs only by a phase shift (or time delay) corresponding to the extra distance the signal had to travel from the farther microphone. This distance is easy to calculate because it reduces to straightforward trigonometry, as illustrated below:



If two microphones are separated by a distance d , then, assuming the wavefront is a plane, the signal travels an extra distance $d \sin \theta$ to one microphone, where θ is the

angle of arrival. To determine the relative delay in units of time, one divides that difference in distance by the speed of sound.

64. In contrast, in some applications the source of the target signal is too close for the far-field approximation to work well. In such cases, it was well known by 2001, that the design of the signal processor would need to more accurately account for amplitude and phase differences in the signals received by the array microphones resulting from the spherical shape of the wavefront. By 2001, the propagation of sound in the “near field” was well understood, along with methods to perform beamforming with sound sources in the near field. By 2001, I had worked on projects with near-field applications of microphone arrays; for these projects, the target-signal components of the microphone signals needed to be time-aligned, and I was well aware that the far-field approximation would not be appropriate for these near-field applications.

[continued on next page]

65. Brandstein discloses the standard near-field design, using c to denote the speed of sound:

- 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.

(*Id.*, 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. The τ_i terms above thus provide the difference in time it takes for the signal to propagate from the source to microphone i relative to a reference microphone. Including the τ_i term in the complex exponential shown above and applying that exponential to the signal received at the reference microphone represents delaying the received signal 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 . In other words, the signal

x_i on microphone i can be expressed in terms of the signal x_{ref} on the reference microphone as follows (where $j = \sqrt{-1}$):

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

66. The reasoning behind these equations 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*. (Ex. 1006 (“**McCowan**”), 89-90.) McCowan expressly motivates and teaches incorporating the near-field design into the GSC. (E.g., *id.*, 90-96.)

67. Similar to Brandstein, McCowan defines a near-field propagation vector:

$$\mathbf{d}(f, r, \theta, \phi) = [\alpha_1 e^{-j2\pi f \tau_1} \dots \alpha_i e^{-j2\pi f \tau_i} \dots \alpha_N e^{-j2\pi f \tau_N}]^T. \quad (6)$$

(*Id.*, 90.)

68. Based on this near-field propagation vector, McCowan introduces a near-field compensation unit (highlighted below) into the standard GSC:

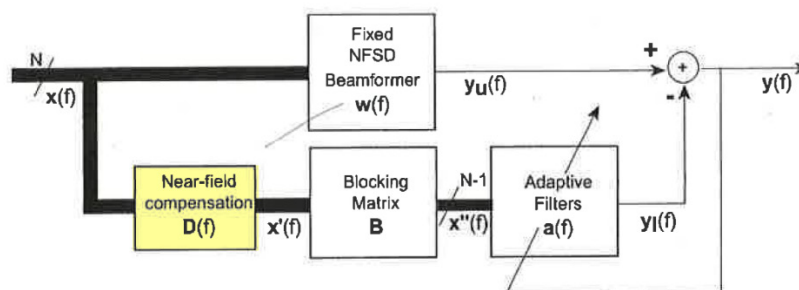


FIG. 3. Near-field adaptive beamformer.

69. McCowan states: “To ensure complete blocking [by the blocking matrix], the desired signal must both be time aligned and have equal amplitudes across all channels.” (*Id.*, 93.) McCowan’s near-field compensation unit achieves this alignment and equalization. McCowan defines a diagonal matrix:

$$\mathbf{D}(f) = [\text{diag}(\mathbf{d}(f))]^{-1}, \quad (16)$$

where $\mathbf{d}(f)$ is the near-field propagation vector from Eq. (6). In this paper we define the diagonal operator, $\text{diag}()$, to produce a diagonal matrix from a vector parameter. Conversely, if invoked with a matrix parameter, it produces a row

(*Id.*, 93 (text faded).) “The near-field compensation can be applied as”

$$\mathbf{x}'(f) = \mathbf{D}(f)\mathbf{x}(f). \quad (17)$$

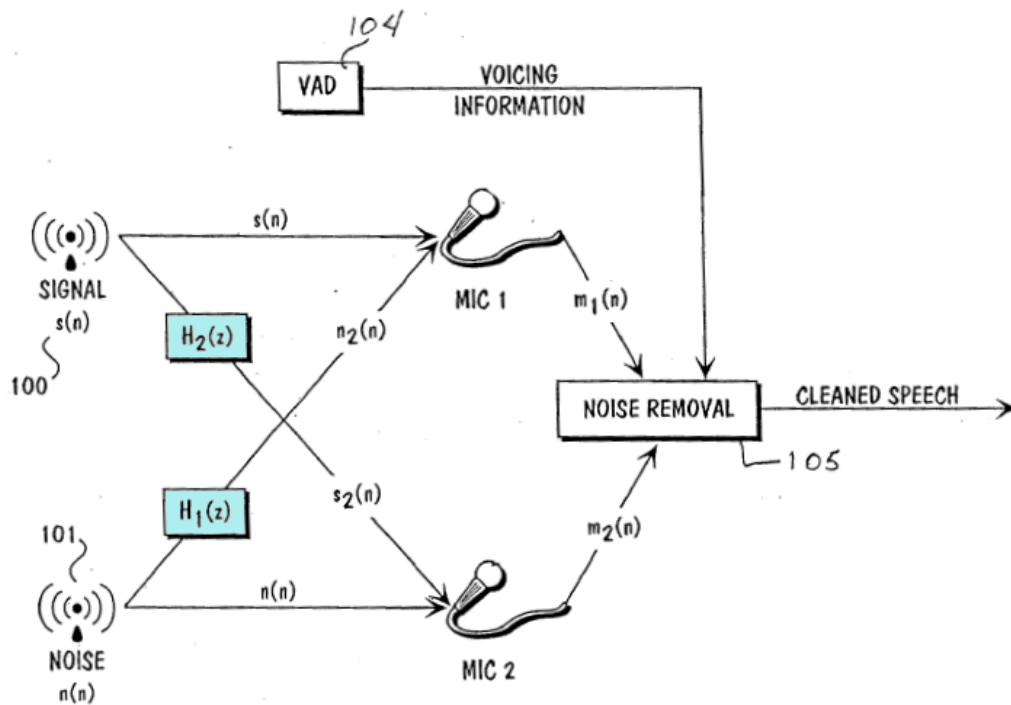
(*Id.*, 93.) Since McCowan defines τ_i to be the negation of Brandstein’s τ_i , applying $\mathbf{D}(f)$ as in equation 17 results in delaying microphone signal x_i by Brandstein’s τ_i (applying an exponential with (Brandstein’s) $-\tau_i$ in the exponent) and scaling the microphone signal’s amplitude by $1/a_i$ (McCowan’s α_i is the same as Brandstein’s a_i).

70. McCowan states: “Once this near-field compensation has been performed, a standard GSC blocking matrix can be employed to block the desired signal from the adaptive path.” (*Id.*, 93.)

F. [2003] Burnett '848

71. I understand Exhibit 1009 is U.S. Patent Publication No. 2003/0128848, which I will refer to as “**Burnett '848**,” since the named inventor is Gregory Burnett. I understand this is the same individual named as inventor on the '357 patent.

72. Figure 2 of Burnett '848 shows a two-microphone system:



(Ex. 1009, Fig. 2 (highlighting added).) In the figure, H_1 denotes the transfer function of an acoustical path from a noise source (101) to the first microphone, while H_2 denotes the transfer function of an acoustical path from the source of the target signal (100) to the other microphone. (*Id.*, [0022].) Burnett '848 also states: “The transfer functions from the signal source 100 to MIC 1 and from the noise

source 101 to MIC 2 are assumed to be unity. . . . The assumption of unity transfer functions does not inhibit the generality of this algorithm, as the actual relations between the signal, noise, and microphones are simply ratios and the ratios are redefined in this manner for simplicity.” (*Id.*, [0022].)

73. Using M_1 and M_2 to denote the input to microphones 1 and 2 respectively, S to denote the target signal, and N to denote the noise signal, Burnett ’848 discloses that the following equations hold:

$$\begin{aligned}M_1(z) &= S(z) + N(z)H_1(z) \\ M_2(z) &= N(z) + S(z)H_2(z)\end{aligned}$$

(*Id.*, [0024], Eq. 1.) Burnett ’848 states: “This is the general case for all two microphone systems.” (*Id.*, [0025].)

74. Rearranging terms, Burnett ’848 identifies the following formula for recovering the target signal S :

$$S(z) = \frac{M_1(z) - M_2(z)H_1(z)}{1 - H_2(z)H_1(z)}.$$

(*Id.*, [0033], Eq. 2.) Burnett ’848 thus teaches that knowing H_1 and H_2 can help recover S . As Burnett ’848 puts it: “If the transfer functions $H_1(z)$ and $H_2(z)$ can be described with sufficient accuracy, then the noise can be completely removed and the original signal recovered.” (*Id.*, [0034].)

75. Burnett '848 discloses that it is appropriate to estimate these transfer functions by making certain assumptions that allow the equations to be simplified. For example, if speech (target signal) is occurring, Burnett '848 states it is appropriate to estimate noise (N) as zero, resulting in:

$$H_2(z) = \frac{M_{2s}(z)}{M_{1s}(z)}$$

where the s subscript indicates the assumption of a speech-only scenario. (*Id.*, [0030].) Similarly, H_1 can be estimated by the inverse ratio in an assumed noise-only scenario. (*Id.*, [0026]-[0027], [0030], [0031].) Burnett '848 further teaches that it is reasonable to assume that “that when one of $H_1(z)$ and $H_2(z)$ are being calculated the other does not change substantially” (*id.*, [0034]), which allows one to use the ratios shown above to estimate the transfer functions (*id.*, [0031]). Burnett '848 thus discloses estimating H_1 and H_2 using transfer functions of acoustical paths between the two microphones.

V. THE '357 PATENT

76. The '357 patent is directed to forming virtual microphones from a dual omnidirectional microphone array. Figure 2 illustrates an array:

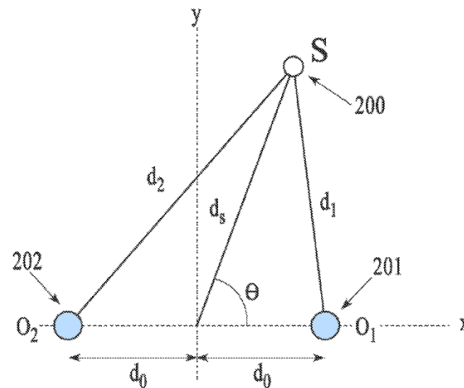


FIG.2

(Ex. 1001, Fig. 2 (highlighting added).) The figure shows two physical microphones O_1 and O_2 (201 and 202). (Ex. 1001, 8:40-42.) The source S (200) of the target signal is also pictured. (*Id.*)

77. Figure 3 illustrates combining the two physical microphones to create a virtual microphone.

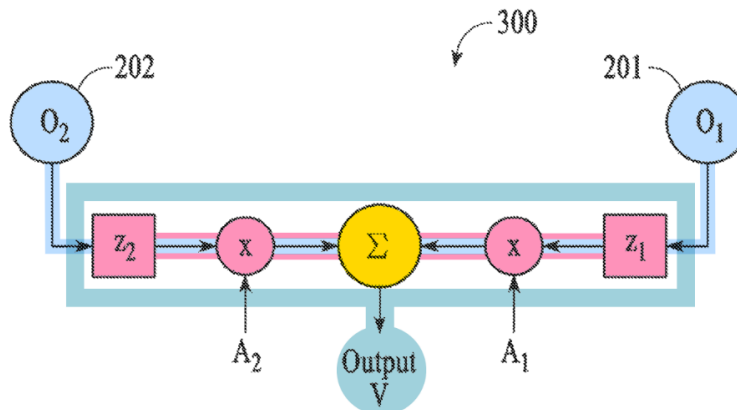


FIG.3

(*Id.*, Fig. 3 (highlighting added).) “The **output from each microphone** 201 and 202 can be **delayed (Z_1 and Z_2)**, **multiplied by a gain (A_1 and A_2)**, and then **summed** with the other as demonstrated in FIG. 3.” (*Id.*, 8:50-53.) “The output of the array is or forms at least one virtual microphone” (*Id.*, 8:53-54.)

78. These signal-processing operations to form virtual microphones were well known in the art long before the '357 patent. The '357 patent acknowledges this: “By varying the magnitude and sign of the delays and gains, a wide variety of virtual microphones (VMs), also referred to herein as virtual directional microphones, can be realized. *There are other methods known to those skilled in the art for constructing VMs but this is a common one and will be used in the enablement below.*” (*Id.*, 8:55-61 (emphasis added).)

79. Figure 4 illustrates the same concept with two processing paths to form two virtual microphones, V_1 and V_2 , as shown to the right. (*Id.*, Fig. 4 (highlighting added).)

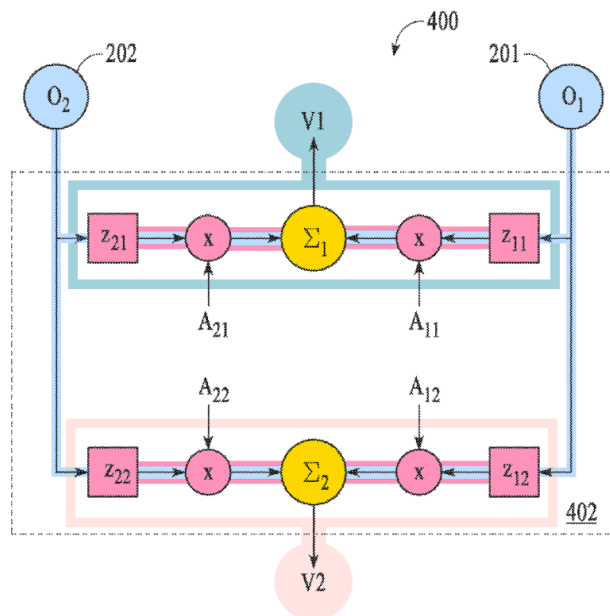


FIG.4

80. The '357 patent discloses using two virtual microphones, V_1 and V_2 , defined by the following formulas:

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

(Ex. 1001, 12:20); and

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

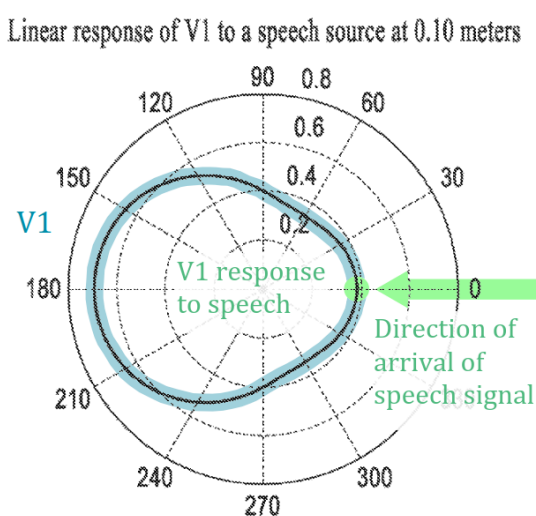
(*id.*, 11:8). In these formulas, O_1 and O_2 are the outputs of the two physical microphones, and:

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

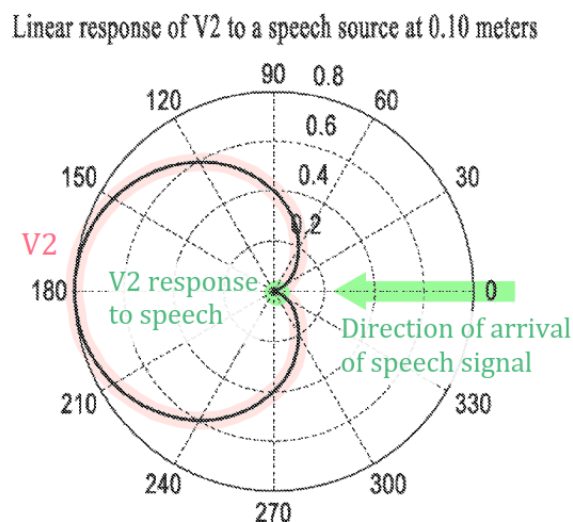
(*Id.*, 11:10-20.) “The distances d_1 and d_2 are the distance from O_1 and O_2 to the speech source (see FIG. 2), respectively, and γ is their difference divided by c , the speed of sound, and multiplied by the sampling frequency f_s .” (*Id.*, 11:19-22.) Note that γ is simply the difference in propagation time to the microphones taken from the standard near-field design equations (denoted τ_i in Brandstein; see ¶65, above), just 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 Brandstein; see ¶65, above).

81. According to the '357 patent, these formulas for V_1 and V_2 produce supposedly 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 Figures 9 and 11 (annotated):



Ex. 1001, Fig. 11



Ex. 1001, Fig. 9

82. As shown, V_1 exhibits a response to the speech signal, while V_2 has a null in the expected direction of the speech signal “and will therefore exhibit minimal response to the speech.” (*Id.*, 11:50-52, 12:39-43.) In contrast, the '357 patent explains that V_1 and V_2 will supposedly have “very similar” responses to noise originating from a meter away from the array. (*Id.*, 12:44-48.)

83. Figure 1 of the '357 patent is materially identical to Figure 2 of Burnett '848. Likewise, the mathematical discussion in the '357 patent from column 6 through column 7 is essentially the same as the discussion in Burnett '848 at paragraphs 20-34 and 54-55 disclosing estimating transfer functions with ratios of microphone signals representing transfer functions for acoustical paths between the microphones.

VI. CLAIMS 1-20 OF THE '357 PATENT WOULD HAVE BEEN OBVIOUS

A. Claims 1-20 Would Have Been Obvious in View of Brandstein and Gannot

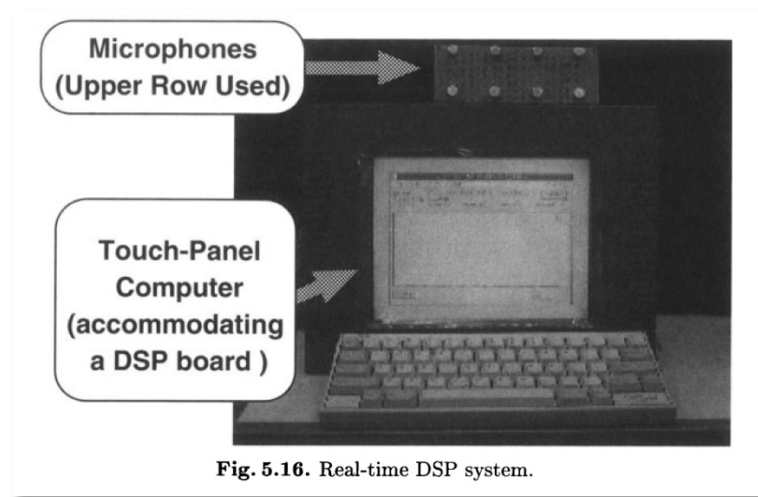
84. For at least the reasons I discuss below, Brandstein and Gannot render claims 1-20 obvious.

1. Claim 1

a. Preamble

85. The preamble of claim 1 recites: “A device, comprising”

86. Brandstein discloses a device with a microphone array and signal processing that implements the GSC. (Ex. 1003 (Brandstein), Fig. 5.1.) Brandstein also discloses the use of adaptive beamforming like the GSC in devices like hearing aids and telephones. (*Id.*, 87.) Brandstein also discusses the results of experiments performed with a real-world implementation of a GSC on a system illustrated in Figure 5.16:



(*Id.*, Fig. 5.16 (annotations in original).) “The system comprises a microphone array and a compact touch-panel personal computer which includes a floating point DSP, the ADSP-21062.” (*Id.*, 104 (internal citation omitted).) Brandstein therefore discloses a device. Alternatively, a person of ordinary skill in the art would have understood that Brandstein’s GSC shown in Figure 5.1 would have been implemented in a device, and therefore it would have been obvious to such a person.

**b. 1[a]: First Virtual Microphone
Comprising a Combination of Signals
from First and Second Physical
Microphones**

87. Claim element 1[a] recites “a first virtual microphone comprising 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.” Brandstein discloses this limitation.

88. The '357 patent states:

The term “virtual microphones (VM)” or “virtual directional microphones” means a microphone constructed using two or more omnidirectional microphones and associated signal processing.

(Ex. 1001, 6:7-10.) My conclusion that Brandstein discloses claim element 1[a] applies regardless of whether this statement is definitional.

89. Figure 5.1 of Brandstein appears below with annotations. **First and second physical microphones** generate **first and second microphone signals** (e.g., x_0 and x_1) and a **first virtual microphone** along the top branch of the GSC.² The first virtual microphone comprises a first combination of the microphone signals:

² Which virtual microphone is the “first” versus the “second” does not matter for claim 1. My analysis applies equally if the labels of “first” and “second” virtual microphone are switched with respect to the top and bottom branches of the GSC. For some dependent claims, the labeling may matter, as I discuss below.

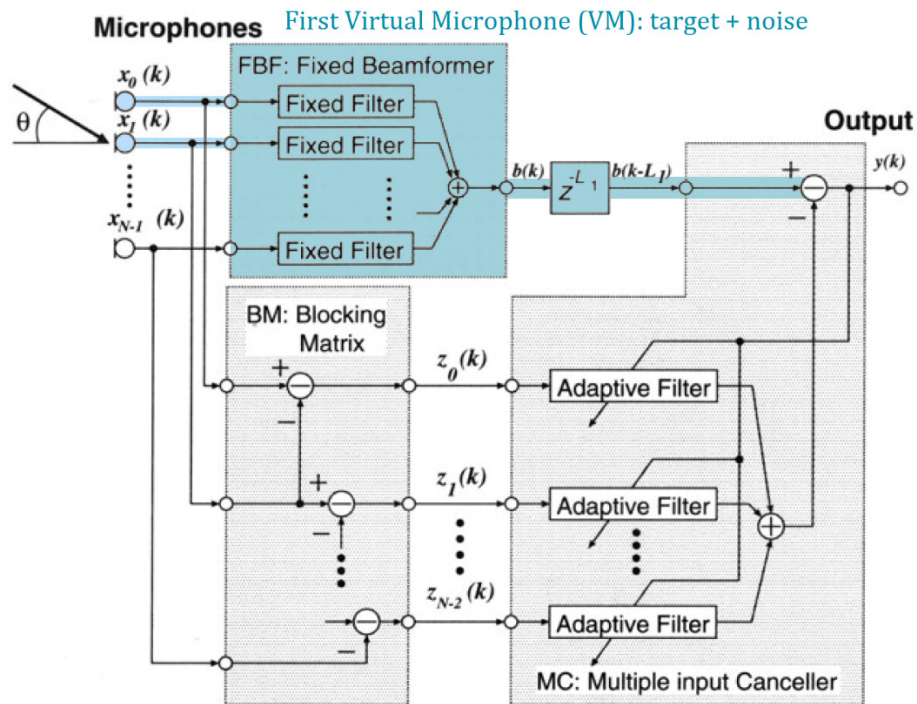


Fig. 5.1. Griffiths-Jim beamformer. It comprises a fixed beamformer (FBF), a multiple-input canceller (MC), and a blocking matrix (BM).

(*Id.*, 89, Fig. 5.1 (annotated).) Brandstein shows the top branch includes signal processing, including filtering and summing. Brandstein discloses that the physical microphones may be omnidirectional microphones. Describing an array used with the GSC, Brandstein states, “Four omni-directional microphones without calibration were mounted on a universal printed circuit board with an equal spacing of 4.1 cm.” (*Id.*, 101.) In any event, to the extent claim 1 requires the use of omnidirectional microphones, omnidirectional microphones were well known to persons of ordinary skill in the art and would have been an obvious choice to use because they typically are more cost-effective and easier to use than directionally sensitive microphones. Indeed, omnidirectional microphones would typically be used in implementing

arrays like those used for the GSC because omnidirectional microphones are generally the cheapest microphones and the easiest to analyze. Therefore, to the extent the statement quoted in paragraph 88 above is definitional, Brandstein discloses or renders obvious claim element 1[a].

c. 1[b][1]: Second Virtual Microphone

90. Claim element 1[b] begins by reciting: “a second virtual microphone comprising a second combination of the first microphone signal and the second microphone signal, wherein the second combination is different from the first combination.” Brandstein discloses this limitation. Again, my conclusion applies regardless of whether the statement in the ’357 patent that I quote in paragraph 88 above is definitional.

[continued on next page]

91. Figure 5.1 shows a **second virtual microphone** along the bottom branch of the GSC comprising a different combination of the **first and second microphone signals** (e.g., x_0 and x_1):

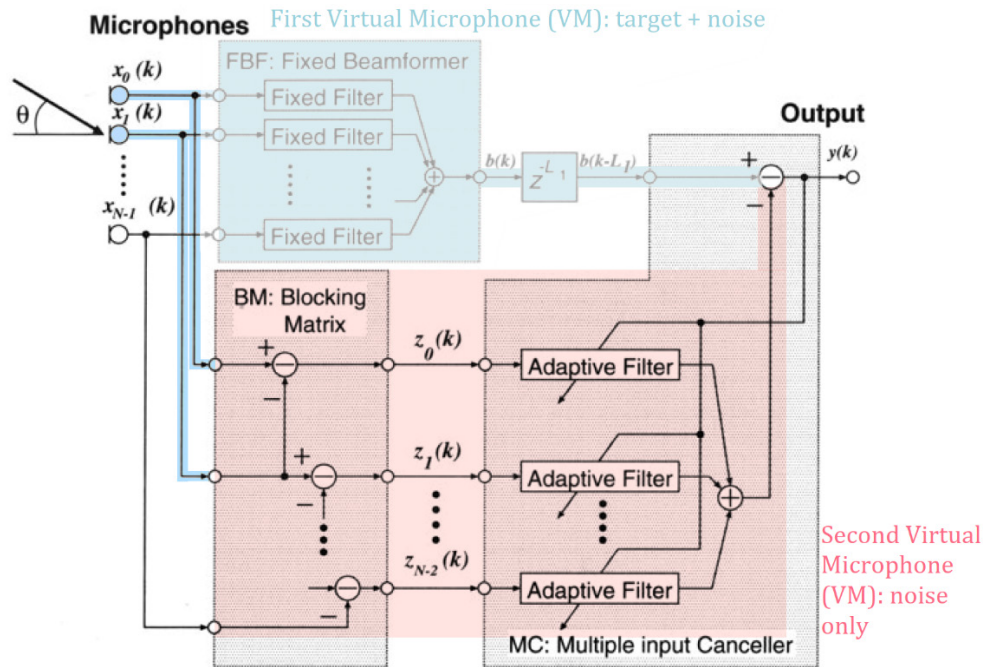


Fig. 5.1. Griffiths-Jim beamformer. It comprises a fixed beamformer (FBF), a multiple-input canceller (MC), and a blocking matrix (BM).

(*Id.*, 89, Fig. 5.1 (annotated).) Figure 5.1 shows the second virtual microphone includes signal processing, including summing/subtraction operations and adaptive filtering. And as discussed above for claim element 1[a], Brandstein discloses or renders obvious that the microphones can be omnidirectional.

d. 1[b][2]: Substantially Similar Responses to Noise and Substantially Dissimilar Responses to Speech

92. Claim element 1[b] further recites: “wherein the first virtual microphone and the second virtual microphone are distinct virtual directional

microphones with substantially similar responses to noise and substantially dissimilar responses to speech.” Brandstein discloses this limitation.

93. The ’357 patent states: “The term ‘speech’ means desired speech of the user.” (Ex. 1001, 5:63.) The ’357 patent also states: “The term ‘noise’ means unwanted environmental acoustic noise.” To the extent either statement is definitional, it does not change my analysis. To the extent I discuss the prior-art references’ disclosure of speech and it is not explicitly stated that the speech is desired speech of the user, it would have been obvious to one of ordinary skill in the art that the reference was discussing target speech of a user or that the reference’s teachings would apply to the fundamental use case of enhancing the speech of a user. Similarly, to the extent I discuss noise in connection with the prior art and it is not explicitly stated that the noise is unwanted environmental acoustic noise, it would have been obvious to one of ordinary skill in the art that the reference was discussing interfering environmental acoustic noise or that the reference’s teachings would apply to the fundamental use case of reducing interference from environmental acoustic noise. (This paragraph applies to all claims.)

94. The virtual microphone formed in the top branch of Brandstein’s Figure 5.1 is distinct from the virtual microphone formed in the bottom branch. If they were identical, the output of the GSC would be zero. The two virtual microphones have very different responses to the target signal: the first virtual microphone is designed

to form a beam that captures the target signal, while the second virtual microphone includes a blocking matrix designed to block the target signal. (Ex. 1003 (Brandstein), 88.) Brandstein discloses that the target signal can be speech. For example, in the set-up illustrated in Figure 5.13, the “target source was an English male speech signal.” (*Id.*, 101; *also id.*, 103, Figs. 5.13, 5.14.) And just before identifying the GSC as 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.” (*Id.*, 87.)

95. The two virtual microphones also have similar responses to noise so that in the final subtraction output the noise cancels out. The different responses to the target signal and similar responses to noise are evident from Figure 5.2, which shows the directivity pattern for the final output of an example Griffiths-Jim beamformer:

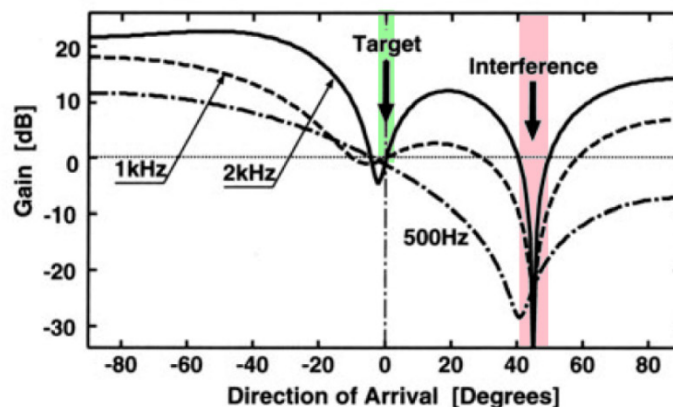


Fig. 5.2. Example directivity pattern of the Griffiths-Jim beamformer.

(*Id.*, 89, Fig. 5.2 (highlighting added).)

96. As highlighted in green, in the direction of the target signal, the Griffiths-Jim beamformer achieves “almost constant gains close to 0 dB.” (*Id.*, 90.) A gain of 0 dB means subtracting the second virtual microphone’s noise-only output from the first virtual microphone’s target-plus-noise output will produce the target signal in the direction of the target, reflecting the two virtual microphones’ different responses to the target signal. As Brandstein states, Figure 5.2 shows that the gains of essentially 0 dB “are obtained over a wide range of frequencies,” with 500 Hz, 1 kHz, and 2 kHz specifically shown. (*Id.*, 90.) Although frequencies between 500 Hz and 3 kHz are often focused on for purposes of speech signal processing, speech can include frequencies across a wide range, including at least between 50 Hz to 7 kHz.

97. As highlighted in red in Figure 5.2 above, “in the direction of the interference, a deep null is formed.” (*Id.*, 90.) The deep null reflects that subtracting the two virtual microphones’ outputs cancels the interference, which confirms that the virtual microphones’ responses to the interference are similar.

**e. 1[c][1]: Signal Processor Operative to
Combine Microphone Signals by Filtering
and Summing in the Time Domain**

98. Claim element 1[c] recites “a signal processor coupled with the first and second microphone signals and operative to combine the first and second

microphone signals by filtering and summing in the time domain.” Brandstein discloses this limitation.

99. Figure 3 the '357 patent illustrates a virtual microphone formed from **filtering** and **summing** physical-microphone signals:

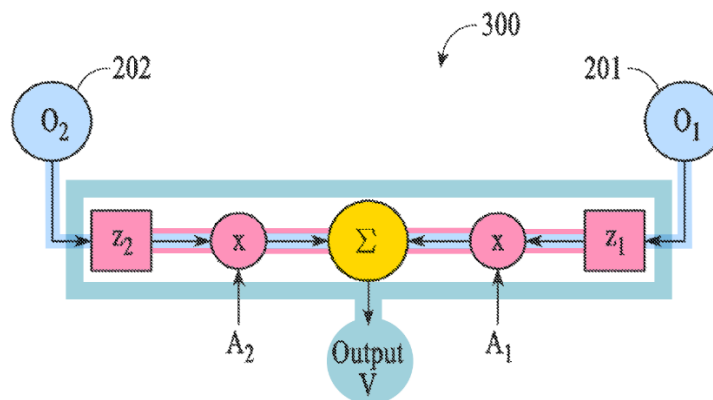
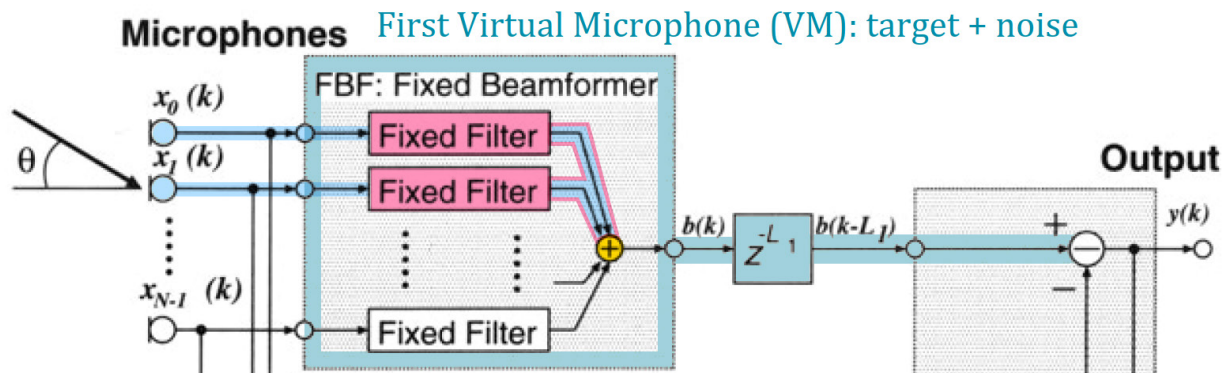


FIG.3

(Ex. 1001, Fig. 3 (highlighted).) Figure 4 illustrates the same concept with two processing paths to form two virtual microphones. (*Id.*, Fig. 4.) Consistent with Figures 3 and 4, in claim 1, each virtual microphone may be formed by the recited filtering and summing. The '357 patent acknowledges this was a “common” technique known in the art for constructing virtual microphones. (*Id.*, 8:55-61; see also ¶¶77-79 above.)

100. In Brandstein, the virtual microphones in the top and bottom branches of Figure 5.1 are both formed by a signal processor combining **two physical-microphone signals** by **filtering** and **summing**. The top branch is illustrative:



(*Id.*, 89, Fig. 5.1 (annotated excerpt).) As shown, the fixed beamformer combines the microphone signals in the top branch by applying fixed filters and summing. Since Figure 5.1 shows the microphone signals as functions of time (functions of k), it shows the filtering and summing as occurring in the time domain.

101. Even if Brandstein's Figure 5.1 had not disclosed filtering and summing *in the time domain*, it would have been obvious to one of ordinary skill in the art to implement the GSC's filtering and summing in the time domain because performing those operations in the time domain would be simply a design choice between two well-known options: the time domain or a transform domain (also called the "frequency domain"). Microphones output signals in the time domain, and the time domain has historically been the default domain in which to implement filters, so filtering and summing in the time domain would be a natural choice.

Lindemann also discloses using the GSC in a microphone array for hearing aids was a “time domain approach.” (Ex. 1007 (Lindemann), 1:51-52.)

102. It would have been obvious, if not inherent, that this processing of the microphone signals would be performed by a signal processor coupled with the signals. Brandstein’s title is *Microphone Arrays: Signal Processing Techniques and Applications*. As that title indicates, Brandstein focuses on signal processing with microphone arrays. And Brandstein discloses implementing a Griffiths-Jim beamformer (specifically, the GJBF-CNA (see pages 97-99)) on a system comprising “a microphone array and a compact touch-panel personal computer which includes a floating point DSP, the ADSP-21062.” (Ex. 1003 (Brandstein), 104 (internal citation omitted); *see also id.*, Fig. 5.16.) Here, “DSP” is an abbreviation for digital signal processor. One of ordinary skill in the art would have understood that some signal processor, such as the ADSP-21062, would perform the operations on the microphone signals shown in Figure 5.1 and that the processor was thus coupled to the microphone signals.

f. 1[c][2]: Applying a Varying Linear Transfer Function

103. Claim element 1[c] then recites the signal processor is operative to “apply a varying linear transfer function between the first and second microphone signals.”

104. Gannot discloses applying a varying linear transfer function between the first and second microphone signals. As discussed above (see paragraphs 54-61), Gannot sets forth a more-robust GSC solution that can handle source-to-microphone acoustical paths with arbitrary transfer functions.

105. The transfer functions Gannot addresses are linear and time-varying. (Ex. 1004 (Gannot), 1615-1617 (“time-varying TFs” and represented by linear systems).) To account for the arbitrary acoustical paths, Gannot incorporates the corresponding transfer functions into the GSC algorithm and applies them. For example, the blocking matrix $\mathcal{H}^\dagger$ is used to create noise reference signals U_m that apply the linear transfer functions A_m of the acoustical paths:

$$U_m(e^{j\omega}) = Z_m(t, e^{j\omega}) - \frac{A_m(e^{j\omega})}{A_1(e^{j\omega})} Z_1(t, e^{j\omega})$$
$$m = 2, \dots, M. \quad (20)$$

(*Id.*, 1618-20.) By incorporating these terms, Gannot’s more-general GSC accounts for arbitrary transfer functions. In particular, by using the ratio of the transfer function for microphone m to the transfer function for microphone 1, Gannot’s algorithm applies a time-varying linear transfer function for an acoustical path between the two microphones.

106. Gannot cites the original Griffiths-Jim GSC paper. (*Id.*, 1614 (citing reference [2]).) Gannot starts with the original GSC in a microphone-array setting

for speech applications, which is precisely what Brandstein discloses. Gannot then proposes a modification to apply a time-varying linear transfer function between microphone signals to arrive at a more-robust GSC. Gannot's discussion goes beyond merely proposing the concept of this more-robust GSC solution. Gannot sets forth the complete underlying algorithm and mathematics for the solution and also reports findings of a real-world implementation. (*See generally id.*) Gannot thus not only suggests modifying a standard microphone-array GSC as shown in Brandstein but also explains exactly how to perform the modification *and* demonstrates that the modification succeeds. A person of ordinary skill in the art would therefore be motivated to combine the teachings of Gannot and Brandstein and would have a reasonable expectation of success.

g. 1[c][3]: Generating an Output Signal with Reduced Noise

107. Claim element 1[c] further recites that the signal processor is operative to “generate an output signal having noise content that is attenuated with respect to speech content.” Brandstein discloses this limitation.

108. The overall **output** of Brandstein's GSC is the target-plus-noise output of the first virtual microphone minus the noise-only output of the second virtual microphone, as shown below:

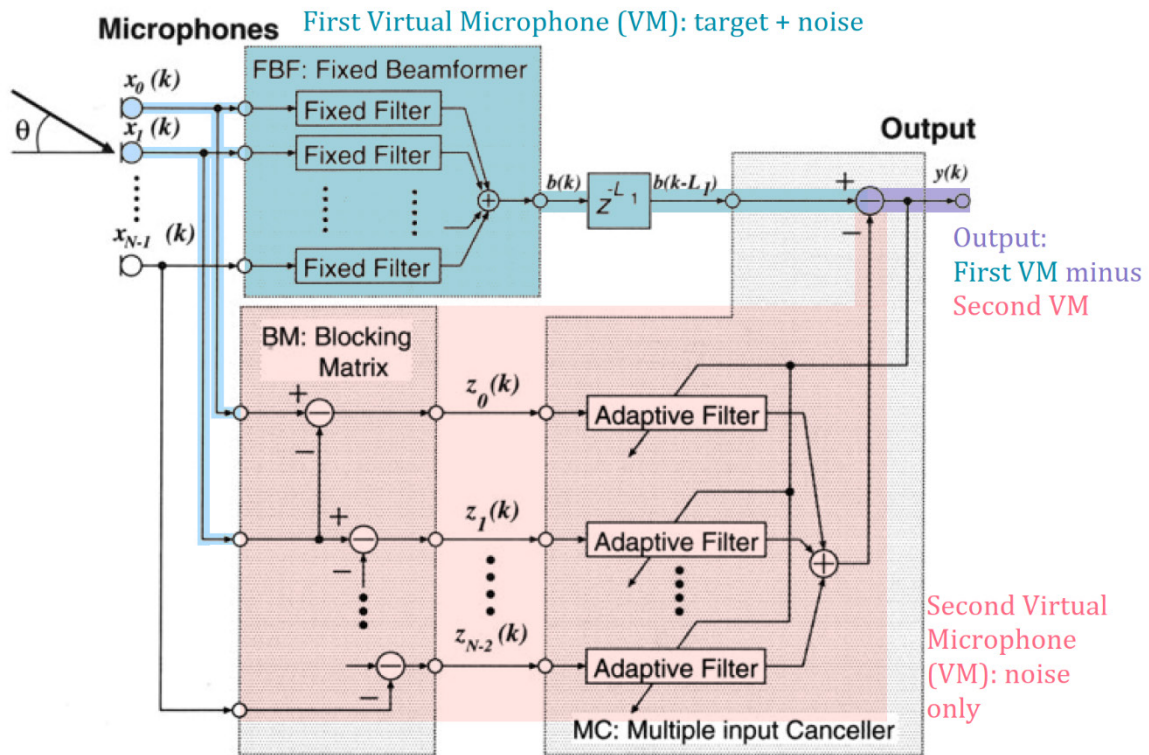


Fig. 5.1. Griffiths-Jim beamformer. It comprises a fixed beamformer (FBF), a multiple-input canceller (MC), and a blocking matrix (BM).

(Ex. 1003 (Brandstein), 89, Fig. 5.1 (annotated).)

109. The result is that, “in the subtracter output $y(k)$, the target signal is enhanced and undesirable signals such as ambient noise and interferences are suppressed.” (*Id.*, 88-89.)

h. Further Comments on Claim 1 in View of Brandstein and Gannot

110. In summary, both Brandstein and Gannot disclose a microphone-array GSC for reducing noise and enhancing speech signals. Brandstein discloses the original Griffiths-Jim GSC in this microphone-array context. Gannot proposes a more-robust GSC designed to handle source-to-microphone acoustical paths with

varying linear transfer functions. Gannot expressly teaches successfully modifying the original Griffiths-Jim GSC to apply ratios of transfer functions, with the ratios representing transfer functions of acoustical paths between microphones. Brandstein discloses all other limitations of claim 1. Because Gannot not only expressly proposes modifying the original GSC but explains exactly how to modify the original GSC and reports the successful results of a real-world implementation, claim 1 would have been obvious to one of ordinary skill in the art in view of the prior art.

2. Claim 2

111. Claim 2 depends from claim 1 and further recites “wherein the signal processor comprises one or more digital signal processors (DSPs).” Brandstein discloses using a DSP to implement a GSC. (Ex. 1003 (Brandstein), 104, 105, Fig. 5.16; *see also* ¶¶98-102 (claim element 1[c][1]), above.) Given Brandstein’s disclosure of using sampling with the GSC of Figure 5.1 (see use of discrete-time variable k and footnote 1 of Brandstein Chapter 5 discussing “ L_1 -sample delay”), it would have been obvious in any event to use a DSP. Thus, claim 2 would have been obvious.

3. Claim 3

112. Claim 3 depends from claim 1 and further recites “wherein the noise content comprises acoustic noise and the speech content comprises human speech.” Brandstein discloses this limitation.

113. Brandstein teaches that noise can be acoustic noise like “airconditioner [sic] noise.” (Ex. 1003 (Brandstein), 101.) Brandstein discloses that, simultaneously, the target speech signal can be human speech, like “an English male speech signal.” (*E.g., id.*) Figure 5.13 illustrates this, with the “White Noise” illustrating noise such as interference from an air conditioner:

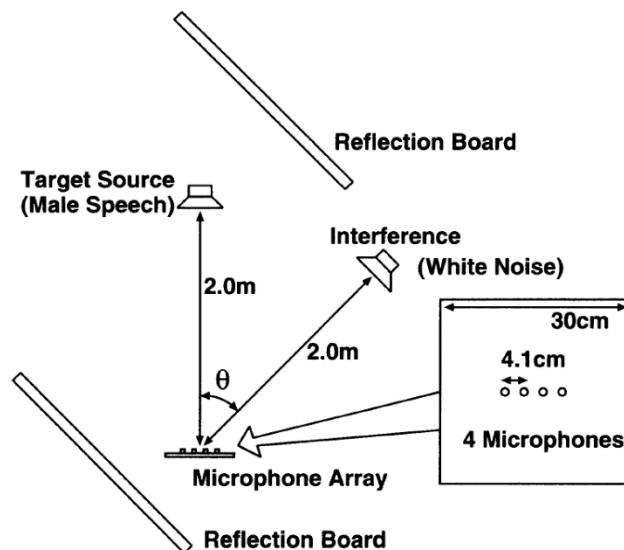


Fig. 5.13. Experimental set-up.

(*Id.*, Fig. 5.13; *see also id.*, Fig. 5.14.)

114. Thus, claim 3 would have been obvious.

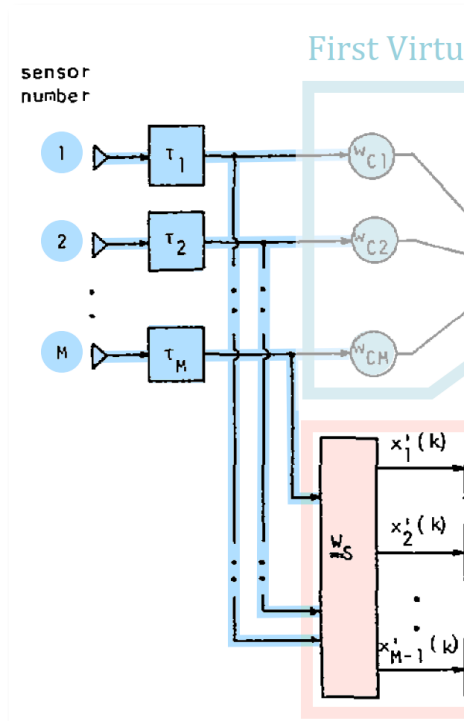
4. Claim 4

115. Claim 4 depends from claim 1 and further recites “wherein the signal processor is operative to add a delay to the first microphone signals.” Brandstein discloses this limitation.

116. Brandstein explains that the “BM [i.e., blocking matrix] in Fig. 5.1 has a simple delay-and-sum structure” (*id.*, 90), also called a “delay-and-subtract” structure since the signals are subtracted (summed with a negative sign) to remove the target signal (*id.*, 88-89). In the case illustrated in Figure 5.1, Brandstein explains, the look direction is perpendicular to the array surface, so “no delay element is necessary.” (*Id.*) In other words, when θ in Figure 5.1 is zero, the blocking matrix’s delay-and-subtract structure can simplify to “a set of subtractors” (*Id.*, 89.) As Brandstein explains: “This structure is actually the one shown in Fig. 5.1.” (*Id.*) But as Brandstein implies, one of ordinary skill in the art would have understood that, when the target signal is expected to arrive from a non-perpendicular direction (i.e., when θ in Figure 5.1 is nonzero, as Figure 5.1 itself indicates is possible), the delay-and-subtract structure is not simplified. Rather, before subtracting, delays are used to time-align the target-signal components of the physical-microphone signals (for brevity, below I sometimes just refer to this as time-aligning signals). (*See id.*, 159 (“Delay-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.”.) Independently of Brandstein, one of ordinary skill in the art would have been familiar with delay-and-sum or -subtract techniques and would have known the purpose of the “delay” part of the technique was to time-align the target-signal components of the microphone signals. The time alignment allows the subtraction operations in the blocking matrix to block the target signal.

117. While Brandstein’s reference to a “delay-and-subtract” structure for the blocking matrix implies this time alignment, the original Griffiths-Jim paper—which Brandstein cites and discusses (e.g., Ex. 1003, 88-90, 108)—shows the time alignment explicitly.



(Ex. 1005 (Griffiths-Jim), Fig. 4 (annotated, highlighted excerpt).) The elements labeled $\tau_1, \dots, \tau_M$ are delays used to time-align the sensor signals. (*See id.*, 29.) Griffiths-Jim thus expressly shows aligning the sensor signals before feeding them to the blocking matrix, just as Brandstein contemplates. This technique would have been well known to one of ordinary skill in the art and would have formed part of the background knowledge of one of ordinary skill in the art. One of ordinary skill in the art reading Brandstein would have thus immediately understood that Brandstein's reference to a "delay-and-subtract" structure for the blocking matrix means the signals should be aligned in time, which can be addressed through conventional delays of the kind Griffiths-Jim describes.

118. Brandstein also expressly teaches modifying the original GSC to include additional delays to the microphone signals to compensate for errors in the assumed time alignment. (*See, e.g.*, Ex. 1003 (Brandstein), 90-91 (motivating modification), 92-93 & Fig. 5.4 (introducing z^{-L_2} elements to delay microphone signals by L_2 samples).)

119. Thus, claim 4 would have been obvious.

5. Claim 5

120. Claim 5 depends from claim 4 and further recites "wherein the signal processor is operative to raise the delay 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.”

121. Claim 5 encompasses standard time-alignment for near-field sources. The '357 patent explains that “ γ is the time difference between arrival of speech at V_1 compared to V_2 ” (Ex. 1001, 15:20-21), where the “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 ” (*id.*, 11:19-22). This appears in equation form in the '357 patent (below right), though there are a few typos (e.g., f_x instead of f_s); a provisional application (below left) incorporated by reference has the same equation without typos:

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

Ex. 1011 ('377 provisional), 13 of 42

$$\gamma_- = \frac{d_2 - d_1}{c} \cdot f_x \quad (\text{samples})$$

Ex. 1001, 11:11 (with typos)

In other words, the '357 patent states the time difference between arrival of speech at V_1 compared to V_2 is just the time difference between arrival of speech at O_1 compared to O_2 . As explained in paragraphs 64-70 and 80 above and paragraph 126 below, this difference in arrival times is taken from the standard near-field equations and is equal in magnitude to τ_i in Brandstein. Standard near-field time-alignment can be accomplished by applying a complex exponential term of the form $e^{-j\omega\tau_i}$ to

the frequency response of a microphone signal, which delays the signal in proportion to τ_i . (¶¶64-70, above, and ¶126, below.) Claim 5 thus encompasses delaying a signal in accordance with standard near-field time-alignment.

122. Brandstein discloses the limitation claim 5 recites in addition to those of claim 4. Alternatively, at a minimum, the limitation would have been obvious from Brandstein’s teachings. As I discuss above in connection with claim 4 (see ¶116 above), Brandstein teaches that the blocking matrix 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, i.e., “to compensate for the propagation delays in the arrival of the source signal at each microphone.” (Ex. 1003 (Brandstein), 159.) When the original Griffiths-Jim paper published in 1982, delay-and-sum beamforming had already been well developed in the art, and those of ordinary skill in the art would have understood why Griffiths-Jim disclosed delay elements to time-align signals before feeding them to the blocking matrix. By 2007, persons of ordinary skill in the art would have understood for over 20 years the role time-alignment plays in the lower branch of the GSC.

123. 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.” (*Id.*, 290 (internal citation omitted).) Indeed, by 2007, persons of ordinary skill in the art would have been long aware of far-field versus near-field

distinctions and how to time-align signals from near-field sources. (See ¶¶ 62-70 (far-field and near-field discussion), above.) Near-field compensation, including in the context of the GSC, was well known to persons of ordinary skill in the art. Applying near-field time-alignment for the lower branch of the GSC would have been part of the background knowledge possessed by a person of ordinary skill in the art because McCowan disclosed exactly that in 2002. (See ¶¶ 66-70 (McCowan discussion), above.) McCowan teaches time-aligning signals in the GSC's blocking matrix in a near-field situation:

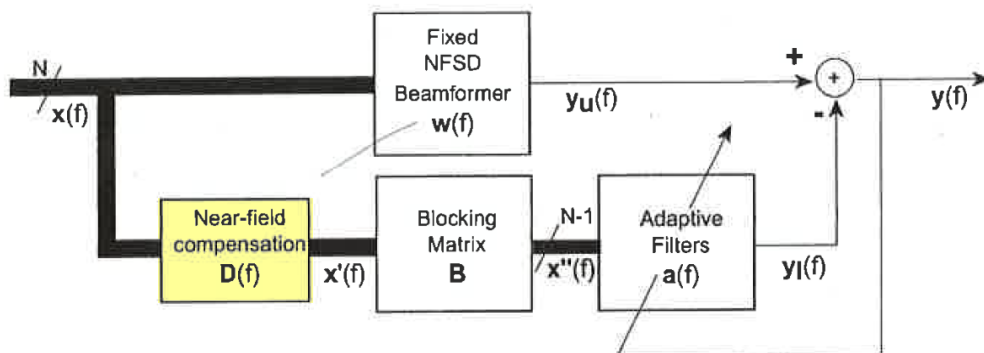


FIG. 3. Near-field adaptive beamformer.

(Ex. 1006 (McCowan), 91, Fig. 3 (highlighting added).) As McCowan discloses, to ensure that the standard GSC's blocking matrix can fulfill its purpose of completely blocking the desired signal, "the desired signal must . . . be time aligned . . ." (*Id.*, 93.) For near-field sources, McCowan thus teaches introducing a near-field compensation unit that time-aligns the microphone signals by applying to each microphone output a complex exponential of the form taught in Brandstein's

equation 2.28 (equation 2.28 discussed above in ¶¶65 and reproduced below in next paragraph). (*Id.*, 93 (see discussion of equation 16 and reference to equation 6); see also ¶¶66-70 (McCowan overview), above.) “Once this near-field compensation has been performed, a standard GSC blocking matrix can be employed to block the desired signal from the adaptive path.” (*Id.*, 93; *see also id.*, 95 (“The blocking matrix and adaptive noise canceling filters are taken from the standard GSC technique [citation]” (citing original Griffiths-Jim paper)).)

124. In view of this and the other background knowledge described above, claim 5 would have been obvious to a person of ordinary skill in the art based on Brandstein’s discussion of the GSC in Chapter 5 alone. A person of ordinary skill in the art would have known that some applications involve near-field target signals (see ¶¶64 (near field), above), that the GSC’s blocking matrix requires time-alignment (see ¶¶116 (claim 4), above), that near-field time-alignment involves applying delays encompassed by claim 5 (see ¶¶65-70, above), and that McCowan discloses such near-field time-alignment followed by a standard GSC blocking matrix (see ¶¶65-70, 123 above). The motivation and knowledge to use near-field time-alignment to implement the delays for the delay-and-subtract blocking matrix in the original Griffiths-Jim GSC (as described in Brandstein’s Chapter 5) would have been immediately apparent to a person of ordinary skill in the art. Alternatively, assuming for argument’s sake that such near-field time-alignment would require modifying

the GSC described in Brandstein's Chapter 5, the 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). The combination also 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.

125. Alternatively, claim 5 would have been obvious in view of the combination of Brandstein's discussion of the GSC and near-field time-alignment in Chapters 2 and 5. In Chapter 5, Brandstein describes the GSC blocking matrix as having 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. 1003, 90); and the "simplest structure for the BM is a delay-and-subtract beamformer which was described in the previous section" (Ex. 1003, 88). The "previous section" referred to is Chapter 2, which is the last chapter of Brandstein before Chapter 5 to discuss delay-and-sum/subtract structures and the GSC blocking matrix.

126. In Chapter 2, section 2.3, Brandstein discloses time-aligning signals for near-field beamforming takes into account the 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.

(*Id.*, 25.) As I discuss above (§65, above), τ_i is the time difference in arrival between microphone i and the reference microphone. Put another way, γ in the '357 patent is just τ_i in Brandstein's notation for the standard near-field design, just expressed in different units of time (samples versus seconds). Under Brandstein's near-field propagation model, the signal x_i on microphone i can be expressed in terms of the signal x_{ref} on the reference microphone as follows (where $j = \sqrt{-1}$):

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

To time-align the x_i signals, the inverse delay is applied: applying the complex-exponential term specified in Brandstein equation 2.28, $e^{-j\omega\tau_i}$, to x_i (i.e., $e^{-j\omega\tau_i}x_i$), delays x_i in proportion to τ_i to time-align the signals.

127. In section 2.4, Brandstein states: “all designs presented in section 2.3 can be implemented by using the GSC-structure.” (Ex. 1003, 35.) Specifically, a

“multi-channel *time-aligned* input signal ... is projected onto the noise-only subspace^[footnote 5] by a blocking matrix ***B***.” (*Id.*, 34 (italics added).) Footnote 5 to that sentence reads: “Which means that the desired signal is spatially filtered out (blocked).” Brandstein explains that a “well-known example” of ***B*** “is the original Griffith[s]-Jim matrix which subtracts two adjacent channels.” (*Id.*) The conclusion section to Chapter 2 summarizes the chapter and states: “we emphasized the close relationship between superdirective arrays and adaptive beamformers and their well-known implementation as a generalized sidelobe canceler,” as well as that “nearfield assumptions ... can be easily included.” (*Id.*, 36-37.) Chapter 2 therefore provides express motivation to use Brandstein’s disclosure of near-field time-alignment in section 2.3 to implement the delay for the GSC’s blocking matrix. 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. 1003, 290 (internal citation omitted).) A person of ordinary skill in the art 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 blocking matrix has a delay-and-sum (or delay-and-subtract) structure; (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 Chapters 2 and 5 to arrive at claim 5 amounts to a simple addition of one known element (near-field time alignment) to another known element (the GSC) to obtain predictable results (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. (*Id.*) Once more, a person of ordinary skill in the art would have had a reasonable expectation of success, particularly in light of McCowan. Claim 5 would therefore have been obvious in view of Brandstein's Chapters 2 and 5 and the background knowledge of a person of ordinary skill in the art. For all the foregoing reasons, claim 5 would have been obvious.

6. Claim 6

128. Claim 6 depends from claim 4 and further recites “wherein the signal processor is operative to raise the delay to a power that 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 the first physical microphone and a speech source of the speech and the fourth distance being between the second physical

microphone and the speech source.” Brandstein discloses this limitation. Alternatively, it would have been obvious in view of Brandstein’s teachings.

129. As I discuss above for claim 5, Brandstein renders obvious extending the standard GSC to a near-field design, including by time-aligning the microphone signals fed to the blocking matrix by applying delay term $e^{-j\omega\tau_i}$, whose exponent is proportional to τ_i . (¶¶ 122-126, above.) As discussed above:

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

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. 1003 (Brandstein), 25.) Brandstein’s τ_i is thus “a quantity equal to a third distance subtracted from a fourth distance, the third distance being between the first physical microphone and a speech source of the speech and the fourth distance being between the second physical microphone and the speech source,” divided by the speed of sound. Claim 6 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 . Brandstein’s delay term $e^{-j\omega\tau_i}$ therefore satisfies claim 6.

130. Alternatively, Brandstein renders obvious the limitation that claim 6 adds. While Brandstein discloses the delay term in the form above for time

expressed in seconds (or milliseconds, etc.; see p. 23, speed of sound c is expressed in meters per second), it would have been routine for one of ordinary skill in the art 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, one simply multiplies the delay (expressed in seconds) by the sampling frequency f_s (expressed in samples per second). It is a trivial task. For example, Brandstein's delay term $e^{-j\omega\tau_i}$ becomes $e^{-j\omega\tau_i f_s}$. (Under certain conventions and changes of variables, this may also be written as $z^{-\tau_i f_s}$.) Applying that exponential term delays a signal by $\tau_i f_s$ samples, as desired. (As the '357 patent acknowledges (Ex. 1001, 11:22-24), to the extent that τ_i is a non-integer, "fractional-delay filters (well known to those versed in the art) may be used," as well as other well-known techniques, including phase shifting.)

131. One of ordinary skill in the art would have been motivated to rewrite Brandstein's near-field delay in terms of samples. Brandstein itself provides a reason to rewrite this delay element in terms of samples for use with the GSC, as Figure 5.1 is shown in terms of samples. (Ex. 1003, 88-89 & footnote 1 (discussing effect of z^{-L_1} block in figure as an " L_1 -sample delay").) Brandstein also discloses implementing the GSC with a digital signal processor. (Ex. 1003, 104 & Fig. 5.16.) Additionally, digital signal processing was a known technique. A person of ordinary skill in the art would have reason to implement the GSC digitally because, among

other advantages over analog systems, digital circuits can be reproduced more easily in large quantities at low cost, digital systems are not affected 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.

132. Thus, claim 6 would have been obvious.

7. Claim 7

133. Claim 7 depends from claim 1 and further recites "wherein the first and second physical microphones comprise omnidirectional microphones." Brandstein discloses this limitation. (*Id.*, 101 (implementation using "omni-directional microphones").) See also the discussion in paragraph 89 above. Thus, claim 7 would have been obvious.

8. Claim 8

134. Claim 8 depends from claim 1 and further recites "wherein the first and second physical microphones are included in a microphone array." Brandstein

discloses this limitation. The microphones in Brandstein’s Figure 5.1 are included in a microphone array. This is evident both from the figure itself and from Brandstein’s discussion. (Ex. 1003, 88 (referring to “array surface”), 90 (“With the same microphone array”); *also id.*, Fig. 5.13.) Brandstein is also a reference dedicated to microphone arrays, and the GSC is a technique used with arrays. (*Id.*, 88-90 (referring to “array[s]” in discussing GSC).) Thus, claim 8 would have been obvious.

9. Claim 9

135. Claim 9 depends from claim 1 and further recites “wherein the first physical microphone and the second physical microphones are disposed along an axis and are separated from each other by a first distance.” Brandstein discloses this limitation.

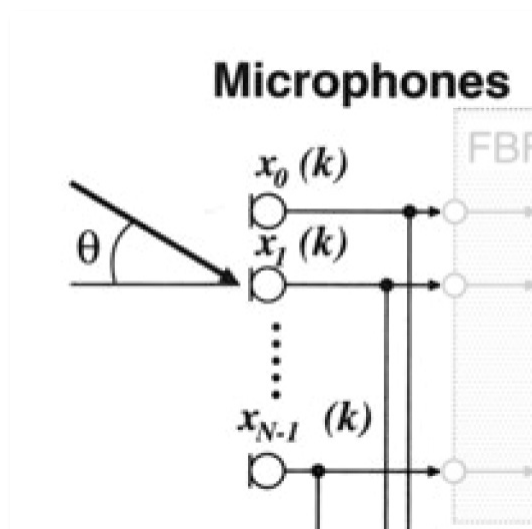
136. In Figures 5.1, 5.13, and 5.16 of Brandstein, the microphones are shown as lying on the same axis, separated by some distance. (*Id.*, 89 (Fig. 5.1), 103 (microphones in Fig. 5.13 shown in linear array with 4.1 cm spacing), 105 (indicating “Upper Row” of microphones used in Fig. 5.16).) In general, most array techniques require or assume the microphones are spaced apart, and a line can be drawn between any two spatially separated microphones. Thus, claim 9 would have been obvious.

10. Claim 10

137. Claim 10 depends from claim 9 and further recites “wherein a midpoint of the axis is a second distance from a speech source that generates the speech, wherein the speech source is located in a direction defined by an angle relative to the midpoint.”

138. Claim 9 assumes the microphones are disposed along an axis and are separated by a first distance. Any distance has a midpoint. Claim 1 discusses speech, which must originate from a source. And any speech source is some distance from the midpoint and at some angle relative to the midpoint. Claim 10 is therefore obvious for the same reasons claim 9 is.

139. In any case, the elements recited in claim 10 would have been obvious in view of Brandstein because Figure 5.1 discloses a linear microphone array along an axis and the target speech signal arriving at an angle θ :



(*Id.*, 89, Fig. 5.1 (excerpt; faded in part for clarity).) Under the far-field assumption, the speech signal’s wavefront is a plane that will arrive at the same angle at all points along the array axis. (*Id.*, 3.)

140. Thus, claim 10 would have been obvious.

11. Claim 11

141. Claim 11 depends from claim 10 and further recites “wherein the first virtual microphone is formed by subtracting the second microphone signal from the first microphone signal.” Brandstein discloses this limitation.

142. As I pointed out in footnote 2 in paragraph 89 above, in claim 1, it does not matter which virtual microphone, or branch, in the GSC is labeled as the “first” versus the “second.” For claim 1, I referred to the top branch as the “first” virtual microphone and the bottom branch as the “second” virtual microphone. But as I noted in footnote 2, my analysis would apply equally if the labels were switched.

[continued on next page]

143. For claim 11, switch the labels so that the bottom branch is the first virtual microphone and the top branch is the second virtual microphone. Then the **first virtual microphone** includes the blocking matrix with its delay-and-subtract structure that **subtracts** the **second microphone signal** from the **first microphone signal** as recited in claim 11:

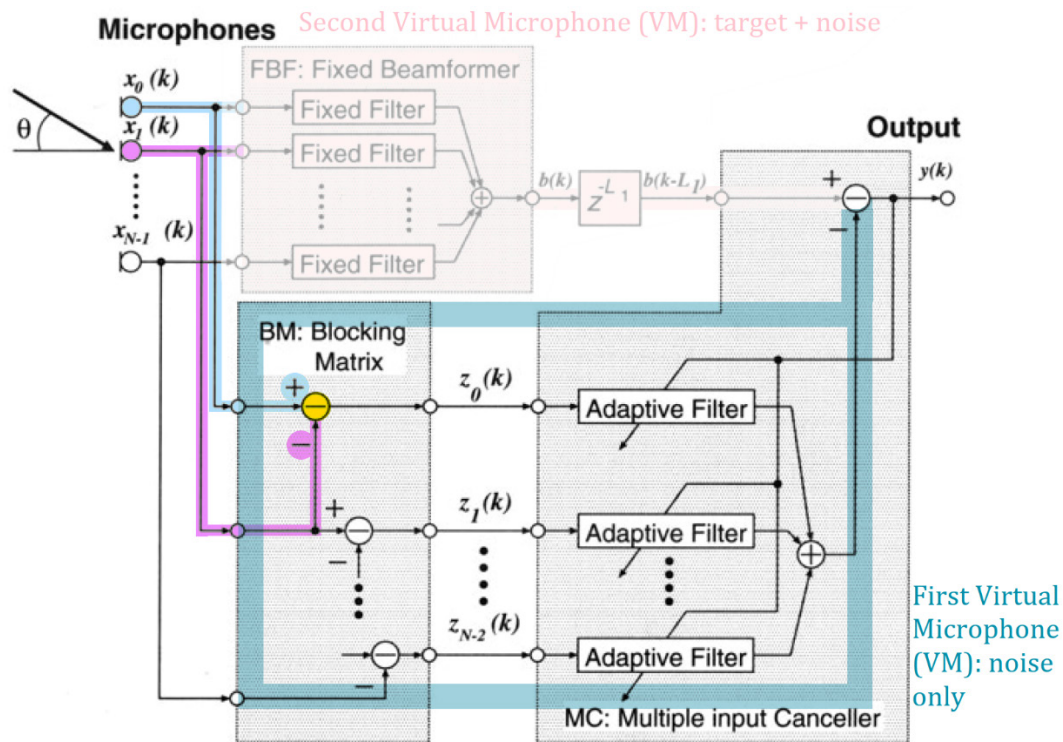


Fig. 5.1. Griffiths-Jim beamformer. It comprises a fixed beamformer (FBF), a multiple-input canceller (MC), and a blocking matrix (BM).

(*Id.*, 89, Fig. 5.1 (annotated).)

144. Thus, claim 11 would have been obvious.

12. Claim 12

145. Claim 12 depends from claim 10 and further recites “wherein the second virtual microphone is formed by subtracting the first microphone signal from the second microphone signal.” Brandstein discloses this limitation.

146. Consistent with how I discussed claim 1 (and claims 9 and 10), I will call the lower branch of the GSC the “second” virtual microphone. For purposes of claim 12, microphone signal x_1 is the “first” microphone signal and microphone signal x_0 is the “second.” With this designation, the **second virtual microphone** includes the blocking matrix with its delay-and-subtract structure that **subtracts** the **first microphone signal** from the **second microphone signal**:

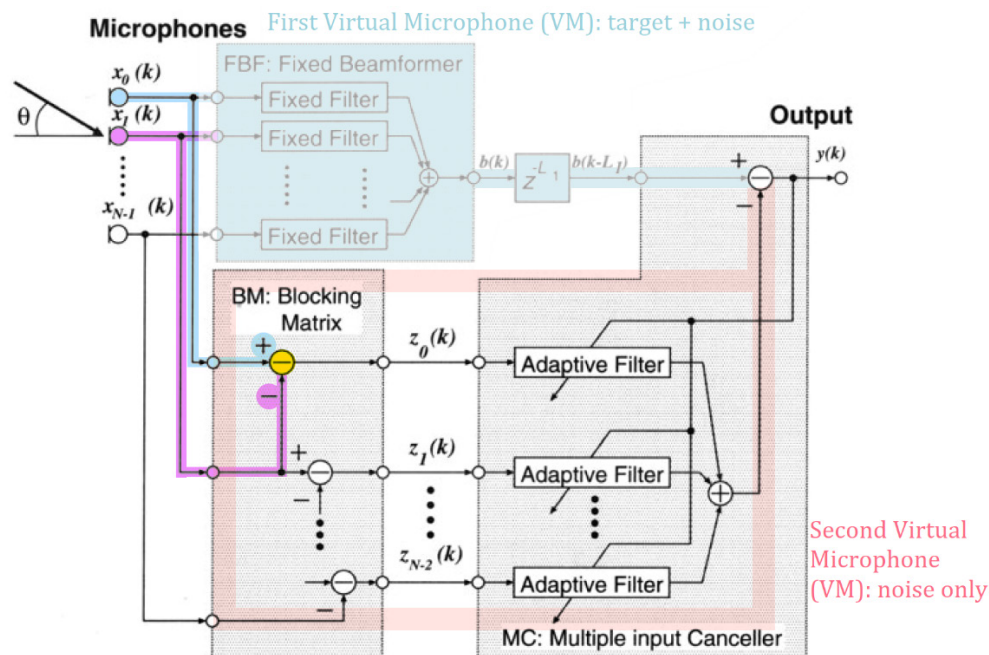


Fig. 5.1. Griffiths-Jim beamformer. It comprises a fixed beamformer (FBF), a multiple-input canceller (MC), and a blocking matrix (BM).

(*Id.*, 89, Fig. 5.1 (annotated).)

147. Thus, claim 12 would have been obvious.

13. Claim 13

148. Claim 13 depends from claim 1 and further recites “wherein the first virtual microphone is formed by subtracting the second microphone signal from a delayed version of the first microphone signal.” Brandstein discloses this limitation or at least renders it obvious.

149. As I pointed out in footnote 2 in paragraph 89 above, in claim 1, it does not matter which virtual microphone, or branch, in the GSC is labeled as the “first” versus the “second”—my analysis would apply equally if the labels were switched. For claim 13, the bottom branch of the GSC will be the “first virtual microphone.”

150. As I discuss for claim 4, Brandstein discloses or renders obvious that, if the target signal is expected to arrive from a direction not perpendicular to the array surface, the non-reference microphone signals will be delayed to be time-aligned before performing the subtraction in the blocking matrix that blocks the target signal—hence, Brandstein’s statement that the blocking matrix has a “delay-and-subtract” structure. (¶¶116-118(claim 4), above.) When the second microphone signal x_1 is the reference signal, a delay term is applied to the first microphone signal x_0 to time-align it, and typically no delay is applied to the second microphone signal x_1 . Thus, Brandstein discloses or renders obvious a **first virtual microphone** (that

includes the blocking matrix with its delay-and-subtract structure) that **subtracts** the **second microphone signal** from a **delayed version of the first microphone signal**.

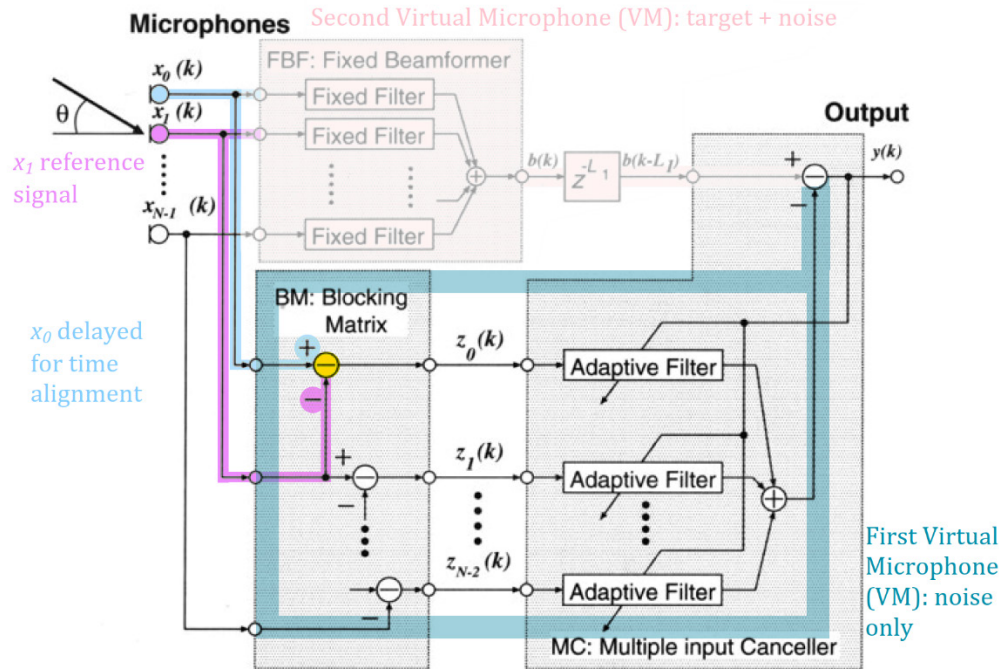


Fig. 5.1. Griffiths-Jim beamformer. It comprises a fixed beamformer (FBF), a multiple-input canceler (MC), and a blocking matrix (BM).

(Ex. 1003 (Brandstein), 89, Fig. 5.1 (annotated).)

151. Thus, claim 13 would have been obvious.

14. Claim 14

152. Claim 14 depends from claim 1 and further recites “wherein the second virtual microphone is formed by subtracting the first microphone signal from a delayed version of the second microphone signal.” Brandstein discloses or renders obvious this limitation.

153. The analysis for claim 14 is the same as for claim 13, except that the bottom branch of Brandstein’s GSC is the “second virtual microphone,” and

designation of the “first” and “second” microphone signals is reversed. Thus, the claim is obvious for the same reasons as claim 13.

154. Taking the bottom branch of the GSC as the “second virtual microphone” and the first microphone signal (x_1) as the reference signal, a delay is applied to the second microphone signal (x_0) to time-align it, and typically no delay is applied to the first microphone signal. Thus, Brandstein discloses or renders obvious that the **second virtual microphone** (which includes the blocking matrix with its delay-and-subtract structure) **subtracts** the **first microphone signal** from a **delayed version of the second microphone signal** as shown below:

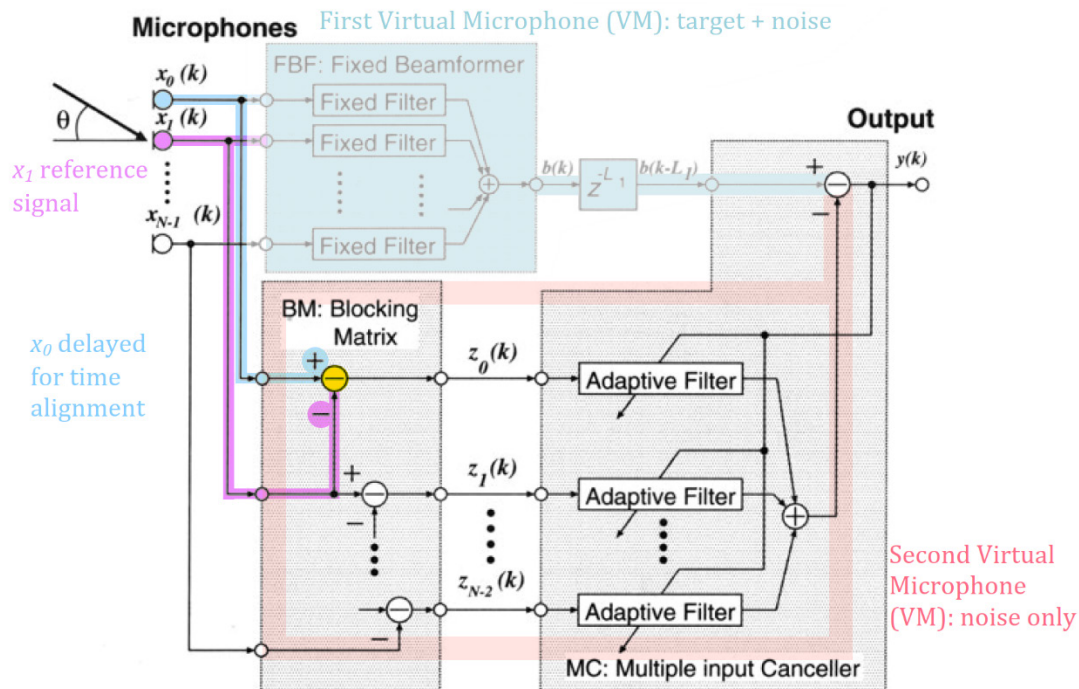


Fig. 5.1. Griffiths-Jim beamformer. It comprises a fixed beamformer (FBF), a multiple-input canceller (MC), and a blocking matrix (BM).

(Ex. 1003 (Brandstein), 89, Fig. 5.1 (annotated).)

155. Thus, claim 14 would have been obvious.

15. Claim 15

a. Preamble

156. The preamble recites “a device, comprising” Brandstein discloses this limitation. (¶¶85-86 (claim 1 preamble), above.)

b. 15[a]: First Virtual Microphone

157. Claim element 15[a] recites “a first virtual microphone comprising 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.” As discussed for claim 1, Brandstein discloses this limitation. (¶¶87-89 (claim element 1[a]), above.)

c. 15[b][2]: Second Virtual Microphone

158. Claim element 15[b] recites “a second virtual microphone comprising a second combination of the first microphone signal and the second microphone signal, wherein the second combination is different from the first combination.” As discussed for claim 1, Brandstein discloses this limitation. (¶¶90-91 (claim element 1[b][1]), above.)

d. 15[b][2]: Substantially Similar Responses to Noise and Substantially Dissimilar Responses to Speech

159. Claim element 15[b] further recites “wherein the first virtual microphone and the second virtual microphone are distinct virtual directional microphones with substantially similar responses to noise and substantially dissimilar responses to speech.” As discussed for claim 1, Brandstein discloses this limitation. (¶¶92-97 (claim element 1[b][2]), above.)

e. 15[c]: Virtual Microphone Array with a Single Null

160. Claim element 15[c] recites “a virtual microphone array including the first and second virtual microphones and having a single null oriented in a direction toward a source of speech.” Brandstein discloses this limitation.

161. Brandstein discloses that the first and second virtual microphones (top and bottom branches of the GSC) output separate signals that, as is common in array processing, are combined with each other. The output of the bottom branch is subtracted from the output of the top branch. As shown below, only the bottom branch (second virtual microphone) has a null oriented in the direction of the target speech signal:

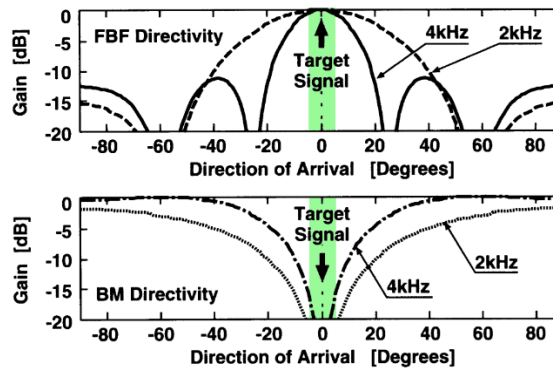


Fig. 5.3. Directivity pattern of a fixed beamformer (FBF) and a blocking matrix (BM).

(Ex. 1003 (Brandstein), 90, Fig. 5.3 (highlighting added).) Whereas the second virtual microphone has a null in the direction of the target speech, the first virtual microphone (top branch) does not. Brandstein thus discloses a virtual microphone array including the first and second virtual microphones and having a single null oriented in a direction toward a source of speech.

f. 15[d][1]: Signal Processor

162. Claim element 15[d] begins by reciting “a signal processor coupled with the first and second microphone signals and operative to combine the first and second microphone signals by filtering and summing in the time domain.” As discussed for claim 1, Brandstein discloses or renders obvious this limitation. (¶¶98-102 (claim element 1[c][1]), above.)

g. 15[d][2]: Applying a Varying Linear Transfer Function

163. Claim element 15[d] further recites the signal processor is operative to “apply a varying linear transfer function between the first and second microphone signals.” As discussed for claim 1, Gannot discloses this limitation; one of ordinary skill in the art would be motivated to combine Brandstein with Gannon with a reasonable expectation of success; and this limitation is obvious in view of Brandstein and Gannon. (¶¶103-106 (claim element 1[c][2]), above.)

h. 15[d][3]: Generating an Output Signal with Reduced Noise

164. Claim element 15[d] further recites the signal processor is operative to “generate an output signal having noise content that is attenuated with respect to speech content.” As discussed for claim 1, Brandstein discloses this limitation. (¶¶107-109 (claim element 1[c][3]), above.)

i. Further Comments on Claim 15 in View of Brandstein and Gannot

165. In summary, both Brandstein and Gannot disclose a microphone-array GSC for reducing noise and enhancing speech signals. Brandstein discloses the original Griffiths-Jim GSC in this microphone-array context. Gannot proposes a more-robust GSC designed to handle source-to-microphone acoustical paths with arbitrary transfer functions. Gannot expressly teaches modifying the original GSC to apply ratios of transfer functions, with the ratios representing transfer functions

of acoustical paths between microphones. Brandstein discloses all other limitations of claim 15. Because Gannot not only expressly proposes modifying the original GSC but explains exactly how to modify the original GSC and reports the successful results of a real-world implementation, claim 15 would have been obvious to one of ordinary skill in the art in view of the prior art.

16. Claim 16

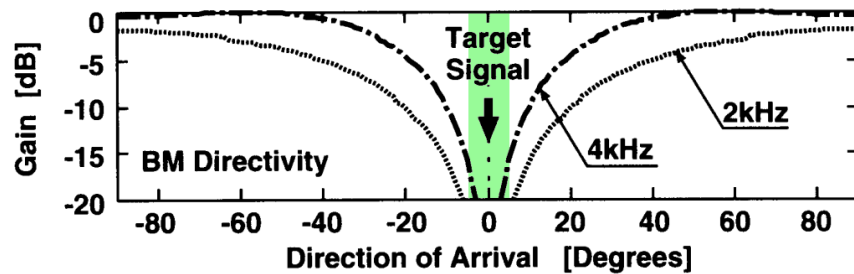
166. Claim 16 depends from claim 15 and further recites “wherein the source of speech comprises human speech.” As discussed for claim 3, which is similar, Brandstein discloses that the speech can be human speech. (¶¶112-114 (claim 3), above.) Thus, claim 16 would have been obvious.

17. Claim 17

167. Claim 17 depends from claim 15 and further recites “wherein the second virtual microphone includes a second linear response to speech and the single null comprises a region of the second linear response to speech having a measured response level that is lower than a measured response level of any other region of the second linear response to speech.” Brandstein discloses this limitation.

168. The '357 patent states: “The term ‘null’ means a zero or minima in the spatial response of a physical or virtual directional microphone.” (Ex. 1001, 5:57-58.) To the extent this statement is definitional, it does not change my analysis, as it covers the null I discuss below.

169. The second virtual microphone (the bottom branch of the GSC) has a null in the direction of the speech signal, as the second diagram in Figure 5.3 illustrates:



(Ex. 1003 (Brandstein), 90, Fig. 5.3 (second diagram; highlighting added).) The second virtual microphone's response to the speech is linear because it can be represented by a linear system. All elements in sound-processing systems, including the microphones and the signal processing, are typically designed to be linear systems to avoid distortion. The equations in Griffiths-Jim that describe the implementation of the GSC are linear equations, and from these equations it follows that each branch of the GSC satisfies the superposition property (or additivity and homogeneity) and is therefore a linear system. Figure 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.

170. Thus, claim 17 would have been obvious.

18. Claim 18

171. Claim 18 depends from claim 15 and further recites “wherein the first virtual microphone is formed by subtracting the second microphone signal from a delayed version of the first microphone signal.” As discussed for claim 13, Brandstein discloses this limitation or at least renders it obvious. (¶¶148-151 (claim 13), above.) Thus, claim 18 would have been obvious.

19. Claim 19

172. Claim 19 depends from claim 15 and further recites “wherein the second virtual microphone is formed by subtracting the first microphone signal from a delayed version of the second microphone signal.” Similar to claim 14 relative to claim 13, claim 19 has the same scope as claim 18; and as discussed for claim 14, Brandstein discloses the limitation added by claim 19 or at least renders it obvious. (¶¶152-155 (claim 14), above.) Thus, claim 19 would have been obvious.

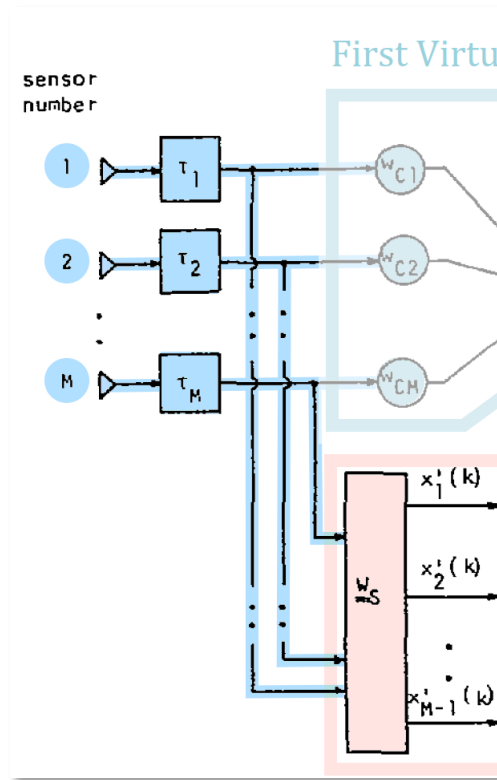
20. Claim 20

173. Claim 20 depends from claim 15 and further recites “wherein the first and second physical microphones are included in a microphone array.” As discussed for claim 8, Brandstein discloses this limitation. (¶134 (claim 8), above.) Thus, claim 20 would have been obvious.

**B. Claims 4, 13, 14, 18, and 19 Would Have Been
Obvious in View of Brandstein, Gannot, and
Griffiths-Jim**

174. Claims 4, 13, 14, 18, and 19 all recite limitations relating to delaying a microphone signal. Above, I discuss why these claims would have been obvious in view of Brandstein and Gannot alone, including in view of Brandstein's disclosure that the GSC's blocking matrix has a "delay-and-subtract" or "delay-and-sum" structure and in view of a person of ordinary skill in the art's understanding that Brandstein implies delays for time-alignment are necessary when the target signal is expected to arrive at an angle that is not perpendicular to the array surface. (E.g., ¶¶115-119 (claim 4), above.)

175. Alternatively, the delay limitations of claims 4, 13, 14, 18, and 19 would have been obvious to one of ordinary skill in the art in view of Brandstein, Gannot, and Griffiths-Jim, as Griffiths-Jim expressly discloses those limitations. While one of ordinary skill in the art would have understood that, when the target signal is expected to arrive from a non-perpendicular direction, delays are used to time-align the physical-microphone signals, and would have understood this from Brandstein alone (see ¶116 (claim 4), above), Griffiths-Jim shows the time alignment explicitly:



(Ex. 1005 (Griffiths-Jim), Fig. 4 (annotated, highlighted excerpt).) The elements labeled $\tau_1, \dots, \tau_M$ are delays used to time-align the sensor signals. (*See id.*, 27.) Griffiths-Jim thus expressly shows aligning the sensor signals before feeding them to the blocking matrix, just as Brandstein contemplates.

176. One of ordinary skill in the art would be motivated to combine Griffiths-Jim’s teachings with Brandstein because Brandstein discusses the same original GSC technique that Griffiths-Jim does—the only differences are the notation and the choices of which details to emphasize. Brandstein explains that the “simplest structure for the BM [i.e., blocking matrix] is a delay-and-subtract beamformer” and that, “[a]ssuming a look direction perpendicular to the array

surface, no delay element is necessary.” (Ex. 1003, 88.) As Brandstein itself suggests, a person of ordinary skill in the art would have understood that, for some applications, the target signal would arrive at an angle not perpendicular to the array surface and thus necessitate delay elements, as shown in Griffiths-Jim, which does not assume a perpendicular look direction. Brandstein’s discussion of Figure 5.1 is based on, and cites, Griffiths-Jim (reference [27] in Brandstein). (*E.g.*, Ex. 1003, 88, 108.) Combining Griffiths-Jim’s teaching of delay elements $\tau_1, \dots, \tau_M$ to time-align the sensor signals (*see* Ex. 1005, 29 & Fig. 4) with Brandstein’s teachings would have constituted applying a well-known technique to a device similar to, if not identical to, the device in Griffiths-Jim in the same way. The result of applying the delay elements to Brandstein would have been immediately foreseeable and predictable to persons of ordinary skill in the art, especially given Brandstein’s discussion of delay-and-sum and delay-and-subtract structures. One of ordinary skill in the art would have a reasonable expectation of success of combining the teachings of Griffiths-Jim and Brandstein because they describe the same solution. For the same reason, adding Griffiths-Jim to the combination does not interfere with the motivation to combine the original GSC (Brandstein) with Gannot or the expectation of success in doing so as discussed above. (¶¶104-106, 110, 163, 165.)

177. Thus, claims 4, 13, 14, 18, and 19 would have been obvious in view of Brandstein, Gannot, and Griffiths-Jim.

C. Claims 4-6, 13, 14, 18, and 19 Would Have Been Obvious in View of Brandstein, Gannot, and McCowan

178. Claims 4, 13, 14, 18, and 19 all recite limitations relating to delaying a microphone signal. Claims 5 and 6 depend from claim 4 and further recite limitations relating to raising the delay to a power that relates to the standard near-field design. (See ¶¶120-132 (claims 5-6), above.)

179. Above, I discuss why these claims would have been obvious in view of Brandstein and Gannot alone. (See discussion of claims 4-6, 13, 14, 18, and 19, above.)

180. Alternatively, the delay limitations of claims 5 and 6, and the more-general delay limitations of claims 4, 13, 14, 18, and 19, would have been obvious to one of ordinary skill in the art in view of Brandstein, Gannot, and McCowan, as McCowan expressly discloses those limitations. McCowan expressly teaches a near-field compensation unit that adapts the GSC for near-field conditions. See the discussion of McCowan generally and in connection with claim 5. (¶¶66-70 (McCowan overview), 126 (claim 5), above.) This near-field compensation is expressed as a matrix in equation 16 and includes a delay term in the form of an exponential, where the exponent 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. (¶¶66-70 (McCowan overview), 126 (claim 5), above; Ex.

1006, 89-93 (discussing equation 16, which relies on equation 6 (note the τ_i terms in the exponents), Eq. 3 and related discussion).) McCowan thus discloses the limitation that claim 5 adds. And for the same reasons I discuss above in claim 6 in connection with Brandstein, it would have been routine for one of ordinary skill in the art to express the near-field compensation in terms of samples. (See ¶¶128-132 (claim 6), above.)

181. One of ordinary skill in the art would be motivated to combine McCowan's teachings with Brandstein. McCowan starts with the same original GSC technique that Brandstein discloses and adapts it for near-field scenarios, which Brandstein expressly contemplates. McCowan explains that the original GSC does not perform well for near-field target signals because, without compensating for near-field effects, the GSC's lower path will not totally block the target signal, leading to target-signal cancellation in the output. (Ex. 1006 (McCowan), 93-94.) McCowan introduces a near-field compensation unit to the original GSC that solves this problem. (*Id.*, 90-96.) "Once 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. 1006 (McCowan), 93; *see also id.*, 95 ("The blocking matrix and adaptive noise canceling filters are taken from the standard GSC technique [citation]" (citing original Griffiths-Jim paper)).) McCowan goes beyond merely proposing modifying the original GSC and in fact explains exactly how to modify

the original GSC and reports the successful results of a real-world implementation. As a result, one of ordinary skill in the art would have been motivated to combine the teachings of McCowan and Brandstein and would have had a reasonable expectation of success of doing so. Combining the teachings would have been straightforward with predictable results because McCowan teaches modifying the original Griffiths-Jim GSC in a microphone-array context, which is what Brandstein describes. Furthermore, the near-field compensation of McCowan is compatible with Gannot and does not interfere with the motivation to combine the original GSC (Brandstein) with Gannot or the expectation of success in doing so as discussed above. (¶¶104-106, 110, 163, 165, above.)

182. Thus, claims 4, 5, 6, 13, 14, 18, and 19 would have been obvious in view of Brandstein, Gannot, and McCowan.

VII. CONCLUSION

183. For the foregoing reasons, it is my opinion that claims 1-20 of the '357 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.

184. 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

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

185. 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.

Amazon.com, Inc. v. Jawbone Innovations, LLC
Declaration of Dr. Richard M. Stern – U.S. Patent 11,122,357

Executed on November 21, 2022 at Pittsburgh, Pennsylvania.

A handwritten signature in black ink, appearing to read "Richard M. Stern", written over a horizontal line.

Richard M. Stern, Ph.D.