

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

GOOGLE LLC,

Petitioner,

v.

JAWBONE INNOVATIONS, LLC,

Patent Owner.

Case IPR2022-01060

Patent No. 8,503,691

DECLARATION OF JEFFREY S. VIPPERMAN, PH.D.

<i>Amazon v. Jawbone</i> U.S. Patent 8,503,691 Amazon Ex. 1003

TABLE OF CONTENTS

I.	INTRODUCTION	1
II.	SUMMARY OF OPINIONS.....	1
III.	BACKGROUND AND QUALIFICATIONS	2
	A. Education.....	2
	B. Experience	3
	C. Compensation.....	7
IV.	MATERIALS CONSIDERED	7
V.	LEGAL STANDARDS	9
	A. Claim Construction	9
	B. Level of Ordinary Skill	10
	C. Obviousness.....	10
VI.	'691 PATENT.....	13
	A. Specification.....	13
	B. Prosecution History	15
VII.	ANALYSIS OF PETITION GROUNDS.....	18
	A. Overview of the Asserted References	18
	1. Kanamori (Ex. 1005)	18
	2. McCowan (Ex. 1006).....	20
	B. Simulations of Virtual Microphone Responses	22
	C. Claim Construction	34
	D. Kanamori and McCowan Render Obvious Claims 1-22 and 29-40	34

1.	Independent Claim 1	34
a.	[1P]: “A microphone array comprising:”	34
b.	[1A]: “a first virtual microphone comprising a first combination of a first microphone signal and a second microphone signal,”	35
c.	[1B]: “the first virtual microphone having a first linear response to speech and a first linear response to noise,”	39
d.	[1C]: “the first linear response to speech being substantially similar across a plurality of frequencies for a speech source located at an angle relative to an axis of the microphone array,”	45
e.	[1D]: “wherein the first microphone signal is generated by a first physical microphone and the second microphone signal is generated by a second physical microphone; and”	47
f.	[1E]: “a second virtual microphone comprising a second combination of the first microphone signal and the second microphone signal,”	48
g.	[1F]: “the second virtual microphone having a second linear response to speech and a second linear response to noise,”	50
h.	[1G]: “the second linear response to noise being substantially similar to the first linear response to noise, and the second linear response to speech being substantially dissimilar to the first linear response to speech,”	54
i.	[1H]: “wherein at least one of the first linear response to noise and the second linear response to noise is being non-zero in a direction towards a source of noise.”	68
2.	Dependent Claims 2-22.....	69

U.S. Patent No. 8,503,691
Declaration of Jeffrey S. Vipperman, Ph.D.

a.	Claim 2: “The microphone array of claim 1, wherein the first and second physical microphones are omnidirectional.”	69
b.	Claim 3: “The microphone array of claim 1, wherein the first linear response to speech is devoid of a null, wherein the speech is human speech.”	69
c.	Claims 4-7.....	70
i.	Claim 4: “The microphone array of claim 3, wherein the second linear response to speech of the second virtual microphone includes a single null oriented in a direction toward a source of the speech.”	71
ii.	Claim 5: “The microphone array of claim 4, wherein the single null is a region of the second linear response to speech having a measured response level that is lower than the measured response level of any other region of the second linear response to speech.”	71
iii.	Claim 6: “The microphone array of claim 4, wherein the second linear response to speech includes a primary lobe oriented in a direction away from the source of the speech.”	72
iv.	Claim 7: “The microphone array of claim 6, wherein the primary lobe is a region of the second linear response to speech having a measured response level that is greater than the measured response level of any other region of the second linear response to speech.”	72
d.	Claim 8: “The microphone array of claim 4, wherein the first physical microphone and the	

- second physical microphone are positioned along
an axis and separated by a first distance.” 72
- e. Claim 9: “The microphone array of claim 8,
wherein a midpoint of the axis is a second distance
from the speech source that generates the speech,
wherein the speech source is located in a direction
defined by an angle relative to the midpoint.” 74
- f. Claim 10: “The microphone array of claim 9,
wherein the first virtual microphone comprises the
second microphone signal subtracted from the first
microphone signal.” 76
- g. Claim 11: “The microphone array of claim 10,
wherein the first microphone signal is delayed.” 77
- h. Claim 12: “The microphone array of claim 11,
wherein the delay is raised to a power that is
proportional to a time difference between arrival of
the speech at the first virtual microphone and
arrival of the speech at the second virtual
microphone.” 78
- i. Claim 13: “The microphone array of claim 11,
wherein the delay is raised 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 the speech
source and the fourth distance being between the
second physical microphone and the speech
source.” 81
- j. Claim 14: “The microphone array of claim 10,
wherein the second microphone signal is
multiplied by a ratio, wherein the ratio is a ratio of
a third distance to a fourth distance, the third
distance being between the first physical
microphone and the speech source and the fourth

	distance being between the second physical microphone and the speech source.”	82
k.	Claim 15: “The microphone array of claim 9, wherein the second virtual microphone comprises the first microphone signal subtracted from the second microphone signal.”	88
l.	Claim 16: “The microphone array of claim 15, wherein the first microphone signal is delayed.”	90
m.	Claim 17: “The microphone array of claim 16, wherein the delay is raised to a power that is proportional to a time difference between arrival of the speech at the first virtual microphone and arrival of the speech at the second virtual microphone.”	91
n.	Claim 18: “The microphone array of claim 16, wherein the power is proportional to a sampling frequency multiplied by a quantity equal to a third distance subtracted from a fourth distance, the third distance being between the first physical microphone and the speech source and the fourth distance being between the second physical microphone and the speech source.”	91
o.	Claim 19: “The microphone array of claim 18, wherein the first microphone signal is multiplied by a ratio, wherein the ratio is a ratio of the third distance to the fourth distance.”	91
p.	Claim 20: “The microphone array of claim 4, wherein the single null is located at a distance from at least one of the first physical microphone and the second physical microphone where the source of the speech is expected to be.”	93
q.	Claim 21: “The microphone array of claim 1, wherein the first virtual microphone comprises the	

	second microphone signal subtracted from a delayed version of the first microphone signal.”	95
r.	Claim 22: “The microphone array of claim 21, wherein the second virtual microphone comprises a delayed version of the first microphone signal subtracted from the second microphone signal.”	96
3.	Independent Claim 29	97
a.	[29P]: “A device comprising:”	97
b.	[29A]: “a first physical microphone generating a first microphone signal;”	97
c.	[29B]: “a second physical microphone generating a second microphone signal; and”	97
d.	[29C]: “a processing component coupled to the first microphone signal and the second microphone signal,”	98
e.	[29D]: “the processing component generating a virtual microphone array comprising a first virtual microphone and a second virtual microphone,”	100
f.	[29E]: “the first virtual microphone having a first linear response to speech and a first linear response to noise,”	101
g.	[29F]: “the first linear response to speech being substantially similar across a plurality of frequencies for a speech source at an angle relative to an axis of the microphone array,”	101
h.	[29G]: “the second virtual microphone having a second linear response to speech and a second linear response to noise,”	102
i.	[29H]: “the second linear response to noise being substantially similar to the first linear response to noise, and the second linear response to speech	

being substantially dissimilar to the first linear response to speech,”	102
j. [29I]: “wherein at least one of the first linear response to noise and the second linear response to noise is non-zero in a direction towards a source of noise”	102
k. [29J]: “wherein the first virtual microphone comprises the second microphone signal subtracted from a delayed version of the first microphone signal,.....	102
l. [29K]: “wherein the second virtual microphone comprises a delayed version of the first microphone signal subtracted from the second microphone signal.”	103
4. Dependent Claims 30-40.....	103
a. Claim 30: “The device of claim 29, wherein the first linear response to speech is devoid of a null, wherein the speech is human speech.”	103
b. Claim 31: “The device of claim 30, wherein the second linear response to speech of the second virtual microphone includes a single null oriented in a direction toward a source of the speech.”	103
c. Claim 32: “The device of claim 31, wherein the single null is a region of the second linear response to speech having a measured response level that is lower than the measured response level of any other region of the second linear response to speech.”	103
d. Claim 33: “The device of claim 31, wherein the second linear response to speech includes a primary lobe oriented in a direction away from the source of the speech.”	104

- e. Claim 34: “The device of claim 33, wherein the primary lobe is a region of the second linear response to speech having a measured response level that is greater than the measured response level of any other region of the second linear response to speech.” 104
- f. Claim 35: “The device of claim 31, wherein the first physical microphone and the second physical microphone are positioned along an axis and separated by a first distance.” 104
- g. Claim 36: “The device of claim 35, wherein a midpoint of the axis is a second distance from the speech source that generates the speech, wherein the speech source is located in a direction defined by an angle relative to the midpoint.” 104
- h. Claim 37: “The device of claim 36, wherein one or more of the first microphone signal and the second microphone signal is delayed.” 105
- i. Claim 38: “The device of claim 37, wherein the delay is raised to a power that is proportional to a time difference between arrival of the speech at the first virtual microphone and arrival of the speech at the second virtual microphone.” 105
- j. Claim 39: “The device of claim 38, wherein the power is proportional to a sampling frequency multiplied by a quantity equal to a third distance subtracted from a fourth distance, the third distance being between the first physical microphone and the speech source and the fourth distance being between the second physical microphone and the speech source.” 105
- k. Claim 40: “The device of claim 36, wherein one or more of the first microphone signal and the second microphone signal is multiplied by a gain factor.” 106

I, Jeffrey S. Vipperman, declare as follows:

I. INTRODUCTION

1. I have been retained as an independent expert by Google LLC (“Petitioner” or “Google”) in connection with an *inter partes* review of U.S. Patent No. 8,503,691 (the “’691 patent”) (Ex. 1001). I have prepared this declaration in connection with Google’s Petition (Paper 1).

2. Specifically, this document contains my opinions about the technology claimed in claims 1-22 and 29-40 of the ’691 patent (the “Challenged Claims”) and Google’s ground of unpatentability for these claims.

II. SUMMARY OF OPINIONS

3. This declaration considers the Challenged Claims of the ’691 patent. Below I set forth the opinions I have formed, the conclusions I have reached, and the bases for these opinions and conclusions.

4. In forming my opinions, I have assumed that the priority date of the ’691 patent is June 13, 2007, which is the filing date of U.S. Application No. 60/934,551, as listed on the cover page of the ’691 patent. Ex. 1001, Cover. I understand the ’691 patent purports to claim the benefit of the filing date of this application. Ex. 1001, Cover, 1:7-10.

5. Based on my experience, knowledge of the art, analysis of the asserted ground and references, and understanding that a person of ordinary skill in

the art (“POSITA”) would have had of the claims, it is my opinion that the Challenged Claims of the ’691 patent would have been obvious to a person of ordinary skill in the art as of 2007, based on the asserted ground.

III. BACKGROUND AND QUALIFICATIONS

6. I believe that I am well qualified to serve as a technical expert in this matter based upon my educational and work experience, which I summarize below. I understand that my curriculum vitae, which includes a more detailed summary of my background, experience, patents, and publications, is attached as Ex. 1004.

A. Education

7. I received my Ph.D. in Mechanical Engineering from Duke University in 1997. Previously, I obtained Master of Science and Bachelor of Science degrees in Mechanical Engineering from the Virginia Polytechnic Institute and State University (“Virginia Tech”) in 1992 and 1990, respectively. My dissertation at Duke was titled “Adaptive Piezoelectric Sensoriactuators for Multivariable Structural Acoustic Control.” My dissertation addressed the development of a hybrid analog/digital circuit and adaptation method to permit piezoelectric transducers to be used simultaneously as a sensor and an actuator. Doing so provides an array of truly “co-located” sensor/actuator pairs with minimum phase, such that stability of the multichannel feedback system is greatly enhanced. These were demonstrated for active structural acoustic control.

B. Experience

8. I am a Professor of Mechanical Engineering, Bioengineering, and Communication Sciences and Disorders. I also currently serve as Vice Chair of the Mechanical Engineering and Materials Science Department at the University of Pittsburgh.

9. I first began research in acoustics and sound systems in 1989 as an undergraduate student. My masters research concerned adaptive feedforward control of broadband structural vibration, and my Ph.D. research concerned the development of arrays of self-sensing piezoelectric transducers that could be used for active structural-acoustic control. I have also developed a number of algorithms for active control of noise and vibration.

10. My acoustics research has included a mix of theory, analytical and numerical modeling, and measurement of acoustic and vibration systems. Aside from the previously mentioned array research, my acoustics research has included transducer and controls development, transducer modeling/fabrication/testing, analog/digital signal processing, embedded systems, active and passive noise and vibration control, development of various types of metamaterials (e.g., phononic crystals, resonant lattices, layered media, and pentamode materials) for acoustical filtering and cloaking, development of noise classifiers to discern types of military noise or for incorporation into surgical devices as surgical aids, development of

thermoacoustic engines, refrigerators, and sensors (e.g., a wireless, “in-core” thermoacoustic sensor that can measure temperature and neutron flux inside a nuclear reactor). Additional topics of my research include developing structural acoustic models (i.e., concerned with sound radiation from vibrating structures) of sound transmission through finite cylinders, various methods of passive and active control of noise, vibration, and structural-acoustic radiation (i.e., controlling sound radiation of a vibrating structure by introducing additional vibrations to make it an inefficient radiator), hearing loss prevention, and modeling of ear response and damage to the inner ear for impulsive and ultrasound sources. During the early stages of the microelectromechanical systems (MEMS) revolution, I worked on producing some of the earliest silicon-on-insulator (MEMS) microphones through electronic fabrication methods.

11. As a professor, I have developed and taught three graduate courses directly related to acoustics and signal processing, including “Measurement and Analysis of Vibroacoustic Systems,” “Fundamentals of Acoustics and Vibration,” and “Measurement and Analysis of Random Data from Dynamical Systems.” The latter two courses cover acoustical arrays. I have also taught three mechanical measurements courses, a dynamic systems and introductory undergraduate and graduate mechanical vibrations course, and an advanced (Ph.D. candidate level) vibrations course, as well as related courses such as controls, undergraduate and

graduate dynamics, kinematics, mechanical measurements, and electrical circuits.

Further, I have developed and given a short course at the American Controls Conference on “Active Control of Sound, Vibration, and Structural Acoustics,” as well as two other short courses for local industry on “Acoustical Theory and Measurements” and “Noise and Vibration Measurements.”

12. I also have a consulting business (Blue Ridge Consulting) and am Vice President of Atlas Medtech, LLC, a University of Pittsburgh licensed startup company.

13. I have worked on Department of Defense (“DoD”) projects as a Principal Investigator and Co-Principal Investigator on projects that involve acoustic arrays. In one project, a microphone array and cross-correlation methods (time difference of arrival or TDOA methods) were used to determine the bearing angle for acoustic plane waves associated with various forms of military and natural noise. Multiple arrays were used to triangulate the location of the noise source. In conjunction, we developed machine learning algorithms to classify the noise source, which provided additional help for noise management programs around U.S. military bases. A corporate partner commercialized the array and research into a product. In another project, I helped co-develop a method for localizing sound using small arrays of unidirectional (e.g., “shot-gun”) microphones. The methods worked in both the time and frequency domains.

Another military project funded by DoD involved the development of 2-D and 3-D source parametric arrays for steering heterodyned ultrasound for communications systems.

14. Some of my professional activities include chairing an American National Standards Institute (ANSI) Committee to revise the ANSI S1.1 Acoustical Terminology Standard. I am also a Fellow in the American Society of Mechanical Engineers (ASME) and a former Chair of the Noise Control and Acoustics Division of ASME. I also chaired the Per Bruel Gold Medal in Acoustics Award selection committee for ASME. I have organized nine conference sessions on acoustics and was a Track Organizer (over multiple conference sessions) for nine ASME conferences, as well as Technical Program Chair over all acoustics-related conference sessions at the ASME International Mechanical Engineering Congress and Exposition (IMECE) in 2009. I also participated on a National Research Council (National Academies) panel to evaluate the hearing loss prevention component of the mining program for the National Institute for Occupational Safety and Health (NIOSH) research programs.

15. I have published numerous technical papers, book chapters, reports, and the like related to acoustic sensors and acoustic signal processing.

C. Compensation

16. I am being compensated for services provided in this matter at my usual and customary rate of \$400 per hour plus travel expenses. My compensation is not conditioned on the conclusions I reach as a result of my analysis or on the outcome of this matter, and in no way affects the substance of my statements in this declaration.

17. I am not aware of any financial interest that I have in the Patent Owner, or any of its subsidiaries or affiliates. Likewise, I am not aware of any financial interest that I have in Petitioner, or any of its subsidiaries or affiliates. I do not have any financial interest in the '691 patent or any proceeding involving the '691 patent.

IV. MATERIALS CONSIDERED

18. In forming my opinions, I have analyzed the following, including the '691 patent, its file history, the prior art listed in this declaration and in the Petition ground, and the materials listed in this declaration.

Exhibit	Description
Ex. 1001	U.S. Patent No. 8,503,691 ("the '691 patent")
Ex. 1002	Prosecution History of the '691 Patent
Ex. 1005	U.S. Patent Publication No. 2004/0185804 to Kanamori et al. ("Kanamori")

Exhibit	Description
Ex. 1006	Iain A. McCowan et al., <i>Near-field Adaptive Beamformer for Robust Speech Recognition</i> , 12 Digital Signal Processing 87-106 (2002) (“McCowan”)
Ex. 1007	U.S. Patent No. 5,473,701 to Cezanne et al. (“Cezanne”)
Ex. 1008	U.S. Patent Publication No. 2007/0244698 to Dugger et al. (“Dugger”)
Ex. 1009	Certified Translation of Japanese Unexamined Patent Application Publication No. H11-18186 (“Ikeda”)
Ex. 1010	U.S. Patent Application Publication No. 2008/0152167 to Taenzer (“Taenzer”)
Ex. 1011	U.S. Patent Application Publication No. 2006/0120537 to Burnett et al. (“Burnett”)
Ex. 1012	U.S. Patent Application Publication No. 2003/0031328 to Elko et al. (“Elko ’328”)
Ex. 1013	U.S. Patent No. 8,098,844 to Elko (“Elko ’844”)
Ex. 1015	Lawrence E. Kinsler et al., <i>Fundamentals of Acoustics</i> , John Wiley & Sons, Inc. (4th ed. 2000)
Ex. 1016	M. P. Norton et al., <i>Fundamentals of Noise and Vibration Analysis for Engineers</i> , Cambridge Univ. Press (2d ed. 2003)
Ex. 1019	U.S. Patent No. 8,942,387 to Elko et al. (“Elko ’387”)

19. My opinions are based on my experience, knowledge of the relevant art, the documents identified above, and the documents discussed in this declaration.

20. Ex. 1015 is a true and accurate copy of excerpts from Lawrence E. Kinsler et al., *Fundamentals of Acoustics*, John Wiley & Sons, Inc. (4th ed. 2000).

21. Ex. 1016 is a true and accurate copy of excerpts from M. P. Norton et al., *Fundamentals of Noise and Vibration Analysis for Engineers*, Cambridge Univ. Press (2d ed. 2003).

V. LEGAL STANDARDS

22. I am not a lawyer. My understanding of the legal standards to apply in reaching the conclusions in this declaration is based on discussions with counsel for Petitioner, my experience applying similar standards in other patent-related matters, and my reading of the documents submitted in this proceeding. In preparing this declaration, I sought to faithfully apply these legal standards to the challenged claims.

A. Claim Construction

23. I have been instructed that the terms appearing in the '691 patent should be interpreted in view of the claim language itself, the specification, the prosecution history of the patent, and any relevant extrinsic evidence. The words of a claim are generally given their ordinary and customary meaning, which is the meaning that the term would have to a person of ordinary skill in the art at the time of the invention, which I am assuming here is June 13, 2007. While claim limitations cannot be read in from the specification, the specification is the single

best guide to the meaning of a disputed term. I have followed these principles in reviewing the claims of the '691 patent and forming the opinions set forth in this declaration.

B. Level of Ordinary Skill

24. I understand a person of ordinary skill in the art is determined by looking at (i) the type of problems encountered in the art; (ii) prior art solutions to those problems; (iii) rapidity with which innovations are made; (iv) sophistication of the technology; and (v) educational level of active workers in the field.

25. In my opinion, a person of ordinary skill in the art ("POSITA") 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 can substitute formal education and additional formal education can substitute for work experience. I was at least a POSITA as of June 13, 2007.

C. Obviousness

26. I have been told that under 35 U.S.C. § 103, a patent claim may be obvious if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the

time the invention was made to a person having ordinary skill in the art to which said subject matter pertains.

27. I have been told that a proper obviousness analysis requires the following:

- a. Determining the scope and content of the prior art;
- b. Ascertaining the differences between the prior art and the claims at issue;
- c. Resolving the level of ordinary skill in the pertinent art; and
- d. Considering evidence of secondary indicia of non-obviousness (if available).

28. I have been told that the relevant time for considering whether a claim would have been obvious to a person of ordinary skill in the art is the time of invention. For purposes of my analysis, I assumed that the date of invention for the Challenged Claims is June 13, 2007.

29. I have been told that a reference may be modified or combined with other references or with the person of ordinary skill's own knowledge, if the person would have found the modification or combination obvious. I have also been told that a person of ordinary skill in the art is presumed to know all the relevant prior art, and the obviousness analysis may take into account the

inferences and creative steps that a person of ordinary skill in the art would employ.

30. I have been told that whether a prior art reference renders a patent claim obvious is determined from the perspective of a person of ordinary skill in the art. I have also been told that, while there is no requirement that the prior art contain an express suggestion to combine known elements to achieve the claimed invention, and while a suggestion to combine known elements to achieve the claimed invention may come from the prior art as a whole or individually and may consider the inferences and creative steps a person of ordinary skill in the art would employ, as filtered through the knowledge of one skilled in the art, obviousness grounds cannot be sustained by mere conclusory statements and must include some articulated reasoning with some rational underpinning to support the legal conclusion of obviousness.

31. I have been told that there is no rigid rule that a reference or combination of references must contain a “teaching, suggestion, or motivation” to combine references. But I also have been told that the “teaching, suggestion, or motivation” test can be used in establishing a rationale for combining elements of the prior art. I have also been told to be aware of distortions caused by hindsight bias, and that reading into the prior art the teachings of the invention at issue is improper.

32. I am aware that a claim may be obvious where the claim represents nothing more than a combination of prior art elements according to understood methods that yields predictable results. I am further aware that a claim may be obvious where it merely involves the simple substitution of one known element for another to achieve predictable results. I am additionally aware that it may be obvious to try a particular combination of claim features if selecting them requires merely choosing from a finite number of identified, predictable solutions, with a reasonable expectation of success.

VI. '691 PATENT

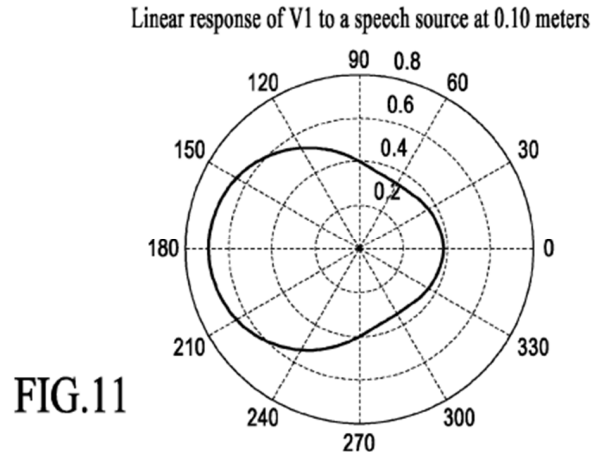
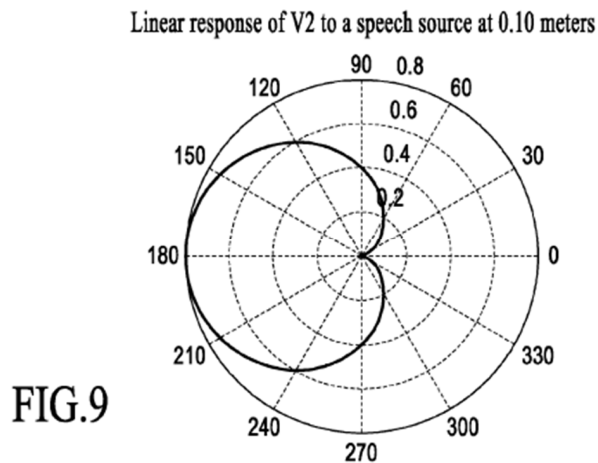
A. Specification

33. The '691 patent discloses generating “virtual microphones” from a pair of physical microphones that “are configured to have very similar noise responses and very dissimilar speech responses.” Ex. 1001, Abstract, 4:63-67. Virtual microphones include those “constructed using two or more omnidirectional microphones and associated signal processing.” Ex. 1001, 5:59-62. The '691 patent discloses that “varying the magnitude and sign of the delays and gains of the processing paths leads to a wide variety of virtual microphones” that “can be realized.” Ex. 1001, 9:5-8.

34. The '691 patent discloses two virtual microphones, where the first virtual microphone is a “virtual directional ‘speech’ microphone” and “has no

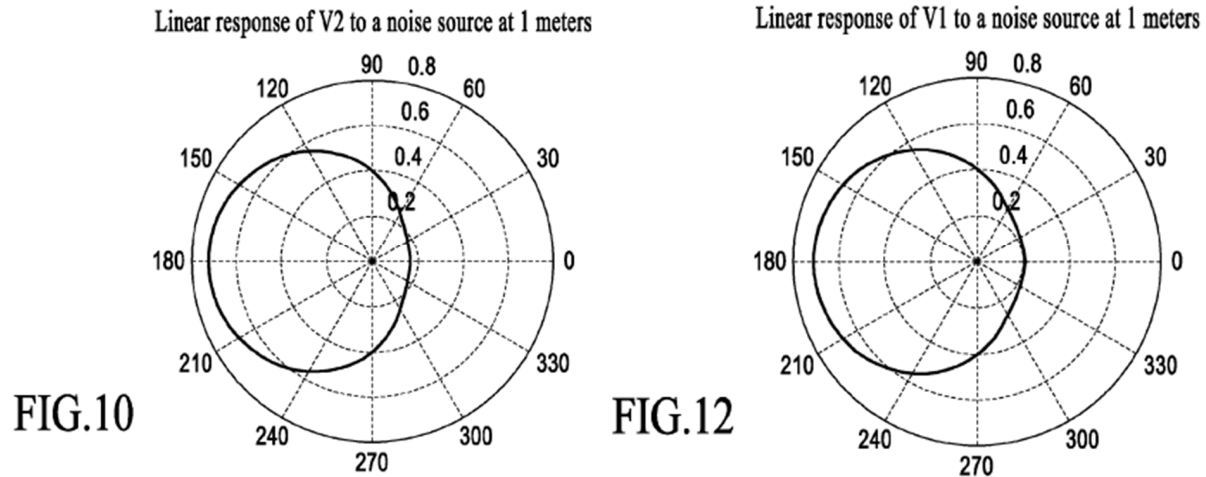
nulls,” and the second virtual microphone is a “virtual directional ‘noise’ microphone” and “has a null for the user’s speech.” Ex. 1001, 5:53-56. The virtual microphones are constructed so that the second virtual microphone “includes relatively small speech response,” the first virtual microphone “includes sufficient speech response,” and the first and second virtual microphones include “substantially similar noise response.” Ex. 1001, 10:26-42.

35. Figures 9 and 11 depict the linear responses of the second and first virtual microphones, respectively, to a speech source. Ex. 1001, 11:21-23, 12:20-22, Figs. 9, 11. In these figures, the speech source is assumed to be in the near field and located at 0° and 0.1 m from the microphone array. Ex. 1001, 12:61-64, 17:5-7. The '691 patent discloses that “the orientation of the speech response” of the first virtual microphone “is completely opposite the orientation of conventional systems, where the main lobe of response is normally oriented toward the speech source.” Ex. 1001, 12:39-43.



Ex. 1001, Figs. 9, 11.

36. Figures 10 and 12 depict the linear responses of the second and first virtual microphones, respectively, to a noise source. Ex. 1001, 11:25-28, 12:25-27, Figs. 10, 12. In these figures, the noise source is assumed to be in the far field and at all angles located 1.0 meter from the microphone array. Ex. 1001, 12:61-64, 17:5-7.



Ex. 1001, Figs. 10, 12.

B. Prosecution History

37. The '691 patent was filed as U.S. Patent Application No. 12/139,333 ("the '333 Application") and claims priority to four provisional applications, the earliest of which was filed June 13, 2007. During prosecution of the '333 Application, the Examiner initially rejected pending claims 1, 2, 21, 22, and 30 as anticipated by U.S. Patent No. 5,473,701 (Ex. 1007, "Cezanne") and the remaining

pending claims as obvious over Cezanne and U.S. Patent No. 7,386,135 (“Fan”).
Ex. 1002, 133-135.

38. In response, the Applicant amended pending independent claim 1 to add “the first virtual microphone having a first linear response to speech, the first linear response being substantially similar across a plurality of frequencies for a speech source located within a predetermined angle relative to an axis of the microphone array,” and made similar amendments to the other pending independent claims. Ex. 1002, 178, 181-184, 186. The Applicant argued that Cezanne and Fan did not disclose or render obvious this limitation. Ex. 1002, 166-167.

39. The Examiner again rejected the pending claims based on the same grounds. Ex. 1002, 191-193. The Examiner reasoned that Cezanne disclosed a first virtual microphone that disclosed the recited first linear response to speech because Cezanne disclosed equations that were similar to those disclosed in the specification of the ’333 Application. Ex. 1002, 193; *see* Ex. 1007 (Cezanne), 5:16-33; Ex. 1002, 16, 20.

40. In response, the Applicant amended pending claim 1, including adding “a first linear response to noise” and “the second virtual microphone having a second linear response to speech and a second linear response to noise, the second linear response to noise being substantially similar to the first linear

response to noise, and the second linear response to speech being substantially dissimilar to the first linear response to speech.” Ex. 1002, 204. The Applicant made similar amendments to the other pending independent claims. Ex. 1002, 208-212, 214-215. The Applicant argued that Cezanne and Fan did not disclose or render obvious these limitations. Ex. 1002, 217-220. The Applicant asserted that Cezanne taught away because “Cezanne at most discloses ‘substantial alignment of the nulls with sources of noise to reduce microphone array output signal level due to noise.’” Ex. 1002, 218 (quoting Ex. 1007 (Cezanne), 2:3-5). The Applicant argued that this is “divergent to the claimed invention,” pointing to the background section’s description in the ’333 Application discussing conventional multi-microphone systems that steer nulls to noise sources. Ex. 1002, 218. Regarding pending independent claim 29, the Applicant also argued that Cezanne did not disclose the claimed “ratio” because “Cezanne at most discloses a linear response to speech based on a ratio of front and back cardioid sensor signals.” Ex. 1002, 220 (citing Ex. 1007 (Cezanne), 7:3-28).

41. The Examiner then issued a notice of allowance. Ex. 1002, 235-236. As reasons for allowance, the Examiner stated that “a dual virtual microphone array as claimed including having a linear response to both speech and noise could not be anticipated nor, in combination, be rendered obvious over the prior art of record.” Ex. 1002, 236.

VII. ANALYSIS OF PETITION GROUNDS

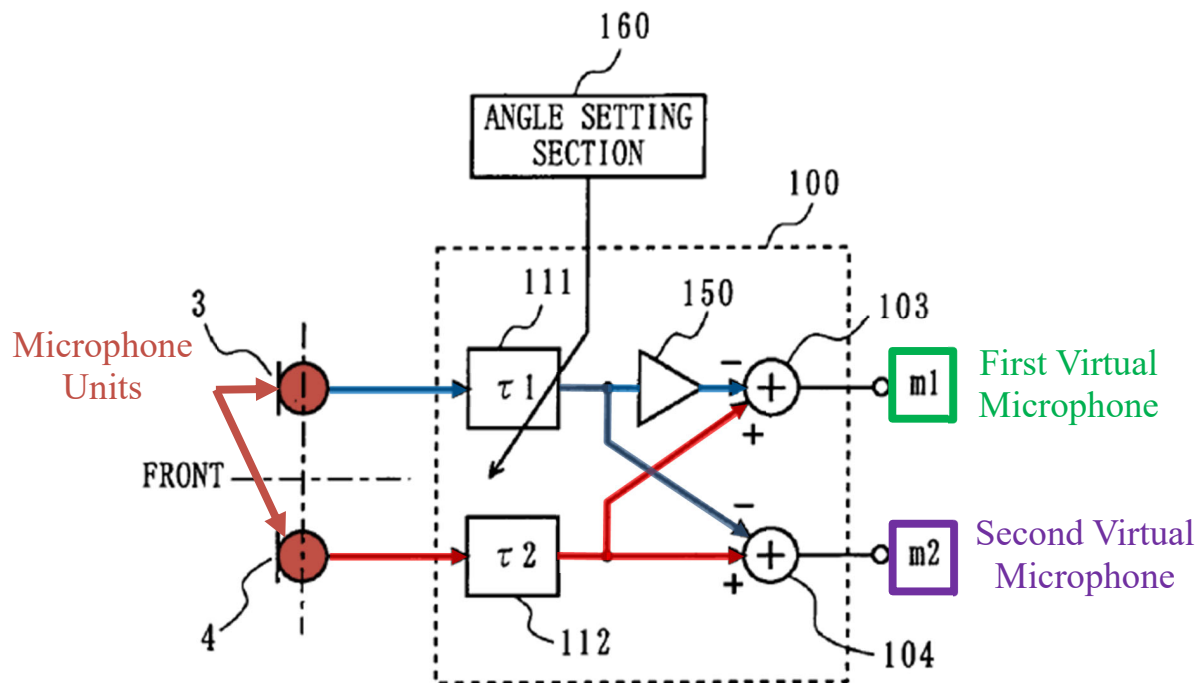
A. Overview of the Asserted References

1. Kanamori (Ex. 1005)

42. U.S. Patent Publication No. 2004/0185804 to Kanamori et al.

(“Kanamori”) discloses microphone devices that “use a plurality of microphone units having non-directivity . . . to obtain directivity by combining signals output from these microphone units.” Ex. 1005, ¶ [0151]. Such devices require the microphones “to be a certain distance (normally, 1 cm to 5 cm) apart,” provide “flexibility in designing directivity,” and allow the use of “a variable characteristic using a digital process.” Ex. 1005, ¶ [0151]. Kanamori discloses using such a directivity combining process “to generate a **main signal** and **noise reference signal**” and uses these signals to detect “a desired sound coming from a specific direction with noise being suppressed.” Ex. 1005, ¶¶ [0002], [0018], [0172]-[0173]. Figure 17A depicts one such directivity combining process that constructs **main signal m1** and **noise reference signal m2** using **microphone units 3 and 4** and associated signal processing. Ex. 1005, ¶¶ [0172]-[0173].

FIG. 17A



Ex. 1005, Fig. 17A (annotated).

43. The directivity patterns of **m1** and **m2** are constructed such that “the directions of a minimum sensitivity in the directivity are in the direction of the target sound” so that “signals **m1** and **m2** include components of the sound coming from a direction other than the direction of the target sound more than components of the target sound.” Ex. 1005, ¶ [0167]. “[I]t is preferable that the directivity patterns of the signals **m1** and **m2** be different in shape from each other only at the portions of directions of a minimum sensitivity in the directivity and be identical in shape to each other at other portions.” Ex. 1005, ¶ [0168]. Kanamori states that this is necessary “to suppress, based on the noise reference signal, noise components

included in the main signal under the circumstances where noise is coming from a plurality of directions.” Ex. 1005, ¶ [0168].

2. McCowan (Ex. 1006)

44. Iain A. McCowan et al., *Near-field Adaptive Beamformer for Robust Speech Recognition*, 12 Digital Signal Processing 87-106 (2002) (“McCowan”) teaches a technique for improving noise cancellation when a speech source is in the near field using **near-field compensation** in the signal path for generating a noise reference signal via a **blocking matrix**, as depicted in Figure 3, below. Ex. 1006, 87, 90-91.

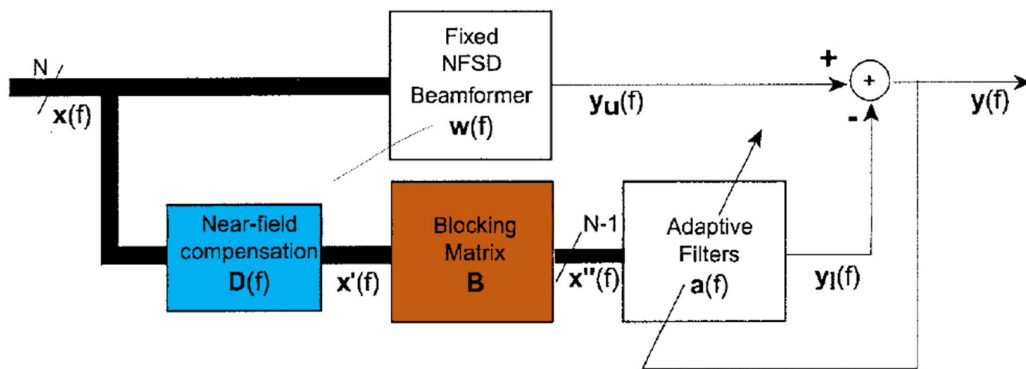


FIG. 3. Near-field adaptive beamformer.

Ex. 1006, Fig. 3 (annotated).

45. According to McCowan, “[m]any microphone array processing techniques assume a planar signal wavefront,” which “is reasonable for a far-field source,” as in Kanamori. Ex. 1006, 89. But when the speech source is in the near field, “a more accurate spherical wavefront model must be employed.” Ex. 1006,

89. McCowan discloses two issues that **near-field compensation** must account for: “phase misalignment” and “amplitude scaling.” Ex. 1006, 89-90, 93.

46. First, “the differences in distance to each sensor can be significant for a near-field source, resulting in phase misalignment across sensors.” Ex. 1006, 89. To properly align the phase, McCowan discloses the following equation representing “[t]he difference in propagation time to each microphone with respect to the reference microphone ($i=1$)”:

$$\tau_i = \frac{d_i - d_1}{c},$$

where d_i represents the distance from the source to the corresponding i^{th} microphone, d_1 is the distance from the source to the first (reference) microphone, and c is the speed of sound. Ex. 1006, 89-90.

47. Second, “the wavefront amplitude decays at a rate proportional to the distance traveled.” Ex. 1006, 90. “The resulting amplitude differences across sensors are negligible for far-field sources, but can be significant in the near-field case.” Ex. 1006, 90. McCowan discloses a microphone-attenuation factor given by the following equation:

$$\alpha_i = \frac{d_1}{d_i},$$

where d_i and d_1 represent the same distance values as above. Ex. 1006, 90.

48. Combining both phase misalignment and amplitude scaling, McCowan discloses that “if $x_1(f)$ is the desired source at the reference microphone, the signal on the i^{th} microphone is given by

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

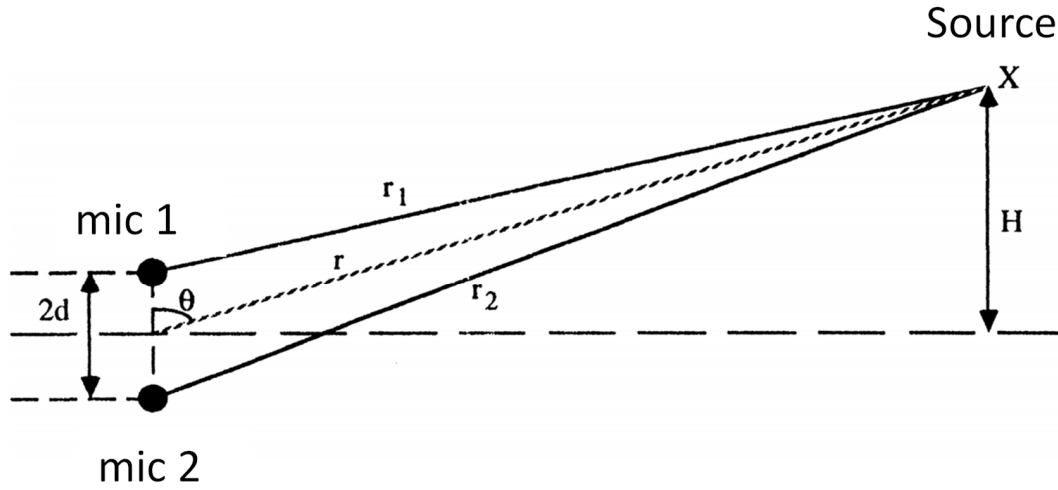
Ex. 1006, 90. McCowan’s **near-field compensation** requires delaying a reference signal x_1 by the difference in propagation time τ_i between x_1 and x_i and multiplying by the ratio α_i of the distance from the source to the reference microphone and the distance from the source to the i^{th} microphone. Ex. 1006, 89-90. “Once this **near-field compensation** has been performed,” a **blocking matrix** can completely “block the desired signal from the adaptive noise estimate” because the desired signal is both “time aligned and ha[s] equal amplitudes across all channels,” ensuring that “each row of the blocking matrix sums to zero.” Ex. 1006, 91, 93. “The **near-field compensation** effectively ensures that a true null exists in the beam-pattern of each blocking matrix row in the direction *and* distance corresponding to the desired source.” Ex. 1006, 94.

B. Simulations of Virtual Microphone Responses

49. The equation for sound radiation from a spherical source of small dimension is given by

$$p(r, t) = \frac{A}{r} e^{j(\omega t - kr)}, \tag{1}$$

where A is the amplitude of the pressure, r is the distance from the source, $e^{j\omega t}$ is the time dependence, and $\frac{e^{-jkr}}{r}$ is the spherical propagation and divergence of the soundwave. Ex. 1015, 171-172. For the time dependence, ω is equal to $2\pi f$, where f is the frequency of the soundwave, and t is time. For the spherical propagation of the soundwave, k is equal to $\frac{2\pi f}{c}$, where f is the frequency of the soundwave and c is the speed of sound. Equation (1) was well known to a POSITA and was conventionally used to simulate the response of a spherical source. It was also known that the same equation could be used to model an omnidirectional microphone as a result acoustic reciprocity, where $\frac{e^{-jkr}}{r}$ now represents the spherically divergent reception of acoustic waves. *See* Ex. 1015, 172-176 (explaining the concept of acoustic reciprocity). Note there is no angular or θ dependence in equation (1) because the microphone is omnidirectional. However, when two microphone signals from two omnidirectional microphones are processed and summed (or differenced) to create a virtual microphone, then the resulting virtual microphone will have a pickup pattern or linear response that varies with θ by virtue of the spacing between the physical microphones. From the figure below, a two-microphone array can be defined by the microphone spacing ($2d$), average radius, r , and radii between mic 1 and mic 2 and the source location X (at angle θ), given by r_1 and r_2 , respectively. Ex. 1016, 151-152.



Ex. 1016, 152 (modified). Distances r_1 and r_2 can be found from the law of cosines as:

$$r_1 = \sqrt{(r^2 + d^2 - 2rd \cos(\theta))}, \quad (2)$$

$$r_2 = \sqrt{(r^2 + d^2 + 2rd \cos(\theta))}. \quad (3)$$

Ex. 1016, 152.

50. For the two-microphone array in the figure above, the pressure from summing (differencing) the two identical omnidirectional microphones for a source of strength, A , will be:

$$p(r, \theta, t) = p_1(t) \pm p_2(t) = Ae^{j\omega t} \left(\frac{e^{-jkr_1}}{r_1} \pm \frac{e^{-jkr_2}}{r_2} \right), \quad (4)$$

which now has θ -dependence as a result of equations (2) and (3). The response of the virtual microphone given by equation (4) is simply the sum (difference) of the two physical microphone signals as a result of linear superposition. If the sign is

positive, we have a sum array and if negative, a difference array. Removing the time dependence to focus on geometric dependence and substituting equations (2) and (3) into equation (4) gives

$$p(r, \theta) = \left| A \left(\frac{e^{-jk\sqrt{(r^2+d^2-2rd\cos(\theta))}}}{\sqrt{(r^2+d^2-2rd\cos(\theta))}} \pm \frac{e^{-jk\sqrt{(r^2+d^2+2rd\cos(\theta))}}}{\sqrt{(r^2+d^2+2rd\cos(\theta))}} \right) \right|. \quad (5a)$$

Equation (5a) is the linear pressure response of the virtual microphone as a function of (r, θ) . As can be seen from equation (5a), the response is linear because the pressure p of the virtual microphone changes linearly with the amplitude A of the pressure of the sound source. Multiplying equation (5a) by r to normalize the effects of spherical divergence gives the shape of transverse response (just θ dependence at a given r) for the virtual microphone.

$$p(r, \theta) = r \left| \left(p_1 \frac{e^{-jk\sqrt{(r^2+d^2-2rd\cos(\theta))}}}{\sqrt{(r^2+d^2-2rd\cos(\theta))}} \pm p_2 \frac{e^{-jk\sqrt{(r^2+d^2+2rd\cos(\theta))}}}{\sqrt{(r^2+d^2+2rd\cos(\theta))}} \right) \right|, \quad (5b)$$

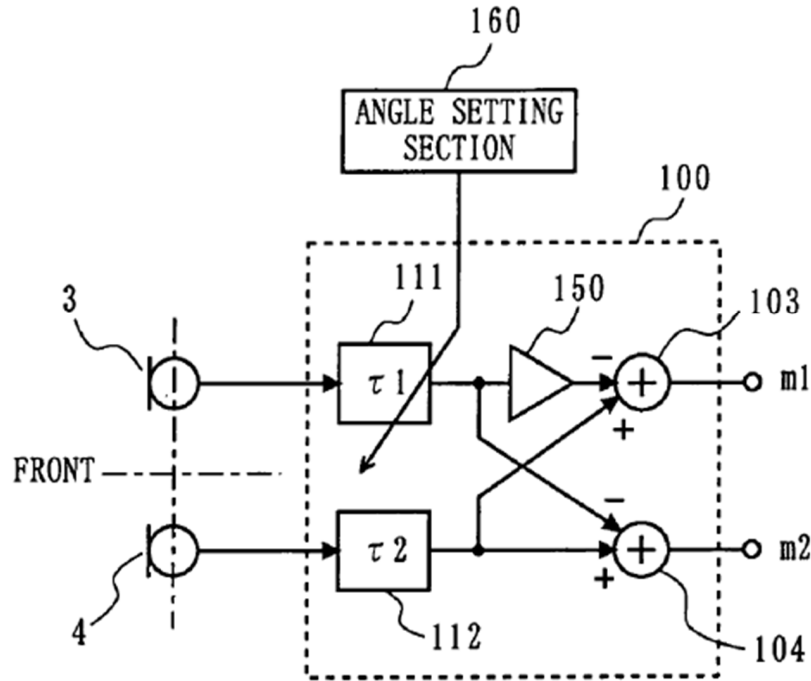
where p_1 and p_2 are the pressures measured at mics m_1 and m_2 , respectively. If we again assume that m_1 and m_2 are identical, then we can let $p_1 = p_2 = A$, then divide equation (5b) by A to get the linear response, which would be the response for a unit acoustic pressure source ($A = 1$).

$$p(\theta) = r \left| \left(\frac{e^{-jk\sqrt{(r^2+d^2-2rd\cos(\theta))}}}{\sqrt{(r^2+d^2-2rd\cos(\theta))}} \pm \frac{e^{-jk\sqrt{(r^2+d^2+2rd\cos(\theta))}}}{\sqrt{(r^2+d^2+2rd\cos(\theta))}} \right) \right|, \quad (5c)$$

51. As discussed above, Kanamori discloses the following system in

Figure 17A:

FIG. 17A



Ex. 1005, Fig. 17A. For the above system, the signal output from microphone unit 3 is amplified in the signal path for m_1 by a gain, g , represented in Figure 17A as signal amplifying section 150; the signal output from microphone unit 3 is delayed by time τ_1 , which can be modeled as $e^{-j\omega\tau_1}$; and the signal output from microphone unit 4 is delayed by time τ_2 , which can be modeled as $e^{-j\omega\tau_2}$.

Ex. 1005, ¶¶ [0166], [0174]; see Ex. 1006, 90. Assuming a unit amplitude source, the linear response of the two virtual mics, $m_1(\theta)$ and $m_2(\theta)$ can be computed considering equation (5c), respectively, as:

$$m_1(\theta) = r \left| \underbrace{-g e^{-j\omega\tau_1} \left(\frac{e^{-jkr_2}}{r_2} \right)}_{\text{unit 3 portion}} + \underbrace{e^{-j\omega\tau_2} \left(\frac{e^{-jkr_1}}{r_1} \right)}_{\text{unit 4 portion}} \right|, \quad (6)$$

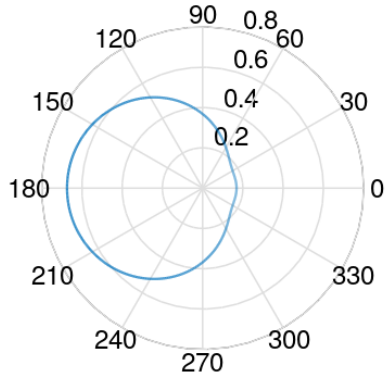
$$m_2(\theta) = r \left| \underbrace{-e^{-j\omega\tau_1} \left(\frac{e^{-jkr_2}}{r_2} \right)}_{\text{unit 3 portion}} + \underbrace{e^{-j\omega\tau_2} \left(\frac{e^{-jkr_1}}{r_1} \right)}_{\text{unit 4 portion}} \right|. \quad (7)$$

In defining the value of θ in equations (6) and (7), microphone unit 4 from Figure 17A of Kanamori corresponds to “mic 1” and microphone unit 3 from Figure 17A of Kanamori corresponds to “mic 2” in the above figure depicting r , r_1 , and r_2 .

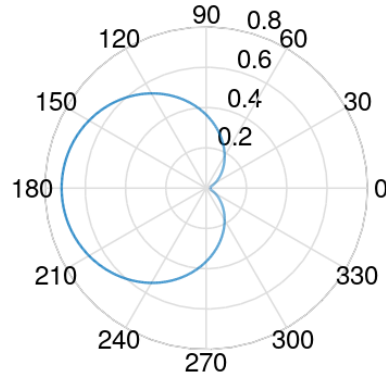
Ex. 1016, 152.

52. Kanamori discloses that the gain is 0.85; the delay τ_1 can be changed “in a range of $0 \leq \tau_1 \leq 2d/c$ (where d is a distance between the microphone units and c is the speed of sound)” to account for different target-sound directions, where $\tau_1=0$ corresponds to a target-sound direction of $\theta = 0^\circ$; and the delay τ_2 is d/c , where d is the distance between the two microphone units and c is the speed of sound. Ex. 1005, ¶¶ [0169], [0175]. Setting $g = 0.85$, $\tau_1 = 0$, and $\tau_2 = \frac{d}{c}$, i.e., the values that Figure 17A of Kanamori associates with a target sound direction of 0° , yields the following linear responses for virtual microphones m_1 and m_2 when the sound source originates in the far field.

m1, Kanamori Fig. 17a at 1m



m2, Kanamori Fig. 17a at 1m

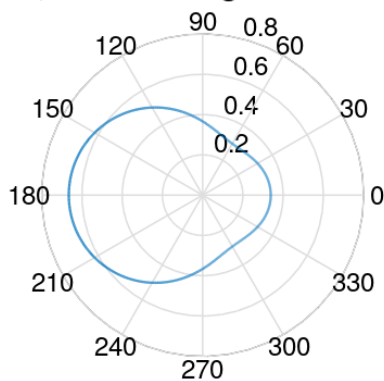


53. In the above linear responses, the two physical microphone units are assumed to be 2 cm apart, consistent with Kanamori's disclosure that "the microphone units are required to be a certain distance (normally, 1 cm to 5 cm) apart from each other." Ex. 1005, ¶ [0151]. Additionally, the frequency of the speech signal was assumed to be 1000 Hz, which is within the typical frequency range for speech. Ex. 1006, 87, 95. The sound source was plotted across all angles and assumed to be located 1 m from the microphone array, which was chosen because it is far enough from the microphone array to make the amplitude differences caused by amplitude decay due to the distance traveled negligible across the two microphones. Ex. 1006, 90. In other words, the distance from microphone unit 4 to the sound source divided by the distance from microphone unit 3 to the sound source is very close to 1 (0.98). Ex. 1006, 90. My simulations of Kanamori's m1 and m2 from Figure 17A are consistent with Kanamori's objective of having "a large difference in sensitivity between the signals m1 and

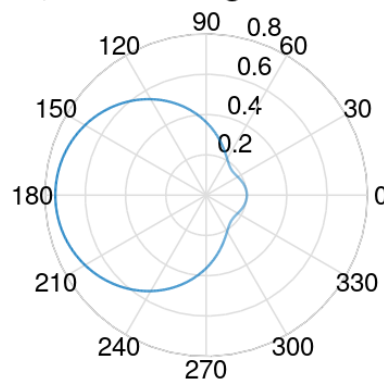
m2” in “the direction of the target sound” and substantially “no difference” in sensitivity between the signals m1 and m2” in all other directions. Ex. 1005, ¶ [0168].

54. If, on the other hand, the speech originates in the near field, the linear responses will look different because of the amplitude differences caused by amplitude decay due to the distance traveled. Ex. 1006, 90. For example, a speech source is typically between 3-5" (about 0.08-0.13 m) from the microphones in a headset, a common device well known before the '691 patent that uses microphones to receive near-field speech from a user of the headset. Ex. 1010 (Taenzer), ¶¶ [0009], [0078]; *infra* § VII.D.1.c ([1B]). Using equations (6) and (7) and changing the distance to the sound source from 1 m to 0.1 m yields the following linear responses for virtual microphones m1 (left) and m2 (right):

m1, Kanamori Fig. 17a at 0.1m



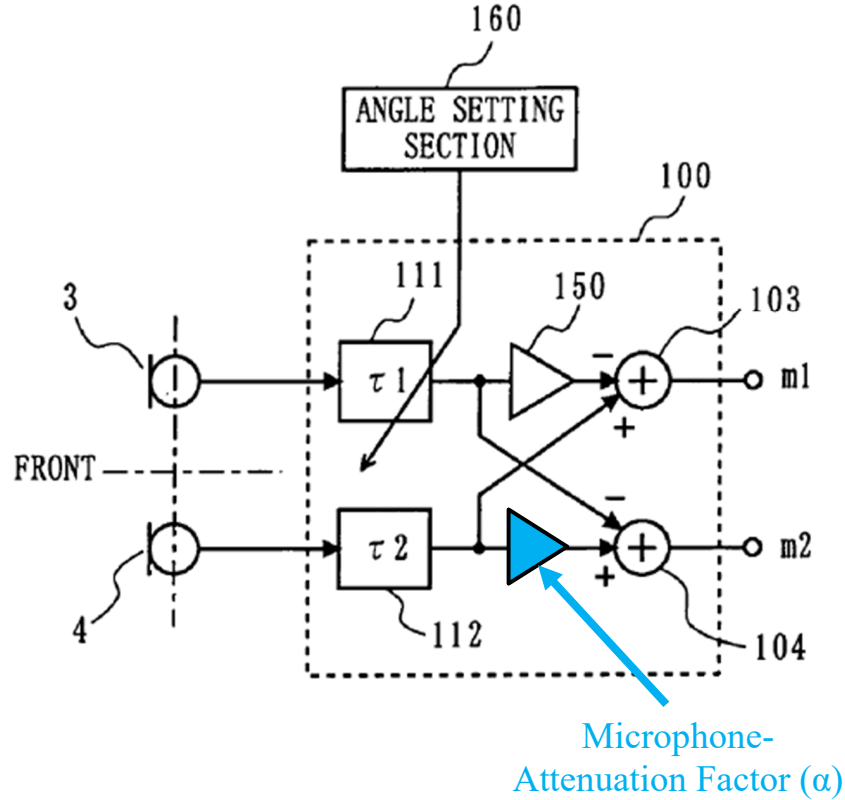
m2, Kanamori Fig. 17a at 0.1m



55. As discussed in further detail below, it would have been obvious to modify Figure 17A of Kanamori to implement McCowan’s [microphone-](#)

attenuation factor α to account for the effects of amplitude decay. The resulting system would look as follows:

FIG. 17A



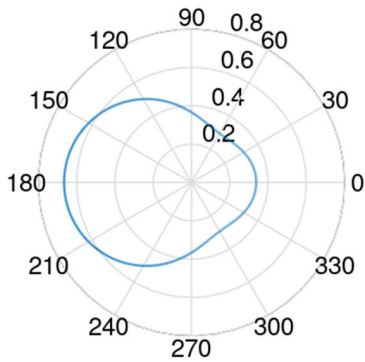
The system in the above figure is identical to Figure 17A, except the signal output from microphone unit 4 is now amplified by microphone-attenuation factor α for signal m2. Assuming a unit amplitude source, the linear response of the two modified virtual mics, m_1 and m_2 can be computed, respectively, using:

$$m_1 = r \left| -g e^{-j\omega\tau_1} \left(\frac{e^{-jkr_2}}{r_2} \right) + e^{-j\omega\tau_2} \left(\frac{e^{-jkr_1}}{r_1} \right) \right|, \quad (8)$$

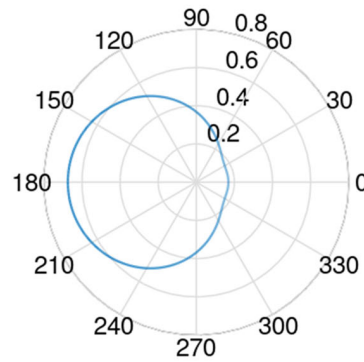
$$m_2 = r \left| -e^{-j\omega\tau_1} \left(\frac{e^{-jkr_2}}{r_2} \right) + \alpha e^{-j\omega\tau_2} \frac{e^{-jkr_1}}{r_1} \right|. \quad (9)$$

56. Equation (8) is identical to equation (6), so the linear responses of m1 to near-field ($r = 0.1\text{m}$) speech (left) and far-field ($r = 1\text{m}$) noise (right) are identical to those depicted above (again assuming a target sound direction of 0° for both).

m1, Modified Kanamori Fig. 17a at 0.1m

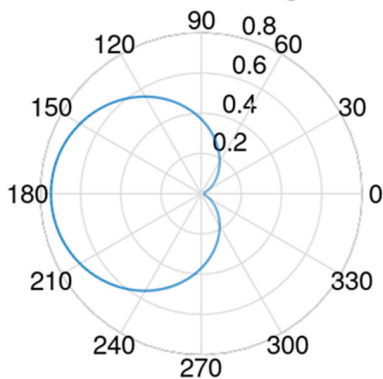


m1, Modified Kanamori Fig. 17a at 1m

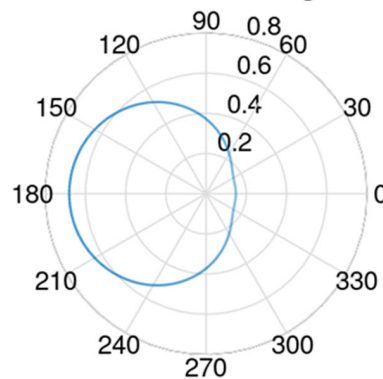


57. Equation (9), however, differs from equation (7) with the inclusion of McCowan's [microphone-attenuation factor](#). The linear responses for virtual microphone m2 to near-field speech (left) and far-field noise (right) are given below.

m2, Modified Kanamori Fig. 17a at 0.1m

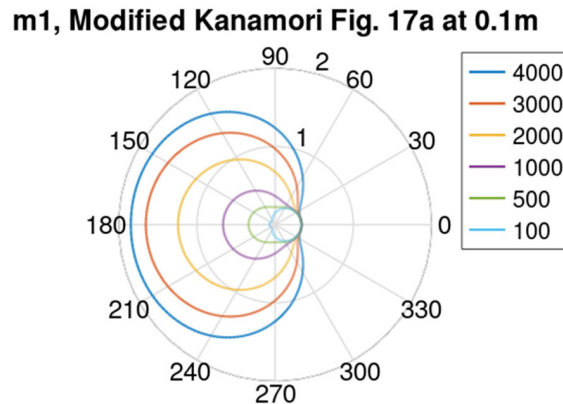


m2, Modified Kanamori Fig. 17a at 1m



58. Comparing the two near-field (speech) responses above, it can be seen that the linear responses are substantially dissimilar. The second virtual microphone m2 has a single null pointed in the direction of speech ($\theta = 0^\circ$) that has a response that is lower than any other region of the linear response to speech. The primary lobe of this linear response points away (180°) from the source of speech, with a response that is higher than any other region of the linear response. In contrast, the two far-field (noise) responses are substantially similar.

59. The near-field linear response of microphone m1 for various frequencies are given below. For a speech source located between $\theta \in [-30^\circ, 30^\circ]$, the linear response is substantially similar across a plurality of frequencies between 100 and 4,000 Hz.



60. These simulations are useful for seeing how a virtual microphone responds to a particular sound source. In a real-world implementation, such as the ones described in Kanamori and McCowan, the microphone signal produced by

each of the omnidirectional microphones is known. Equation (1) above is merely representative of how each omnidirectional microphone responds to a real-world sound source. Similarly, real-world implementations of virtual microphones are often implemented digitally. For example, Kanamori discloses using a “digital process” when performing the directivity combining process. Ex. 1005, ¶ [0151]. Such digital processes rely on discrete, as opposed to continuous, samples from each microphone. *See* Ex. 1005, ¶¶ [0107]-[0108]. Accordingly, while the above simulations assume continuous microphone signals, a POSITA would have understood that a real-world implementation in the digital domain would represent the delay in terms of the number of samples that correspond to the desired time delay, which can be represented as $z^{-\tau f_s}$. This is a function of the desired time delay τ and the sampling frequency f_s . For example, in the above example, the delay τ_2 is set to d/c , where $d=2$ cm and $c=343$ m/s. If the microphones were sampled at 16 kHz, for example, a person of ordinary skill in the art would understand that the signal output from microphone unit 4 would be delayed by about 0.93 samples ($z^{\frac{d}{c} f_s} = z^{\frac{.02}{343} 16000} = z^{0.93}$). As the '691 patent acknowledges, “fractional-delay filters” were “well known to those versed in the art” and would have been a conventional technique for implementing such a delay. Ex. 1001, 11:4-6; Ex. 1019 (Elko '387), 4:58-5:2, 5:60-66, 23:47-50 (disclosing that using “interpolation filtering to obtain the required delay” for a non-integer number of

samples was “relatively simple . . . if the sampling rate cannot be easily set to be equal to the propagation time of sound between the two sensors for on-axis propagation”). A POSITA would understand that, with a sufficiently high sampling frequency, any differences between the linear responses of a real-world digital implementation of the virtual microphones discussed above and the simulations of those linear responses depicted above assuming continuous signals would be so small as to be negligible.

C. Claim Construction

61. I understand that the claim terms of the '691 patent should be given their plain and ordinary meaning as understood by a POSITA, consistent with the patent's disclosure and prosecution history. For purposes of my analysis, I have considered each claim term and applied its plain and ordinary meaning consistent with the patent's disclosure and prosecution history.

D. Kanamori and McCowan Render Obvious Claims 1-22 and 29-40

1. Independent Claim 1

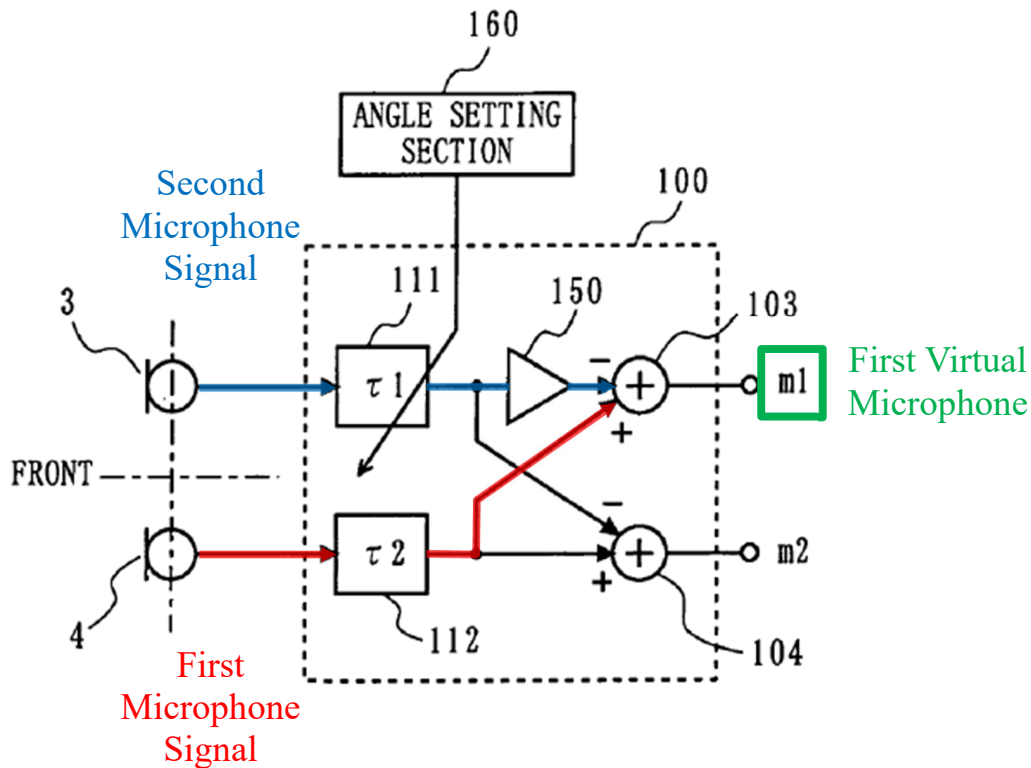
a. [1P]: “A microphone array comprising:”

62. In my opinion, Kanamori discloses this feature. Kanamori discloses a “microphone device” (e.g., an audio recorder or video recorder) that includes two microphone units spaced a certain distance apart. Ex. 1005, ¶¶ [0077], [0151], [0201], Figs. 14-21, 23. The two microphone units constitute a microphone array.

b. [1A]: “a first virtual microphone comprising a first combination of a first microphone signal and a second microphone signal,”

63. In my opinion, Kanamori discloses or renders obvious this feature, consistent with the '691 patent's disclosure that a “virtual microphone[]” is “a microphone constructed using two or more omnidirectional microphones and associated signal processing.” Ex. 1001, 5:59-62. Kanamori discloses that **main signal m1** is a first virtual microphone. Kanamori discloses that devices “often use a plurality of microphone units having non-directivity,” i.e., two or more omnidirectional microphones, “to obtain directivity by combining signals output from these microphone units,” i.e., associated signal processing. Ex. 1005, ¶ [0151]. In Figure 17A, **main signal m1** comprises a combination of “**a signal output from the fourth microphone unit 4**” and “**a signal output from the third microphone unit 3**,” corresponding to the first and second microphone signals, respectively. Ex. 1005, ¶¶ [0166], [0172], [0174].

FIG. 17A



Ex. 1005, Fig. 17A (annotated).

64. In my opinion, Kanamori discloses signal processing associated with constructing m1. As depicted above, a “signal output from the fourth microphone unit 4” is delayed by “the second signal delaying section 112.” Ex. 1005, ¶¶ [0166], [0174]. A “signal output from the third microphone unit 3” is delayed by the “signal delaying section 111” and multiplied by an amplification ratio at the “signal amplifying section 150.” Ex. 1005, ¶¶ [0166], [0169], [0174]. A “first signal subtracting section 103” subtracts the signal output from the signal amplifying section 150 from the signal output from the second signal delaying

section 112 to obtain m1. Ex. 1005, ¶¶ [0166], [0174]. A POSITA would have understood that these aspects of the directivity combining process constitute signal processing associated with the formation of virtual microphone m1.

65. In Figure 17A, “a plurality of microphone units having the same directivity characteristic are used to obtain the main signal.” Ex. 1005, ¶ [0172]. In my opinion, a POSITA would understand that Figure 17A discloses using two omnidirectional microphones as the microphone units. Omnidirectional microphones are examples of microphones having a uniform directivity characteristic across all angles. *See* Ex. 1005, ¶ [0193] (describing Figure 20 as comprising two “non-directional microphone units of the same characteristic”), ¶ [0025] (describing a directivity combining section similar to Figure 17A, where the microphones have an “identical” characteristic). Kanamori discloses that the microphone units can be either omnidirectional or directional microphones and accordingly discloses constructing main signal m1 using two omnidirectional microphones. Ex. 1005, ¶ [0020] (“[T]he main signal can be . . . a signal obtained by performing a predetermined process on a signal output from *any microphone unit* (that is, a non-directional microphone unit or a directional microphone unit whose main axis of directivity is oriented to a predetermined direction).” (emphasis added)), ¶ [0113] (stating that “a unidirectional microphone unit” or “a non-directional” microphone can be used), ¶ [0151] (stating that devices can use “a

plurality of microphone units having non-directivity or directivity of the same characteristic to obtain directivity by combining signals output from these microphone units”), ¶ [0193] (disclosing an embodiment with two “non-directional microphone units of the same characteristic”).

66. Moreover, in my opinion, it would have at least been obvious to use omnidirectional microphones as the microphone units in Figure 17A. Kanamori teaches forming directional microphones by combining signals from multiple omnidirectional microphones. Ex. 1005, ¶¶ [0020], [0113], [0151], [0193]. Kanamori discloses benefits to using omnidirectional microphones instead of directional microphones as the physical microphones, motivating a POSITA to use omnidirectional microphones to avoid “manufacturing variations in characteristics, such as the orientation of the direction of a minimum sensitivity in the directivity and the amount of sensitivity attenuation.” Ex. 1005, ¶ [0083]. Several other references also disclose combining signals from omnidirectional microphones to form virtual microphones, demonstrating that this process was well known and giving a POSITA a reasonable expectation of success in using omnidirectional microphones for the microphone units in Figure 17A. *E.g.*, Ex. 1005, ¶¶ [0151], [0193]; Ex. 1008 (Dugger), ¶ [0036]; Ex. 1009 (Ikeda), ¶ (0011).

- c. [1B]: “the first virtual microphone having a first linear response to speech and a first linear response to noise,”

67. In my opinion, Kanamori discloses or renders obvious this feature. As described for [1A], main signal $m1$ is obtained by subtracting a delayed and amplified version of the signal output from the third microphone unit from a delayed version of the signal output from the fourth microphone unit. Ex. 1005, ¶¶ [0166], [0174]; *supra* § VII.D.1.b. Kanamori discloses that this directivity combining process “is a linear process” and can be performed digitally. Ex. 1005, ¶¶ [0151], [0182]. A POSITA would have understood that $m1$ has linear responses to speech and noise that can be simulated and visualized, as I explain in Section VII.B and further below, by expressing this directivity combining process as an equation, with values for each component of the directivity combining process:

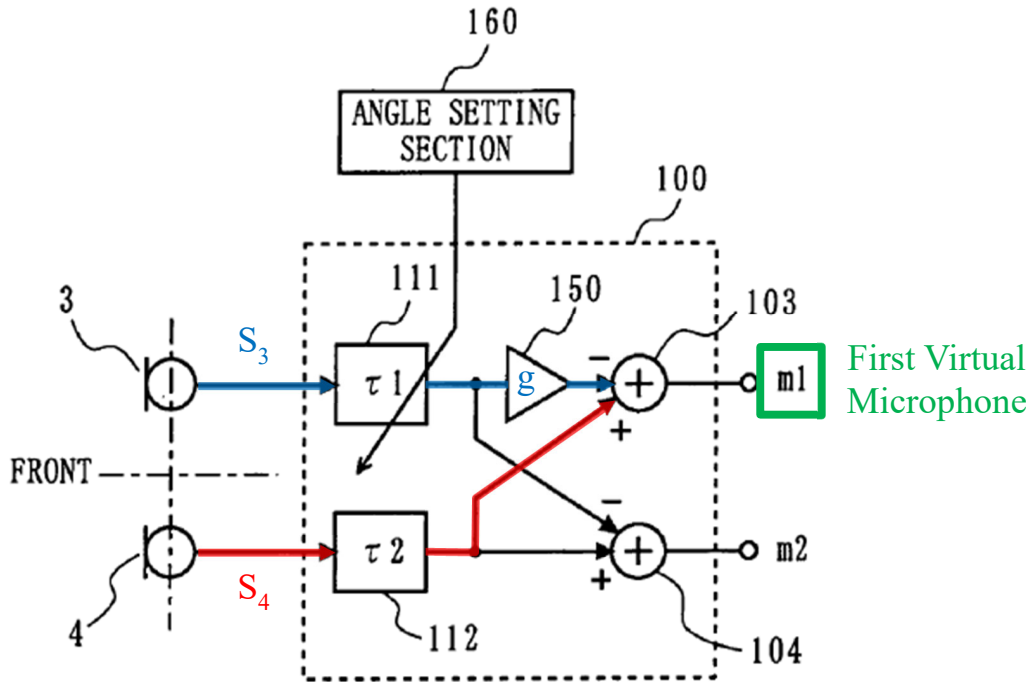
$$m1 = z^{-\tau_2 \cdot f_s} \cdot S_4 - g \cdot z^{-\tau_1 \cdot f_s} \cdot S_3, \quad (10)$$

where $z^{-\tau_2 \cdot f_s}$ refers to a delay of time τ_1 , $z^{-\tau_1 \cdot f_s}$ refers to a delay of time τ_2 , f_s is the sampling rate, g is the gain applied by signal amplifying section 150, S_3 is the “signal output from the third microphone unit 3,” and S_4 is the “signal output from the fourth microphone unit 4.” Ex. 1005, ¶¶ [0166], [0174]-[0175]. Consistent with Kanamori’s disclosure that the directivity combining process is a digital process, this equation is represented in the digital domain, which is evident by expressing the delay in terms of the number of samples (the desired time delay τ multiplied by

the sampling frequency f_s). *Supra* § VII.B; *infra* § VII.D.2.h (claim 12, explaining that z^{-s} refers to a delay of s samples). This equation is further illustrated by

Figure 17A:

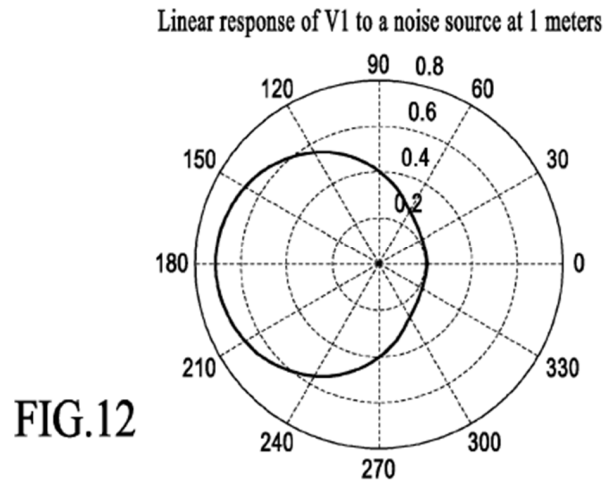
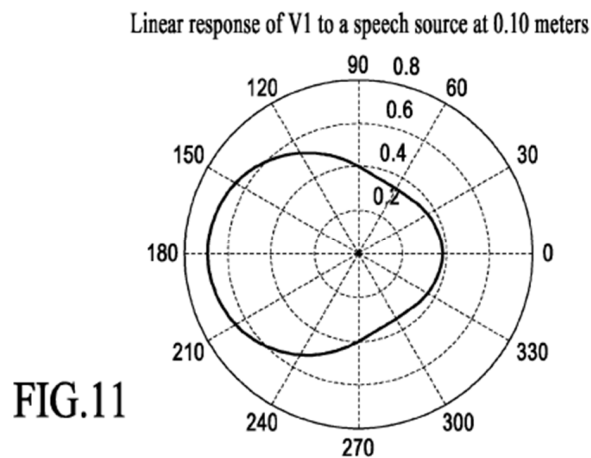
FIG. 17A



Ex. 1005, Fig. 17A (annotated). Kanamori discloses that τ_1 can range between “ $0 \leq \tau_1 \leq 2d/c$ (where d is a distance between the microphone units and c is the speed of sound),” and τ_2 is “set as $\tau_2 = d/c$.” Ex. 1005, ¶ [0175]. The variable delay τ_1 combined with delay τ_2 allows the target sound direction to “be changed in a range of ± 90 degrees with respect to the front direction.” Ex. 1005, ¶ [0175]. “[T]he delay amount τ_1 is determined so that the directions of a minimum sensitivity in the directivity are in the direction of the target sound.” Ex. 1005,

¶ [0167]. A POSITA would have understood from these disclosures that τ_1 is set to zero when speech is expected to approach from an angle of 0° because when τ_1 is set to zero, the directions of a minimum sensitivity face towards 0° (I explain what 0° refers to in Section VII.B and further below). Kanamori discloses that the gain “in the amplifying section is set so as to cause a difference in a sensitivity to the target sound between the main signal and the noise reference signal” and is “set at approximately 0.85.” Ex. 1005, ¶¶ [0025], [0169].

68. The '691 patent discloses plots of linear responses of virtual microphones, where a virtual microphone's linear response to speech is depicted as the directivity plot of the virtual microphone's near-field response, and a virtual microphone's linear response to noise is depicted as the directivity plot of the virtual microphone's far-field response. Ex. 1001, Figs. 9-13, 11:21-29, 12:20-33, 12:61-13:5, 17:5-7 (“[T]he noise (far-field) responses of V_1 and V_2 are very similar while the speech (near-field) responses are very different.”). For example, the '691 patent assumes that “the speech is located at approximately 0 degrees and approximately 10 cm,” and “the noise at all angles is located approximately 1.0 meter away from the midpoint of the array.” Ex. 1001, 12:61-64. Plots of the “linear response of virtual microphone V_1 ” to a “speech source at a distance of 0.1 m” and a “noise source at a distance of 1.0 m” are depicted in Figures 11 and 12, respectively, below. Ex. 1001, 12:20-22, 12:25-27.



Ex. 1001, Figs. 11-12. Kanamori does not explicitly describe **main signal m1** as having different linear responses, depending on the distance of the sound source to the array, but a POSITA, in my opinion, would have understood from Kanamori's description of **main signal m1**, including Equation (10) above, that **main signal m1** has a first linear response to near-field speech and a first linear response to far-field noise. *Supra* § VII.B.

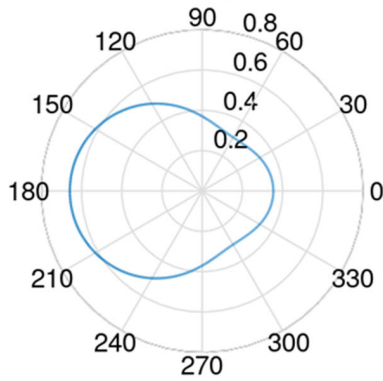
69. Kanamori discloses generic “microphone devices” that “detect[] a desired sound coming from a specific direction with noise being suppressed.” Ex. 1005, ¶ [0002]. In my opinion, it would have been obvious to implement Kanamori's microphone device in a device intended to receive near-field speech, such as a headset. Kanamori discloses various “use environments” and “purposes” for its microphone devices, including calling and voice recognition. Ex. 1005, ¶¶ [0115], [0182], [0207]. A headset was one such conventional microphone device used for calling. Ex. 1008 (Dugger), ¶ [0004]; Ex. 1009 (Ikeda), ¶ (0036);

Ex. 1010 (Taenzer), ¶¶ [0008]-[0009]; Ex. 1011 (Burnett), Abstract, ¶¶ [0048], [0071], [0126]. Such headsets commonly utilized omnidirectional microphones between 3-5" (.08-.13 m) from the user's mouth. Ex. 1010 (Taenzer), ¶¶ [0009], [0078]; Ex. 1011 (Burnett), ¶ [0048]. In these headsets, the microphones are close to the speech source and in the near field, but noise signals "generally arriv[e] from distant locations" and "simultaneously exist in various directions." Ex. 1005, ¶ [0012]; Ex. 1010 (Taenzer), ¶ [0009]; *see also* Ex. 1006, 88 (disclosing "near-field speech sources"), 89 (disclosing that "a planar signal wavefront" is a reasonable assumption "for a far-field source," but a "spherical wavefront model must be employed" "when the desired source is close to the array"); Ex. 1008 (Dugger), ¶ [0006] (disclosing that a noise source ("interferer") is "in the far field"), ¶ [0035] (disclosing that a speaker is "in the near field").

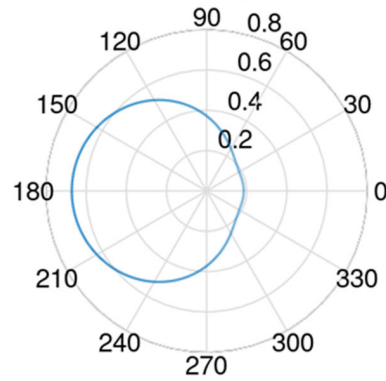
70. When implemented in a device that receives near-field speech, Kanamori's specific structure in Figure 17A discloses that **main signal m1** has a first linear response to near-field speech and a first linear response to far-field noise. Using Equation (10) from Kanamori and the equations described above in Section VII.B for simulating the response of virtual microphones, I simulated and plotted the linear responses of **main signal m1** to a 1000 Hz signal located 0.1 m (speech) from the microphone array and a 1000 Hz signal located 1 m (noise) from

the microphone array when $\tau_1=0$ (linear response to speech (left) and linear response to noise (right)):

m1, Kanamori Fig. 17a at 0.1m



m1, Kanamori Fig. 17a at 1m



See supra § VII.B. These plots are the same types of plots that the '691 patent calls the “linear response[s]” of the virtual microphone. *See* Ex. 1001, Figs. 9-13, 2:41-60, 11:21-29, 12:20-33. Additionally, the equations representing **m1** from Kanamori and the first virtual microphone in the '691 patent are substantially similar. *Compare* Equation (1), *with* Ex. 1001, 11:67. Moreover, a POSITA would understand that Kanamori’s **main signal m1** has linear responses for the reasons explained in Section VII.B and because **m1**’s response to a sound source at a particular distance, frequency, and direction is linearly proportional to its input. For example, as depicted in the above plots, **m1**’s response to a 1000 Hz speech source at 0.1 m and 0° can be determined from the linear response plot by multiplying the input amplitude by about 0.35, and **m1**’s response to a 1000 Hz noise source at 1 m and 90° can be determined from the linear response plot by

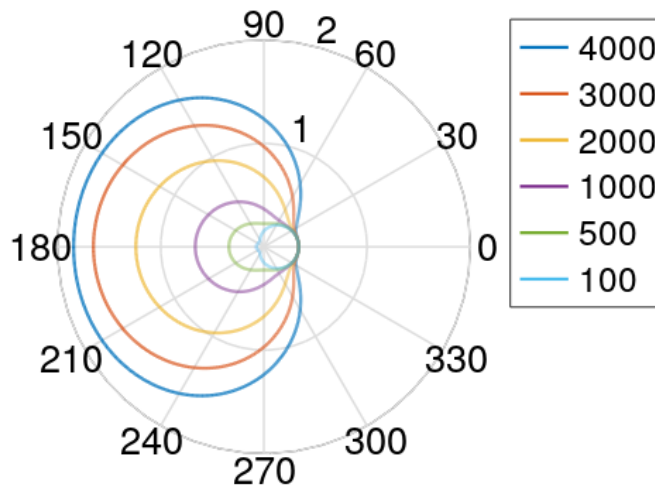
multiplying the input amplitude by about 0.38. This is because the above plots depict the response for a unit acoustic pressure source, as explained in Section VII.B.

71. Accordingly, Kanamori's **main signal m1** discloses a virtual microphone having a first linear response to speech and a first linear response to noise.

- d. **[1C]: “the first linear response to speech being substantially similar across a plurality of frequencies for a speech source located at an angle relative to an axis of the microphone array,”**

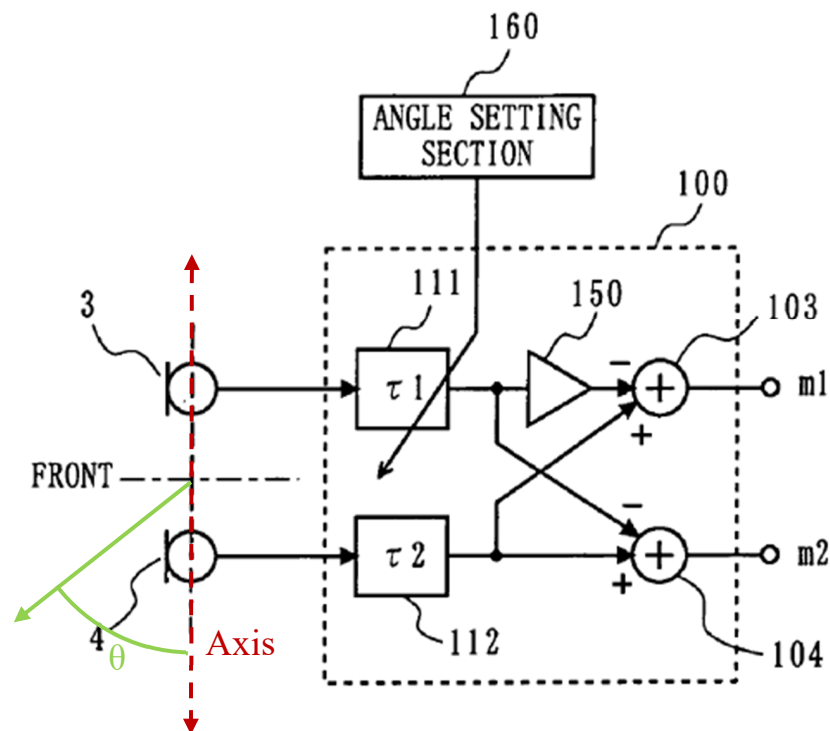
72. In my opinion, Kanamori discloses this feature. As explained for [1B], when implemented in a device that receives near-field speech, Kanamori's **main signal m1** in Figure 17A has a first linear response to near-field speech. *Supra* § VII.D.1.c. Using Equation (10) from Kanamori discussed above and the equations described above in Section VII.B for simulating the response of virtual microphones, I simulated and plotted the linear response of **main signal m1** to a signal containing various frequencies located 0.1 m (speech) from the microphone array when $\tau_1=0$ (legend values have units of Hz):

m1, Kanamori Fig. 17a at 0.1m



In the depictions of **main signal m1**'s linear responses, the angle corresponds to the value of θ in the figure below.

F I G. 1 7 A



Ex. 1005, Fig. 17A (annotated). Kanamori discloses that the direction of minimum sensitivity points towards the direction of speech, which in this instance corresponds to $\theta=0^\circ$ with reference to the above figure, as desired for $\tau_1=0$.

Ex. 1005, ¶¶ [0167], [0175]. As depicted, **m1**'s linear response to speech is substantially similar across a plurality of frequencies for between about ± 30 degrees relative to an **axis** of the microphone array, which includes the direction of speech. In other words, for this direction range, **m1**'s response to a speech source between 100 and 4000 Hz at 0.1 m can be determined from the linear response plot by multiplying the input amplitude by about 0.35. This is because the above plot depicts the response for a unit acoustic pressure source, as explained in Section VII.B.

73. Accordingly, Kanamori's **main signal m1** discloses a virtual microphone having a first linear response to speech that is substantially similar across a plurality of frequencies for a speech source located at an **angle** of 0° relative to an **axis** of the microphone array.

- e. **[1D]: “wherein the first microphone signal is generated by a first physical microphone and the second microphone signal is generated by a second physical microphone; and”**

74. In my opinion, Kanamori discloses this feature. Kanamori discloses a “signal output from the fourth microphone unit 4” (first microphone signal) and a

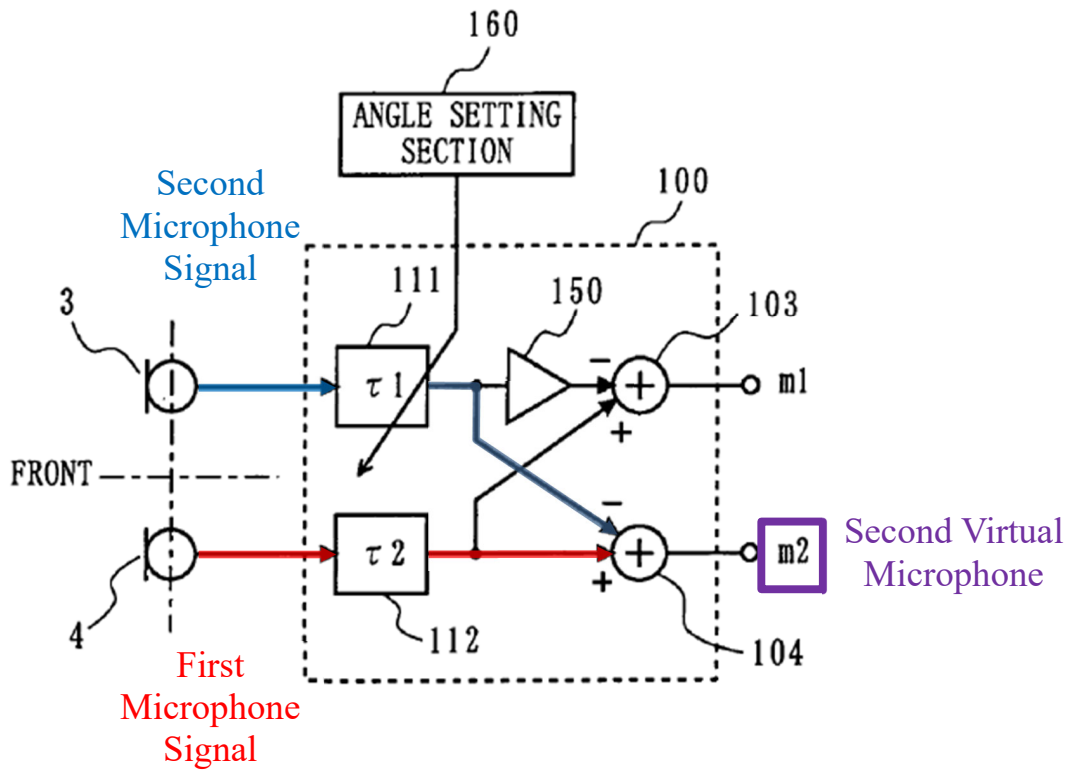
“signal output from the third microphone unit 3” (second microphone signal).

Ex. 1005, ¶¶ [0166], [0174]. Microphone unit 4 is a first physical microphone that generates the corresponding signal output, and microphone unit 3 is a second physical microphone that generates the corresponding signal output. Ex. 1005, ¶ [0151] (disclosing that “the microphone units are required to be a certain distance . . . apart from each other”).

f. [1E]: “a second virtual microphone comprising a second combination of the first microphone signal and the second microphone signal,”

75. In my opinion, Kanamori discloses or renders obvious this feature, consistent with the '691 patent's disclosure that a “virtual microphone[]” is “a microphone constructed using two or more omnidirectional microphones and associated signal processing.” Ex. 1001, 5:59-62. Kanamori discloses that **noise reference signal m2** is a second virtual microphone. Kanamori discloses that devices “often use a plurality of microphone units having non-directivity,” i.e., two or more omnidirectional microphones, “to obtain directivity by combining signals output from these microphone units.” Ex. 1005, ¶ [0151]. In Figure 17A, **noise reference signal m2** comprises a combination of “**a signal output from the fourth microphone unit 4**” and “**a signal output from the third microphone unit 3**,” corresponding to the first and second microphone signals, respectively. Ex. 1005, ¶¶ [0166], [0172], [0174].

FIG. 17A



Ex. 1005, Fig. 17A (annotated).

76. Kanamori also discloses signal processing associated with constructing **m2**. As depicted above, a “**signal output from the fourth microphone unit 4**” is delayed by “the second signal delaying section 112.” Ex. 1005, ¶¶ [0166], [0174]. A “**signal output from the third microphone unit 3**” is delayed by the “signal delaying section 111.” Ex. 1005, ¶¶ [0166], [0174]. A “second signal subtracting section 104” subtracts the **signal output from the signal delaying section 111** from the **signal output from the second signal delaying section 112** to obtain **m2**. Ex. 1005, ¶¶ [0166], [0174]. A POSITA would have understood that

these aspects of the directivity combining process constitute signal processing associated with the formation of virtual microphone **m2**.

77. In Figure 17A, “a plurality of microphone units having the same directivity characteristic are used to obtain . . . the noise reference signal.”

Ex. 1005, ¶ [0172]. As explained *supra* Section VII.D.1.b ([1A]), Kanamori discloses or renders obvious that microphone units 3 and 4 in Figure 17A can be omnidirectional microphones and accordingly discloses or renders obvious constructing **noise reference signal m2** using two omnidirectional microphones.

Ex. 1005, ¶¶ [0020], [0113], [0151], [0193].

- g. **[1F]: “the second virtual microphone having a second linear response to speech and a second linear response to noise,”**

78. In my opinion, Kanamori discloses this feature. As described for [1E], **noise reference signal m2** is obtained by subtracting a delayed version of the **signal output from the third microphone unit 3** from a delayed version of the **signal output from the fourth microphone unit 4**. Ex. 1005, ¶¶ [0166], [0174]; *supra* § VII.D.1.f. Kanamori discloses that this directivity combining process “is a linear process” and can be performed digitally. Ex. 1005, ¶¶ [0151], [0182]. A POSITA would have understood that **m2** has linear responses to speech and noise that can be simulated and visualized, as I explain in Section VII.B and further below, by expressing this

directivity-combining process as an equation, with values for each component of the directivity combining process:

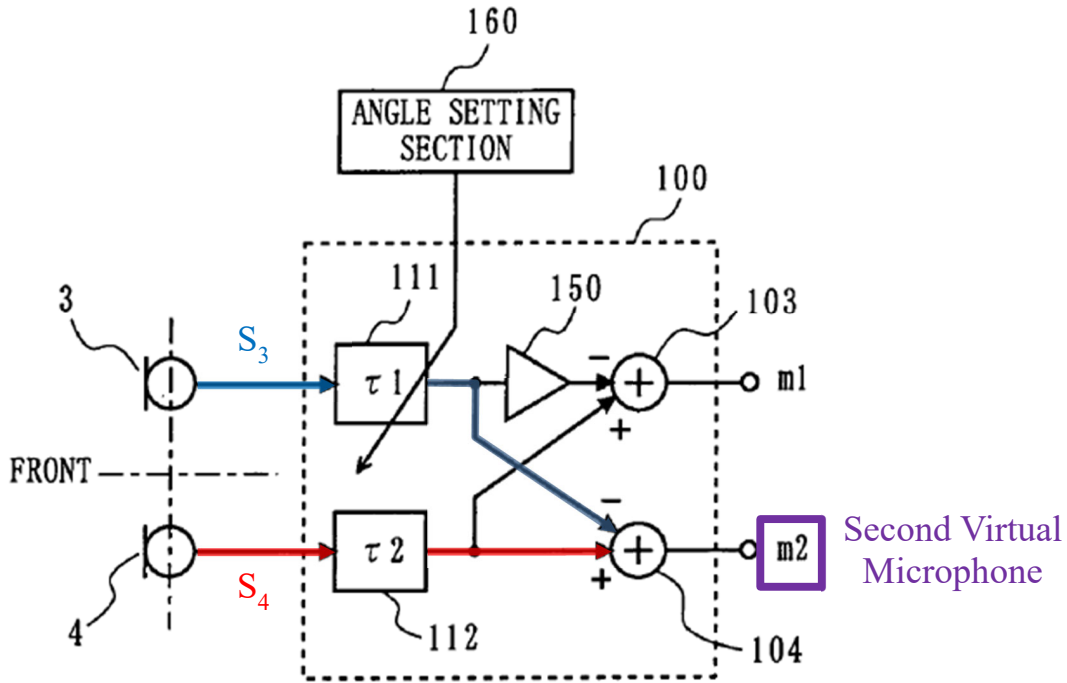
$$m2 = z^{-\tau 2 \cdot f_s} \cdot S_4 - z^{-\tau 1 \cdot f_s} \cdot S_3, \quad (11)$$

where $z^{-\tau 1 \cdot f_s}$, $z^{-\tau 2 \cdot f_s}$, S_3 , and S_4 mean the same thing here as they do for

Equation (10). *Supra* § VII.D.1.c ([1B]); Ex. 1005, ¶¶ [0166], [0174]-[0175].

Consistent with Kanamori's disclosure that the directivity combining process is a digital process, this equation is represented in the digital domain, which is evident by expressing the delay in terms of the number of samples (the desired time delay τ multiplied by the sampling frequency f_s). *Supra* § VII.B. This equation is further illustrated by Figure 17A:

F I G . 1 7 A



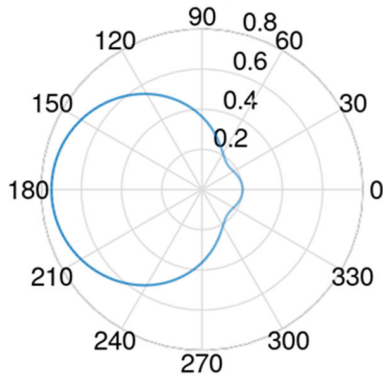
Ex. 1005, Fig. 17A (annotated). As for m1, “the delay amount τ_1 is determined so that the directions of a minimum sensitivity in the directivity are in the direction of the target sound.” Ex. 1005, ¶ [0167].

79. As explained *supra* Section VII.D.1.c ([1B]), the '691 patent discloses that the linear response to speech refers to the virtual microphone's near-field response, and the linear response to noise refers to the virtual microphone's far-field response. Kanamori does not explicitly describe noise reference signal m2 as having different linear responses, depending on the distance of the sound source to the array, but a POSITA, in my opinion, would have understood from Kanamori's description of noise reference signal m2, including Equation (11) above, that noise reference signal m2 has a second linear response to near-field speech and a second linear response to far-field noise.

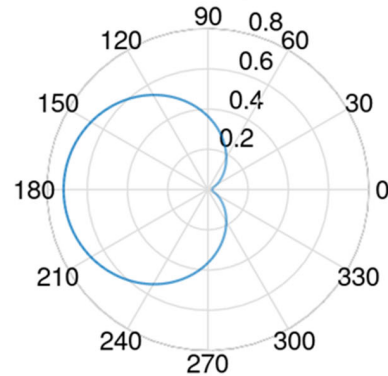
80. As explained *supra* Section VII.D.1.c ([1B]), it would have been obvious to implement Kanamori's microphone device in a device intended to receive near-field speech, such as a headset. When implemented in a device that receives near-field speech, Kanamori's specific structure in Figure 17A discloses that noise reference signal m2 has a second linear response to near-field speech and a second linear response to far-field noise. Using Equation (11) from Kanamori and the equations described above in Section VII.B for simulating the response of virtual microphones, I simulated and plotted the linear responses of noise reference

signal m2 to a 1000 Hz signal located 0.1 m (speech) from the microphone array and a 1000 Hz signal located 1 m (noise) from the microphone array when $\tau_1=0$ (linear response to speech (left) and linear response to noise (right)):

m2, Kanamori Fig. 17a at 0.1m



m2, Kanamori Fig. 17a at 1m



See *supra* § VII.B. These plots are the same types of plots that the '691 patent calls the “linear response[s]” of the virtual microphone. See Ex. 1001, Figs. 9-13, 2:41-60, 11:21-29, 12:20-33. Moreover, a POSITA would understand that Kanamori’s noise reference signal m2 has linear responses for the reasons explained in Section VII.B and because m2’s response to a sound source at a particular distance, frequency, and direction is linearly proportional to its input. For example, as depicted in the above plots, m2’s response to a 1000 Hz speech source at 0.1 m and 0° can be determined from the linear response plot by multiplying the input amplitude by about 0.2, and m2’s response to a 1000 Hz noise source at 1 m and 90° can be determined from the linear response plot by multiplying the input amplitude by about 0.38. This is because the above plots depict the response for a

unit acoustic pressure source, as explained in Section VII.B. As depicted, the direction of minimum sensitivity for both linear responses is towards the source of speech at 0°, as desired for $\tau_1=0$. Ex. 1005, ¶¶ [0167], [0175]; *supra* § VII.B.

81. Accordingly, Kanamori's noise reference signal m2 discloses a virtual microphone having a second linear response to speech and a second linear response to noise.

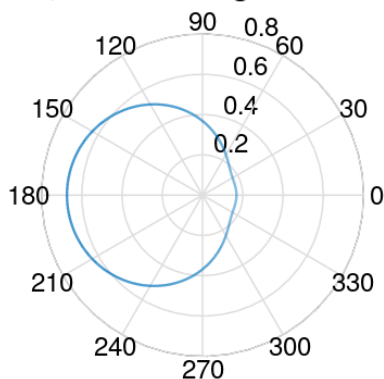
- h. [1G]: “the second linear response to noise being substantially similar to the first linear response to noise, and the second linear response to speech being substantially dissimilar to the first linear response to speech,”**

82. In my opinion, Kanamori and McCowan render obvious this feature. Kanamori discloses a variety of “use environments” and “purposes” for its microphone devices, many of which would not expect speech to always be in the near field. Ex. 1005, ¶ [0182]. For example, Kanamori discloses implementing its microphone device in a device such as a “video recorder, capable of collecting sounds.” Ex. 1005, ¶ [0151]. As such, Kanamori's directivity-combining process for forming m1 and m2 does not explicitly describe factoring in the distance of the speech source to the microphones because the amplitude differences at each microphone caused by amplitude decay from the distance traveled “are negligible for far-field sources.” Ex. 1006, 90; *see also* Ex. 1012 (Elko '328), ¶ [0021] (disclosing that when a sound source is located at a distance from a microphone

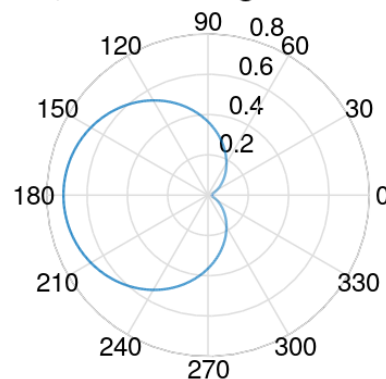
array “where farfield conditions are applicable . . . the audio signal arriving at [the microphone array] can be treated as a plane wave”).

83. For such devices that assume far-field speech, Kanamori discloses that the directivity patterns of **m1** and **m2** are “preferably such that, as for the direction of the target sound, there is a large difference in sensitivity between the signals **m1** and **m2**,” but, “for the directions other than the direction of the target sound, it is preferable that there is no difference [in sensitivity] therebetween.” Ex. 1005, ¶ [0168]. This can be seen in the plots of the linear responses of **main signal m1** and **noise reference signal m2** to a 1000 Hz signal originating 1 m from the microphone array, assuming a target-sound direction of 0°, which I generated as described in VII.B using the directivity combining process of Figure 17A from Kanamori and in the same manner described above for [1B] and [1F], *supra* §§ VII.D.1.c, VII.D.1.g.

m1, Kanamori Fig. 17a at 1m



m2, Kanamori Fig. 17a at 1m

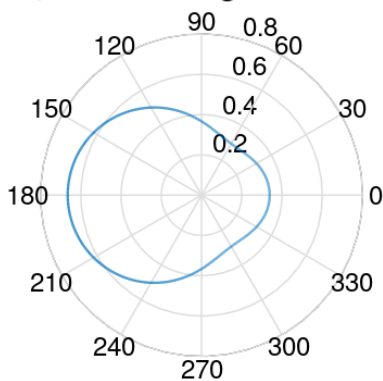


Kanamori discloses constructing m1 and m2 this way because, “to suppress, based on the noise reference signal, noise components included in the main signal under the circumstances where noise is coming from a plurality of directions,” the sensitivity to noise in the directivity patterns of m1 and m2 must be the same “irrespectively of the direction from which noise is coming.” Ex. 1005, ¶ [0168]. Accordingly, in my opinion, when both speech and noise originate in the far field, a POSITA would understand that Kanamori’s microphone device would suppress noise from most directions well but would struggle to suppress noise that originated from the same direction as the speech source because it could not differentiate between speech and noise arriving from the same direction. Ex. 1005, ¶ [0168]. While additional processing may be able to remove some of the noise (e.g., filtering by frequency), the processed signal will inevitably include at least some noise in many situations because, as can be seen in m2’s linear response at 1 m, noise reference signal m2 does not pick up any noise in the direction of speech, so the noise in the direction of speech received by main signal m1 cannot be suppressed using the noise reference signal.

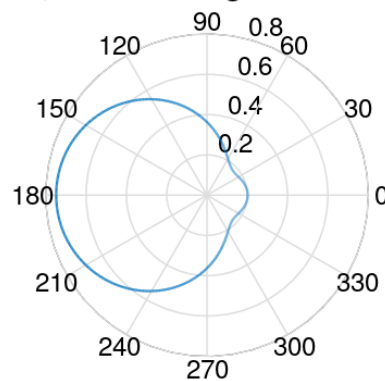
84. As explained *supra* Section VII.D.1.c ([1B]), it would have been obvious to implement Kanamori’s microphone device in a device intended to receive near-field speech, such as a headset. In such a context (e.g., headset), in my opinion, it would have been obvious to modify Kanamori’s noise reference signal

m2 to account for the differences in amplitude decay caused by distance traveled, which “can be significant in the near-field case.” Ex. 1006, 90. In my opinion, the need for such a modification would have been apparent to a POSITA because, unlike the far-field situation, Kanamori’s noise reference signal m2 is relatively sensitive to near-field speech, as depicted in the plots of the linear responses of main signal m1 and noise reference signal m2 to a 1000 Hz signal originating 0.1 m from the microphone array (near field), assuming a target-sound direction of 0°:

m1, Kanamori Fig. 17a at 0.1m



m2, Kanamori Fig. 17a at 0.1m



Supra § VII.B. Because “noise is suppressed mainly through a directivity combining process,” a POSITA would have recognized that noise reference signal m2 would make for a poor noise reference signal at this distance. Ex. 1005, ¶¶ [0083], [0182]. Kanamori discloses that, if the target sound enters m2 (as here), “detrimental effects, such as distortion or reduction in level, will occur in the processed sound.” Ex. 1005, ¶ [0083]. In my opinion, to avoid such detrimental

effects, a POSITA would have looked to McCowan to determine how to modify Kanamori's noise reference signal m_2 for use in a device such as a headset.

85. McCowan recognizes the above problem, stating that “leakage of the desired signal” is a common problem for noise-reference signals and is “particularly problematic” for speech signals. Ex. 1006, 88. McCowan teaches a technique for improving noise cancellation when a speech source is in the near field using near-field compensation in the signal path for generating a noise reference signal via a blocking matrix, as depicted in Figure 3, below. Ex. 1006, 87, 90-91. The purpose of the blocking matrix “is to block the desired signal from the adaptive noise estimate,” which requires that the desired signal “be time aligned and have equal amplitudes across all channels.” Ex. 1006, 93.

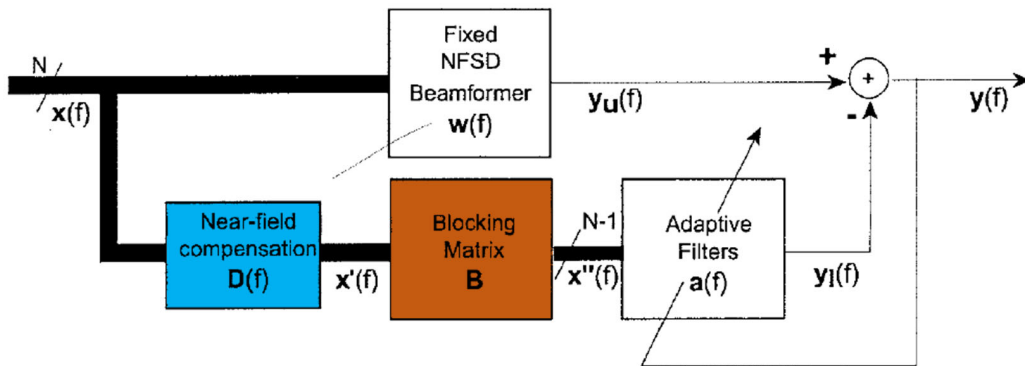


FIG. 3. Near-field adaptive beamformer.

Ex. 1006, Fig. 3 (annotated).

86. As McCowan explains, when the speech source is in the near field, a spherical wavefront model is “more accurate” than a planar signal wavefront

model and “must be employed.” Ex. 1006, 89. “For a near-field desired source, to align the desired signal on all channels, a **near-field compensation** must first be applied to the input channels prior to blocking.” Ex. 1006, 93. McCowan discloses two issues that **near-field compensation** must account for to “ensure[] that a true null exists” in the response of the blocking matrix “in the direction *and* distance corresponding to the desired source”: (1) “phase misalignment” and (2) “amplitude scaling.” Ex. 1006, 89-90, 93-94.

87. First, “the differences in distance to each sensor can be significant for a near-field source, resulting in phase misalignment across sensors.” Ex. 1006, 89. To properly align the phase, McCowan discloses the following equation representing “[t]he difference in propagation time to each microphone with respect to the reference microphone ($i=1$)”:

$$\tau_i = \frac{d_i - d_1}{c},$$

where d_i represents the distance from the source to the corresponding i^{th} microphone, d_1 is the distance from the source to the first (reference) microphone, and c is the speed of sound. Ex. 1006, 89-90.

88. Second, “the wavefront amplitude decays at a rate proportional to the distance traveled.” Ex. 1006, 90. “The resulting amplitude differences across sensors are negligible for far-field sources, but can be significant in the near-field

case.” Ex. 1006, 90. McCowan discloses a microphone-attenuation factor given by the following equation:

$$\alpha_i = \frac{d_1}{d_i},$$

where d_i and d_1 represent the same distance value as above. Ex. 1006, 90.

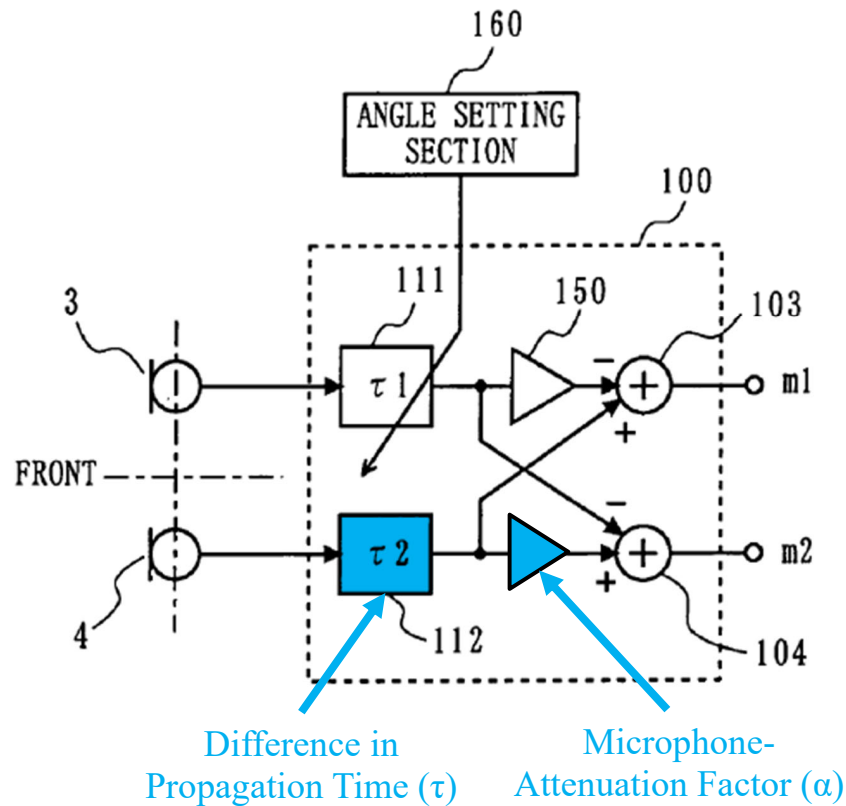
89. Combining both phase misalignment and amplitude scaling, McCowan discloses that “if $x_1(f)$ is the desired source at the reference microphone, the signal on the i^{th} microphone is given by

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

Ex. 1006, 90. In other words, McCowan discloses that **near-field compensation** requires delaying a reference signal x_1 by the difference in propagation time (τ_i) to the reference microphone and the i^{th} microphone, and multiplying by a gain equal to the distance from the source to the reference microphone divided by the distance from the source to the i^{th} microphone. Ex. 1006, 89-90. “Once this **near-field compensation** has been performed,” a **blocking matrix** can completely “block the desired signal from the adaptive noise estimate” because the desired signal is both time aligned and has equal amplitudes across the channels. Ex. 1006, 91, 93. The two channels will be equal in magnitude with respect to “the direction *and* distance corresponding to the desired source” and thus, when subtracting one from the other, the signals cancel each other out, producing “a true null.” Ex. 1006, 93-94.

90. In my opinion, it would have been obvious to modify Kanamori's noise reference signal m2 to incorporate McCowan's near-field compensation to address situations when the speech source is in the near field (e.g., a headset). In my opinion, a POSITA would have been motivated to make such a combination because it would yield substantial improvement in noise suppression in such situations because it would "ensure[] that a true null exists" in m2's directivity pattern "in the direction *and* distance corresponding to the desired source," thereby avoiding the "detrimental effects" caused by speech entering m2. Ex. 1005, ¶ [0083]; Ex. 1006, 93-94. Moreover, such a combination would require minimal modification to Kanamori's Figure 17A, as depicted below. Ex. 1006, 89-90, 93.

FIG. 17A



Ex. 1005, Fig. 17A (annotated).

91. Kanamori already discloses delaying the first microphone signal by d/c , where d is the distance between the microphones and c is the speed of sound.

Ex. 1005, ¶ [0175]. By dividing the distance by the speed of sound, this value corresponds to the time it takes for sound to travel from one microphone to the other along the axis of the microphones, i.e., the difference in propagation time.

Supra § VII.D.1.c ([1B]). When the speech source originates from a direction of 0° , as discussed above, Kanamori's delay $\tau 2$ and McCowan's difference in propagation time are identical because the distance from microphone unit 3 to the

speech source minus the distance from microphone unit 4 to the speech source will equal the distance between microphones 3 and 4. Ex. 1006, 89-90. Accordingly, while a POSITA would be motivated to modify Kanamori's delay τ_2 to McCowan's difference in propagation time τ_i to account for near-field speech originating from other directions, Kanamori's delay τ_2 requires no modification for near-field speech at 0° . In this situation, m_1 is completely unaffected. Because the linear responses of modified m_2 discussed in Section VII.B assumed that speech originated from a direction of 0° , the only necessary modification to Figure 17A was adding the microphone-attenuation factor.

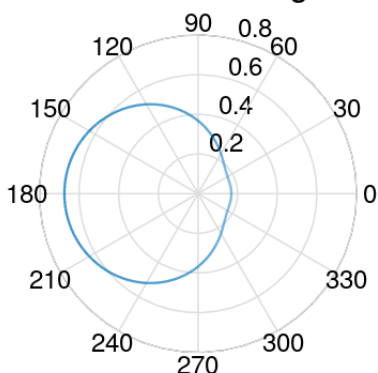
92. Moreover, Kanamori already discloses a signal amplifying section 150 for main signal m_1 . Ex. 1005, ¶¶ [0169], [0174]. Kanamori states that its "signal amplification ratio [is] set at approximately 0.85" to create imbalance in the signals that are subtracted at signal subtracting section 103, thereby increasing the sensitivity in the direction of speech. Ex. 1005, ¶¶ [0166], [0169]. Unlike m_1 , however, a POSITA would have been motivated to create a "true null" in the directivity pattern of m_2 corresponding to "the distance *and* direction" of the speech source to avoid the "detrimental effects" of speech entering the noise reference signal. Ex. 1005, ¶ [0083], Ex. 1006, 93-94. To create this null, McCowan teaches that multiplying the delayed signal by a gain (microphone-attenuation factor) is necessary to account for the difference in amplitude decay

caused by the different distances between the speech source and the microphone units. Ex. 1006, 90, 93-94. In my opinion, a POSITA would have been motivated to add this microphone-attenuation factor between signal delaying section 112 and signal subtracting section 104 so that “the desired signal” (speech) is both “time aligned and ha[s] equal amplitudes across [both] channels” entering signal subtracting section 104, ensuring “complete blocking” of the speech signal, as taught by McCowan. Ex. 1006, 93-94. Adding another signal amplifying section corresponding to McCowan’s microphone-attenuation factor for noise reference signal m2 would therefore have been routine and obvious to account for the amplitude decay due to the different distances traveled. Ex. 1005, ¶ [0169]; Ex. 1006, 90.

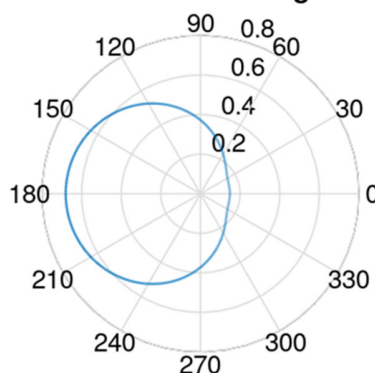
93. In sum, McCowan teaches that incorporating [near-field compensation](#) would align the speech signal in both channels entering signal subtracting section 104 in Kanamori. Ex. 1006, 93. In my opinion, a POSITA would understand that this would produce a null in m2’s linear response to near-field speech. Ex. 1006, 94 (“The near-field compensation effectively ensures that a true null exists in the beam-pattern of each blocking matrix row in the direction *and* distance corresponding to the desired source.”). This aligns with Kanamori’s objective of suppressing “the inclusion of the target sound in the noise reference signal.” Ex. 1005, ¶ [0083].

94. Using the modified Figure 17A above incorporating **near-field compensation**, I simulated and plotted m1's and m2's linear responses, respectively, to a 1000 Hz signal originating 1 m (noise) from the microphone array, assuming a speech direction of 0°:

m1, Modified Kanamori Fig. 17a at 1m



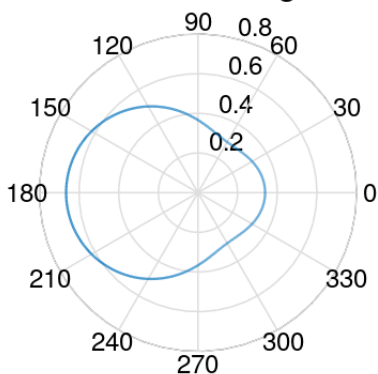
m2, Modified Kanamori Fig. 17a at 1m



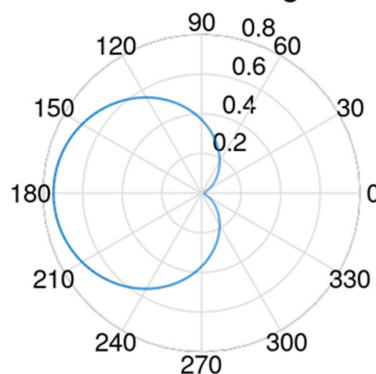
Supra § VII.B. Note that, as explained in Section VII.B, m1's linear response to noise for both original Figure 17A and modified Figure 17A are identical. These plots show that m1's linear response to noise is substantially similar to m2's linear response to noise. This aligns with Kanamori's objective that there is no difference in the sensitivity between m1 and m2 for noise "irrespectively of the direction from which noise is coming." Ex. 1005, ¶ [0168].

95. Using the modified Figure 17A above incorporating **near-field compensation**, I also simulated and plotted m1's and m2's linear responses, respectively, to a 1000 Hz signal originating 0.1 m (speech) from the microphone array, assuming a speech direction of 0°:

m1, Modified Kanamori Fig. 17a at 0.1m



m2, Modified Kanamori Fig. 17a at 0.1m



Supra § VII.B. Note that, as explained in Section VII.B, m1’s linear response to speech for both original Figure 17A and modified Figure 17A are identical. These plots show that m1’s linear response to speech is substantially dissimilar to m2’s linear response to speech in the direction of speech (0°). This aligns with Kanamori’s objective of including “a large difference in sensitivity between the signals m1 and m2” in the direction of speech. Ex. 1005, ¶ [0168].

96. In my opinion, a POSITA would have been motivated to make such a modification because, unlike Kanamori’s Figure 17A, it would allow the cancellation of noise originating from the same direction but at a different distance than the speech. *See* Ex. 1008 (Dugger), ¶ [0035] (disclosing that a “near-field gain factor” that compensates for the speaker being in the near field “allows preservation of a desired source such as the [speaker] based on distance as well as direction relative to [the microphone array], and provides additional attenuation of the [noise source] in the same direction as the speaker, but at a different distance

from the microphone array than the [speaker]”). A POSITA, in my opinion, would have been motivated to adjust the linear responses as described above so that the linear responses to noise of main signal m_1 and noise reference signal m_2 are substantially similar in all directions, making the estimated noise spectrum accurate in all directions, including the direction of the target speech, while maintaining Kanamori’s teaching that the linear responses to speech of m_1 and m_2 should be different in the direction of the speech source. Ex. 1005, ¶¶ [0103]-[0106] (explaining that the estimated noise spectrum “represents the noise components included in the spectrum $X(\omega)$ of the main signal” and is used to eventually calculate a filter coefficient used to filter out the noise components in the main signal and output the denoised signal), [0168], (explaining that the estimated noise spectrum will not be correctly calculated in the directions where the directivity patterns of m_1 and m_2 are different in shape).

97. In my opinion, a POSITA would have had a reasonable expectation of success in making such a combination because it is a combination of familiar elements according to known methods to yield predictable results. For example, Kanamori already discloses techniques for varying the value of delays, including τ_2 , providing a reasonable expectation that similar techniques could be used to adjust the difference in propagation time to account for different speaker positions. Ex. 1005, ¶¶ [0175], [0180], Figs. 17A, 18A. Likewise, Kanamori already

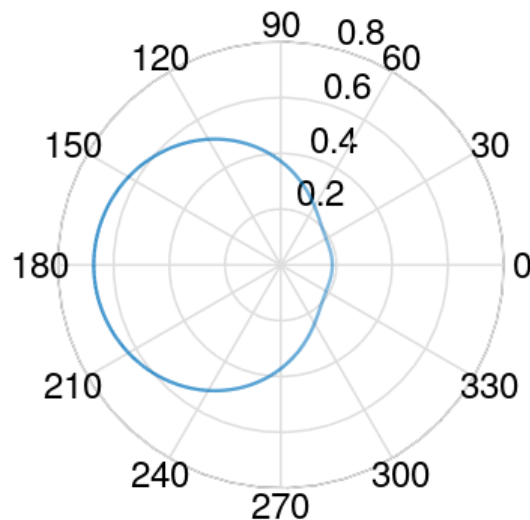
discloses a signal amplifying section 150, providing a reasonable expectation that a similar amplifying section could be added to account for amplitude decay.

Ex. 1005, ¶ [0169]. Moreover, similar to Kanamori's noise reference signal m2, McCowan discloses using this technique to generate a "noise estimate" that completely blocks the desired signal. Ex. 1006, 93.

- i. **[1H]: "wherein at least one of the first linear response to noise and the second linear response to noise is being non-zero in a direction towards a source of noise."**

98. In my opinion, Kanamori discloses this feature. As explained in Section VII.D.1.c ([1B]), Kanamori discloses the following first linear response to noise:

m1, Kanamori Fig. 17a at 1m



As depicted, m1's first linear response to noise is non-zero in all directions and therefore is non-zero in a direction towards a source of noise. *See also* Ex. 1005, ¶ [0167] (disclosing that m1 includes "components of the sound coming from a direction other than the direction of the target sound more than components of the target sound").

2. Dependent Claims 2-22

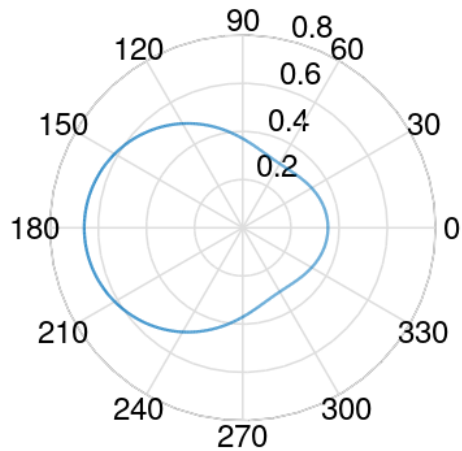
a. Claim 2: "The microphone array of claim 1, wherein the first and second physical microphones are omnidirectional."

99. In my opinion, Kanamori discloses or renders obvious that the first and second physical microphones are omnidirectional. *Supra* §§ VII.D.1.b ([1A]), VII.D.1.e ([1D]).

b. Claim 3: "The microphone array of claim 1, wherein the first linear response to speech is devoid of a null, wherein the speech is human speech."

100. In my opinion, Kanamori discloses this additional feature. As explained *supra* Section VII.D.1.c ([1B]), Kanamori discloses that m1 has the following first linear response to near-field speech:

m1, Kanamori Fig. 17a at 0.1m



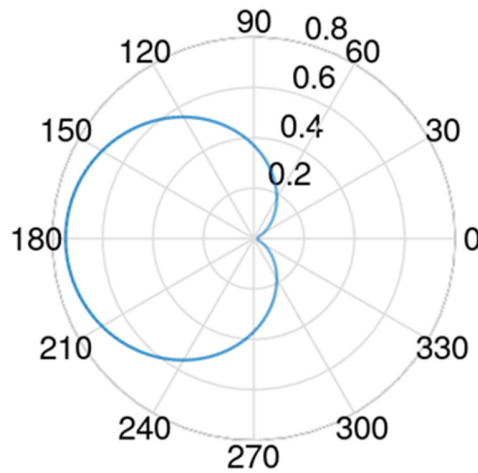
As depicted, m1's linear response to near-field speech lacks a null.

101. Additionally, Kanamori discloses that the speech is human speech. Kanamori discloses using its microphone device to receive "voice from the talker." Ex. 1005, ¶ [0013]; *see also* Ex. 1005, ¶ [0125], Fig. 9 (disclosing a "talker" approximately 50 cm from a microphone array). In my opinion, a POSITA would understand that "voice from the talker" refers to human speech. Ex. 1005, ¶ [0013].

c. Claims 4-7

102. In my opinion, Kanamori and McCowan render obvious the additional features of claims 4-7. As explained *supra* Section VII.D.1.h ([1G]), it would be obvious to modify Kanamori's noise reference signal m2 (the second virtual microphone) to produce the following second linear response to speech:

m2, Modified Kanamori Fig. 17a at 0.1m



- i. **Claim 4: “The microphone array of claim 3, wherein the second linear response to speech of the second virtual microphone includes a single null oriented in a direction toward a source of the speech.”**

103. As depicted, m2’s linear response to speech includes a single null oriented at 0°, which in the above plot is the direction toward a source of the speech. *Supra* §§ VII.D.1.h ([1G]), VII.B.

- ii. **Claim 5: “The microphone array of claim 4, wherein the single null is a region of the second linear response to speech having a measured response level that is lower than the measured response level of any other region of the second linear response to speech.”**

104. As depicted, the null at 0° represents the lowest measured response level of any region in m2’s linear response to speech. *Supra* §§ VII.D.1.h ([1G]),

VII.B; *see* Ex. 1005, ¶ [0167] (disclosing that m2's direction of minimum sensitivity points towards speech).

- iii. **Claim 6: “The microphone array of claim 4, wherein the second linear response to speech includes a primary lobe oriented in a direction away from the source of the speech.”**

105. As depicted, the primary lobe of m2's linear response is oriented towards 180°, away from the source of speech at 0°. *Supra* §§ VII.D.1.h ([1G]), VII.B.

- iv. **Claim 7: “The microphone array of claim 6, wherein the primary lobe is a region of the second linear response to speech having a measured response level that is greater than the measured response level of any other region of the second linear response to speech.”**

106. As depicted, the primary lobe of m2's linear response to speech is oriented towards 180° and is the region having the highest measured response level of any region of the second linear response to speech. *Supra* §§ VII.D.1.h ([1G]), VII.B.

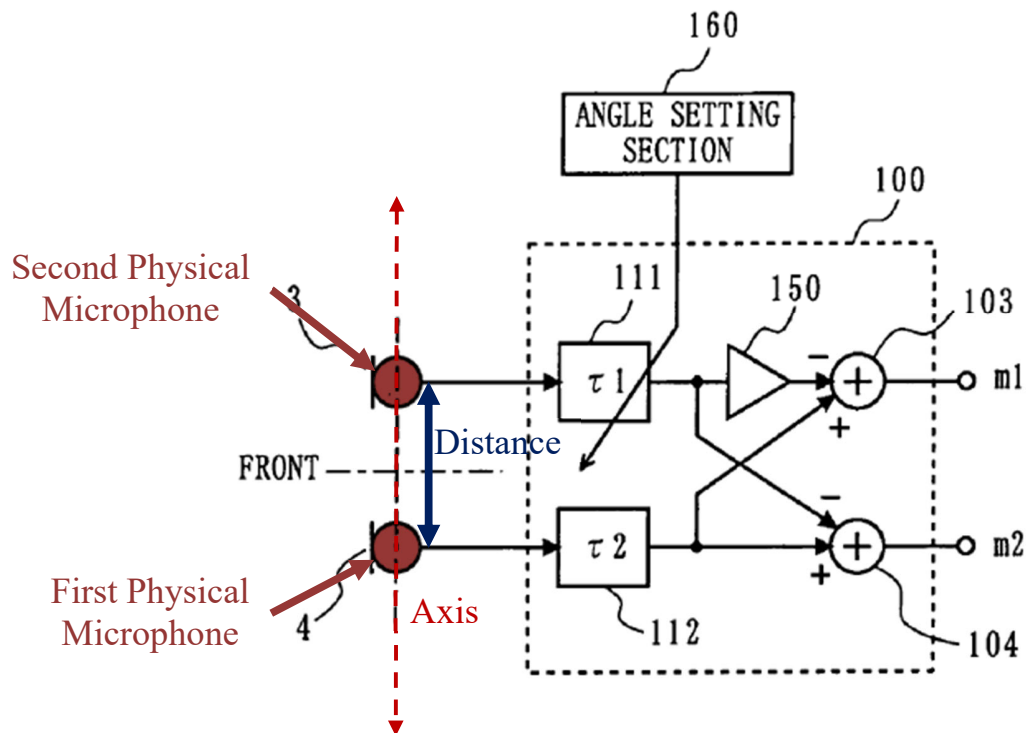
- d. **Claim 8: “The microphone array of claim 4, wherein the first physical microphone and the second physical microphone are positioned along an axis and separated by a first distance.”**

107. In my opinion, Kanamori discloses this additional feature. Kanamori discloses that the “microphone units 3 and 4 are placed on an **axis** (denoted by a one-dot chain line . . .) perpendicular to a straight line oriented to the front.”

Ex. 1005, ¶ [0160]. Kanamori also discloses that “the microphone units are required to be a certain **distance** (normally, 1 cm to 5 cm) apart from each other.”

Ex. 1005, ¶ [0151]. For example, in my simulations of Kanamori above, I set the separation between the first physical microphone and the second physical microphone to 2 cm. *Supra* § VII.B.

F I G. 1 7 A

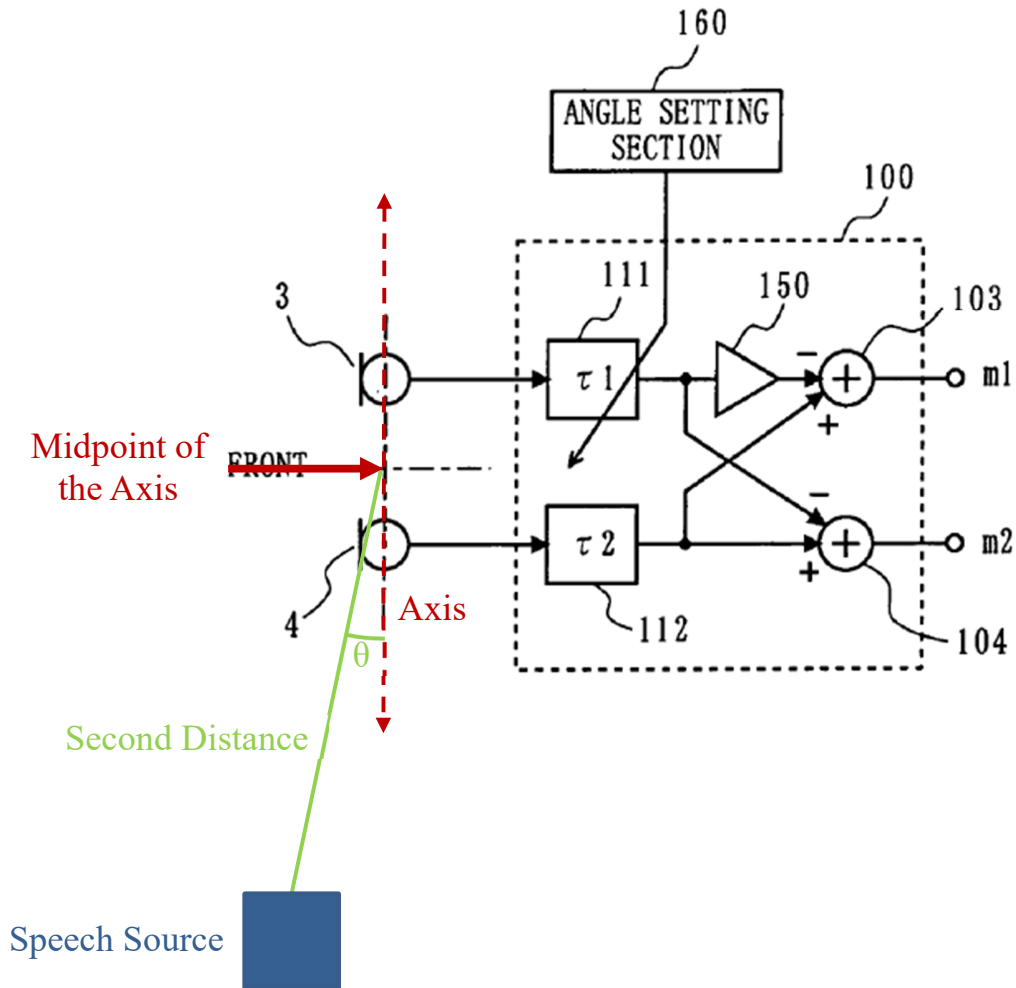


Ex. 1005, Fig. 17A (annotated).

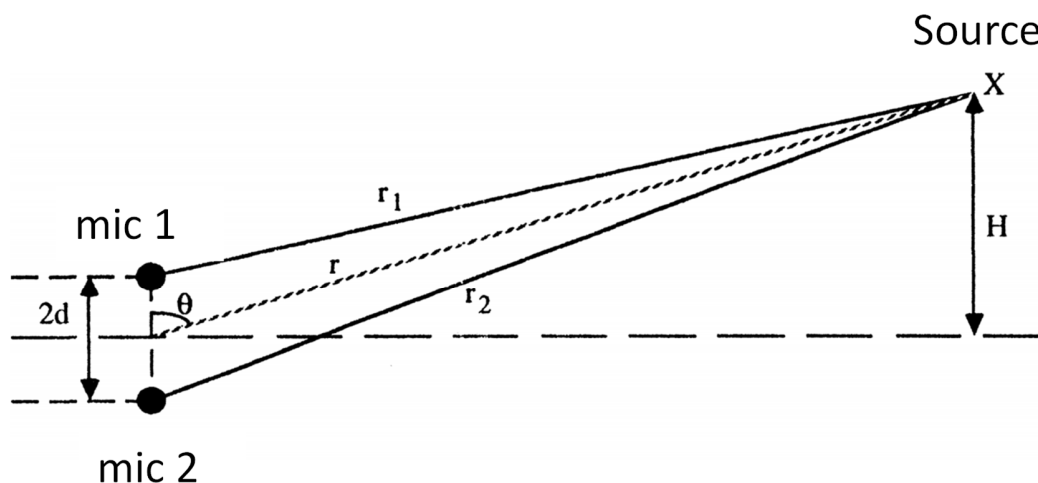
- e. **Claim 9: “The microphone array of claim 8, wherein a midpoint of the axis is a second distance from the speech source that generates the speech, wherein the speech source is located in a direction defined by an angle relative to the midpoint.”**

108. In my opinion, Kanamori discloses this additional feature. Kanamori discloses that the microphone units are “placed on an **axis**” and “required to be a certain distance (normally, 1 cm to 5 cm) apart from each other.” Ex. 1005, ¶¶ [0151], [0160]; *supra* § VII.D.2.d (claim 8). A midpoint of the **axis** is located halfway between the microphones along that **axis**. This midpoint is necessarily located at a distance (**second distance**) from the speech source. Kanamori discloses an “angle setting section” that “can change a signal delay amount τ_1 ” so that “the target sound direction can be changed in a range of ± 90 degrees with respect to the front direction.” Ex. 1005, ¶ [0175]. In my opinion, a POSITA would understand that the speech source is located in a direction defined by an angle relative to the midpoint of the **axis**.

FIG. 17A



Ex. 1005, Fig. 17A (annotated). This is explained further by my discussion of θ in Section VII.B. For example, as I explain in Section VII.B, a two-microphone array can be defined by the microphone spacing ($2d$), average radius, r , and radii between mic 1 and mic 2 and the source location X (at angle θ), given by r_1 and r_2 , respectively. Ex. 1016, 151-152.



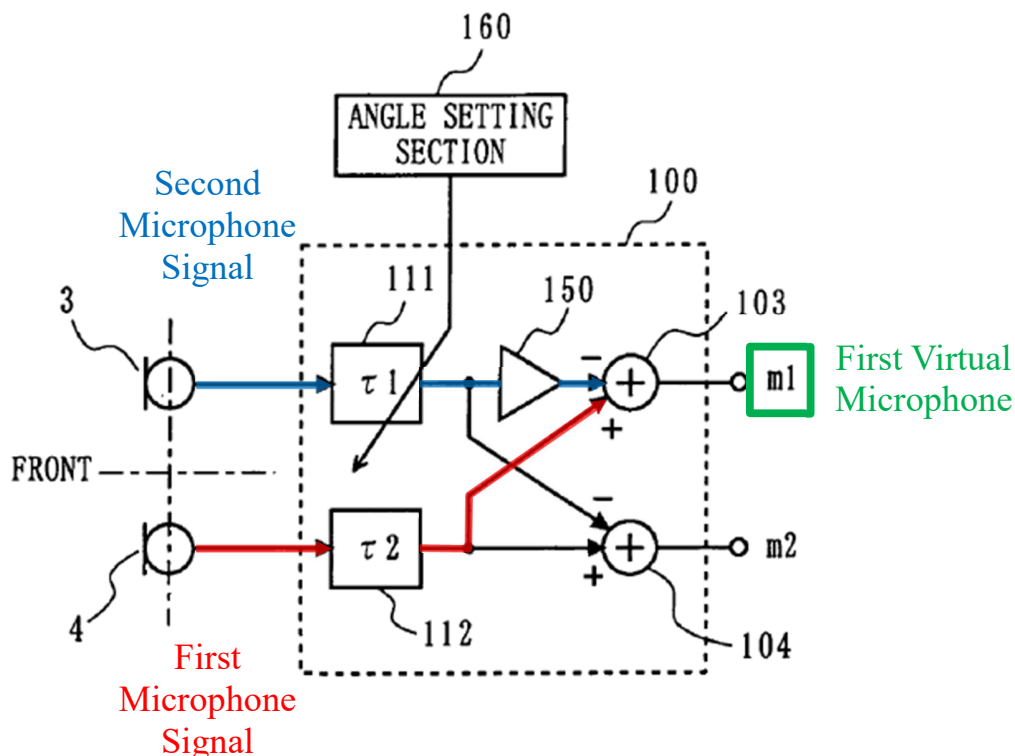
Ex. 1016, 152 (modified). In this figure, the midpoint of the axis is located halfway between mic 1 and mic 2 and is a second distance r from speech source X , which is located in a direction defined by an angle θ relative to the midpoint. In my simulations discussed above for the near-field situation, the speech source is located at a 0.1 m from the midpoint of the axis (the claimed “second distance”) and at an angle of 0° relative to the midpoint. *Supra* §§ VII.D.1.c ([1B]), VII.D.1.d ([1C], VII.B.

- f. **Claim 10: “The microphone array of claim 9, wherein the first virtual microphone comprises the second microphone signal subtracted from the first microphone signal.”**

109. In my opinion, Kanamori discloses this additional feature. As explained for [1A], Kanamori’s signal subtracting section 103 subtracts the second

microphone signal from the first microphone signal to obtain first virtual microphone m1. *Supra* § VII.D.1.b.

F I G. 1 7 A



Ex. 1005, Fig. 17A (annotated).

g. Claim 11: “The microphone array of claim 10, wherein the first microphone signal is delayed.”

110. In my opinion, Kanamori discloses or Kanamori and McCowan render obvious this additional feature. Kanamori discloses that a “**signal output from the fourth microphone unit**” (the first microphone signal) is delayed by “the second signal delaying section 112.” Ex. 1005, ¶¶ [0166], [0174]; *supra*

§ VII.D.1.b ([1A]). Additionally, it would have been obvious to modify the delay to account for phase misalignment in view of McCowan. *Supra* § VII.D.1.h ([1G]).

h. Claim 12: “The microphone array of claim 11, wherein the delay is raised to a power that is proportional to a time difference between arrival of the speech at the first virtual microphone and arrival of the speech at the second virtual microphone.”

111. The '691 patent defines a value γ as follows:

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

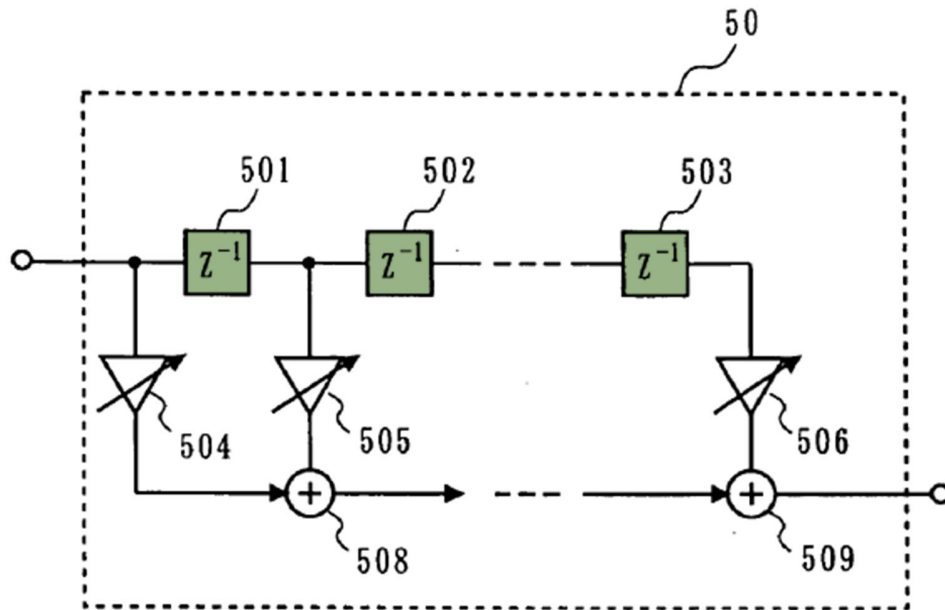
where “[t]he distances d_1 and d_2 are the distance from O_1 and O_2 to the speech source . . . , and γ is their difference divided by c , the speed of sound, and multiplied by the sampling frequency f_s .” Ex. 1001, 10:50-11:4. The '691 patent later refers to γ as “the time difference between arrival of speech at V_1 compared to V_2 .” Ex. 1001, 14:66-15:2.

112. Based on the '691 patent's reference to γ as “the time difference between arrival of speech at V_1 compared to V_2 ,” Ex. 1001, 14:66-15:2, which I have applied as the definition of “a time difference between arrival of the speech at the first virtual microphone and arrival of the speech at the second virtual microphone” for the purpose of this declaration, Kanamori discloses or Kanamori and McCowan render obvious this additional feature. Kanamori discloses that the delay amount τ_2 at signal delaying section 112 is d/c , “where d is a distance between the microphone units and c is the speed of sound.” Ex. 1005, ¶ [0175]. As

explained for [1G], when the speech originates from a direction of 0° , d/c is the time difference between arrival of the speech at microphone unit 3 and microphone unit 4. *Supra* § VII.D.1.h ([1G]). Additionally, as explained for [1G], it would have been obvious to modify the delay τ_2 to McCowan's difference in propagation time τ_i , which, with reference to Figure 17A, is the distance from microphone unit 3 to the speech source minus the distance from microphone unit 4 to the speech source divided by the speed of sound. Ex. 1006, 89-90; *supra* § VII.D.1.h ([1G]).

113. Kanamori discloses that “a digital process can be used” for the “directivity combining process.” Ex. 1005, ¶ [0151]. In digital processes, signals are represented by discrete samples, where the number of samples over a particular time period of the signal depends on the sampling frequency (signal samples per unit time). *See* Ex. 1005, ¶ [0108]; *supra* § VII.B. For digital processes in the microphone-array art, delays are commonly expressed as z^{-s} , where s constitutes the numbers of samples by which the process is delayed. For example, Figure 5 of Kanamori discloses several **signal delaying sections 501, 502, and 503**, where “each signal delaying section is connected in series to each other for delaying a received signal by one sample” consecutively. Ex. 1005, ¶ [0108], Fig. 5.

FIG. 5



Ex. 1005, Fig. 5 (annotated). In Figure 5, z^{-1} refers to a delay by one sample.

Ex. 1005, ¶ [0108], Fig. 5. In my opinion, a POSITA would understand that determining how many samples exist over a particular time period is a simple matter of multiplying the time period by the sampling frequency. *Supra* § VII.B. Because the signal combining process is a digital one, a POSITA would therefore understand that delaying the signal output from microphone unit 4 by time τ_2 or the difference in propagation time refers to delaying this signal by the number of samples corresponding to the delayed time period, which is expressed as

$$z^{-\tau \cdot f_s},$$

where τ is the desired time delay and f_s is the sampling frequency. *Supra* § VII.B

114. Accordingly, in my opinion, a POSITA would understand that, by delaying the first microphone signal by the time it takes for a soundwave to travel the distance between the microphones (either via Kanamori's τ_2 or McCowan's difference in propagation time), the delay z is raised to a power that is proportional to a time difference between arrival of the speech at the first virtual microphone and arrival of the speech at the second virtual microphone.

- i. **Claim 13: "The microphone array of claim 11, wherein the delay is raised 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 the speech source and the fourth distance being between the second physical microphone and the speech source."**

115. In my opinion, Kanamori discloses or Kanamori and McCowan render obvious this additional feature. As explained for claim 12, Kanamori discloses, or Kanamori and McCowan render obvious, delaying the signal output from microphone unit 4 by the following amount:

$$z^{-\tau \cdot f_s},$$

where τ is the desired time delay (either Kanamori's τ_2 or McCowan's difference in propagation time) and f_s is the sampling frequency. *Supra* § VII.D.2.h. As explained for [1G], when the speech source is at 0° , τ_2 is proportional to a distance between the first physical microphone and the speech source subtracted from a

distance between the second physical microphone and the speech source. *Supra* § VII.D.1.h. Also as explained for [1G], McCowan's difference in propagation time is proportional to a distance between the first physical microphone and the speech source subtracted from a distance between the second physical microphone and the speech source. *Supra* § VII.D.1.h.

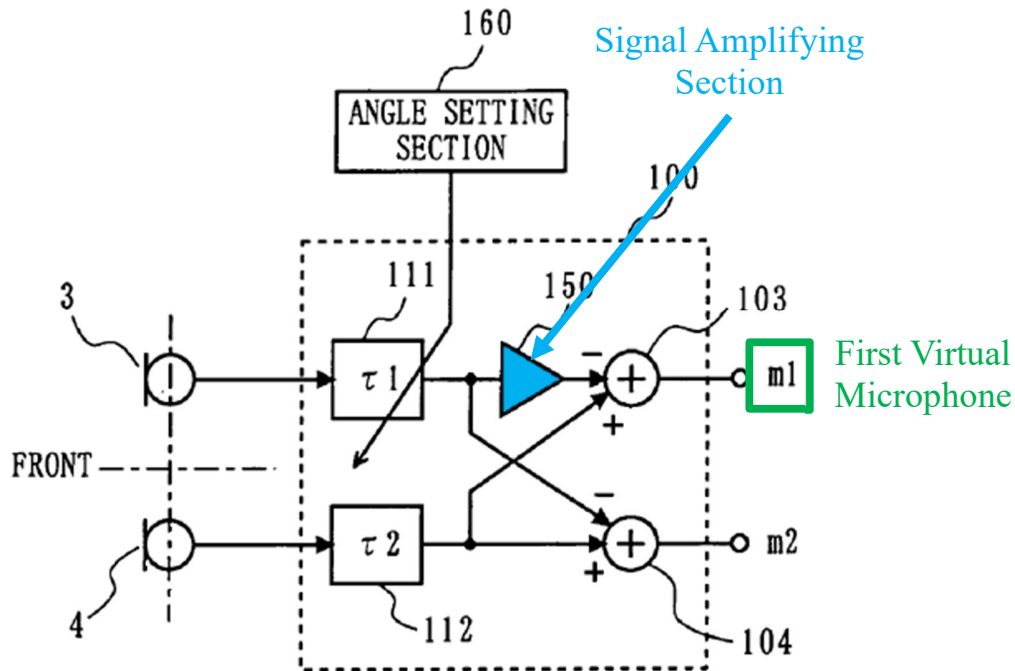
116. Accordingly, in my opinion, a POSITA would understand that Kanamori discloses, or Kanamori and McCowan render obvious, a delay raised to a power that is proportional to a sampling frequency f_s multiplied by a quantity equal to a third distance subtracted from a fourth distance, the third distance being between the first physical microphone and the speech source and the fourth distance being between the second physical microphone and the speech source.

- j. **Claim 14: “The microphone array of claim 10, wherein the second microphone signal is multiplied by a ratio, wherein the ratio is a ratio of a third distance to a fourth distance, the third distance being between the first physical microphone and the speech source and the fourth distance being between the second physical microphone and the speech source.”**

117. In my opinion, Kanamori and McCowan render obvious this additional feature. Kanamori discloses a [signal amplifying section 150](#) for [main signal m1](#). Ex. 1005, ¶¶ [0169], [0174]. Kanamori states that its “signal amplification ratio [is] set at approximately 0.85” to create imbalance in the signals

that are subtracted at signal subtracting section 103, thereby increasing the sensitivity in the direction of speech. Ex. 1005, ¶¶ [0166], [0169].

FIG. 17A



Ex. 1005, Fig. 17A (annotated).

118. Kanamori includes this [signal amplifying section](#) because, assuming a far-field speech source where the differences in amplitude decay are negligible across the microphones, phase-aligned speech signals in the two channels could cancel each other out at signal subtracting section 103, producing a null in the directivity pattern of [m1](#) in the direction of the speech source. Ex. 1005, Fig. 17B (depicting null in m2's directivity pattern in the direction of the speech source); Ex. 1006, 90; *supra* § VII.D.1.h ([1G]). In other words, if a speech signal has the

same amplitude at each microphone by virtue of the speech originating in the far field, and one microphone signal is delayed such that the two signals entering the signal subtracting section 103 are phase-aligned, then the two speech signals will be equal in magnitude and phase and therefore cancel each other out when subtracted. This would dramatically reduce the efficacy of Kanamori's noise-suppression technique, which is premised on "suppress[ing], based on the noise reference signal, noise components included in the main signal under the circumstances where noise is coming from a plurality of directions." Ex. 1005, ¶ [0168]. If **main signal m1** did not include the speech components due to a null in the direction of speech, Kanamori's system would not work.

119. As discussed for [1G], it would have been obvious to combine Kanamori and McCowan to implement McCowan's near-field compensation for when the speech source is in the near field. *Supra* § VII.D.1.h ([1G]). As further discussed for [1G], it would have been obvious to modify Kanamori's τ_2 to implement McCowan's τ_i to account for near-field speech originating from other directions. *Supra* § VII.D.1.h ([1G]). This aligns with Kanamori's disclosure that Figure 17A can change the target sound direction "in a range of ± 90 degrees with respect to the front direction." Ex. 1005, ¶ [0175].

120. Unlike far-field speech, the "resulting amplitude differences across sensors" from wavefront amplitude decay "can be significant in the near-field

case.” Ex. 1006, 90. Due to the effect of amplitude decay, in my opinion, a POSITA would have been motivated to modify Kanamori’s signal amplifying section 150 to implement McCowan’s microphone-attenuation factor to ensure that speech is not attenuated or completely canceled in **main signal m1** when approaching from different directions. Ex. 1006, 90. To demonstrate how this is possible, I will analyze a few scenarios. As discussed for [1G], speech may approach the microphone array from 0°. *Supra* § VII.D.1.h ([1G]). In such a situation, the speech would travel a shorter distance to microphone unit 4 than it would to microphone unit 3, meaning the speech signal would have a greater amplitude at microphone unit 4 than it would at microphone unit 3. Ex. 1006, 90. In such a situation, Kanamori’s amplification ratio of 0.85 would create further imbalance in the two channels.

121. If, on the other hand, speech approached the microphone array from 180°, the speech would travel a shorter distance to microphone unit 3 than it would to microphone unit 4, meaning the speech signal would have a greater amplitude at microphone unit 3 than it would at microphone unit 4. Ex. 1006, 90. In this situation, Kanamori’s amplification ratio would not necessarily ensure that the two channels are imbalanced as they enter signal subtracting section 103 because the larger speech amplitude in the upper channel of Figure 17A would be reduced by a factor of 0.85, driving its amplitude toward the smaller speech amplitude in the

lower channel of Figure 17A. Ex. 1005, ¶ [0169], Fig. 17A. Indeed, if the speech signal was 0.123 m from the microphone array—within the typical range of .08-.13 m for headset devices, *supra* § VII.D.1.c ([1B])—then the speech signal would be completely canceled in main signal m1 (assuming a 2 cm microphone spacing in this example). Ex. 1005, ¶ [0151] (microphone spacing between 1-5 cm); Ex. 1006, 90; Ex. 1010 (Taenzer), ¶¶ [0009], [0078] (disclosing a range between 0.08-0.13 m from the microphones in a headset to a speech source). This is understood from McCowan’s disclosure of how a soundwave’s amplitude decays based on distance. Ex. 1006, 90. At a distance of 0.123 m from the midpoint of the microphone array, the amplitude of the speech signal at microphone unit 4 will be $g = \left(\frac{0.123-0.01}{0.123+0.01} \right) \approx 0.85$ the amplitude at microphone unit 3—exactly the same as the gain at signal amplifying section 150. Ex. 1006, 90. This means that multiplying the signal output from microphone unit 3 by 0.85 will result in two phase-aligned speech signals with equal amplitudes entering the signal subtracting section 103, meaning that the speech signal will be canceled out in the main signal m1. In my opinion, a POSITA would seek to avoid this outcome.

122. The same issue could occur with different combinations of distances and angles. For example, in the examples discussed in previous sections, the speech source is assumed to be located 0.1 m from the midpoint of the array for the near-field case. If the speech source was located at an angle of 144.45°, the speech

signal would again be canceled out in the main signal m1. *See supra* § VII.B (discussing equations for determining distances from the source to each microphone based on the distance to the midpoint and the law of cosines).

123. In my opinion, a POSITA would have been motivated to modify signal amplifying section 150 in Figure 17A to multiply “the signal output from the signal delaying section 111” by McCowan’s microphone-attenuation factor, should the microphone device instead anticipate speech from a direction other than 0°. Ex. 1005, ¶ [0166]. If the speech source approached the array from an angle greater than 90°, it would be closer to microphone unit 3, causing the microphone-attenuation factor to be greater than 1. Ex. 1006, 90. If it approached the array from an angle less than 90°, it would be closer to microphone unit 4, causing the microphone-attenuation factor to be less than 1. Ex. 1006, 90. Modifying the amplification ratio of [signal amplifying section 150](#) to McCowan’s microphone-attenuation factor would therefore ensure that the microphone device could be configured to receive speech from any angle between 0° and 180° without risking a null in the direction of speech, improving m1’s sensitivity to speech. Ex. 1005, ¶¶ [0169], [0175]; Ex. 1006, 90.

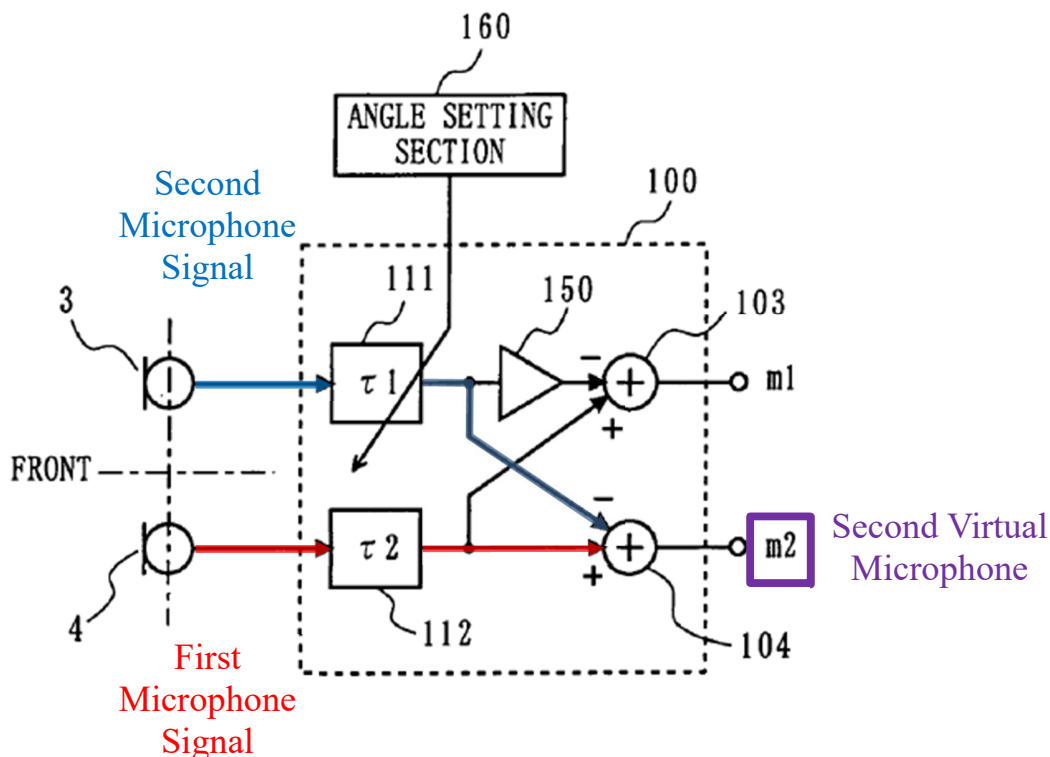
124. A POSITA would have had a reasonable expectation of success in making such a combination because it is a combination of familiar elements according to known methods to yield predictable results. For example, Kanamori

already discloses techniques for varying aspects of the directivity combining section 100, such as the value of delay τ_1 , to account for a target sound originating from various directions, providing a reasonable expectation of success of also adjusting the amplification ratio at signal amplifying section 150 based on known and/or expected distances between the speech source and the microphone distances. Ex. 1005, ¶ [0175]. Moreover, as explained for [1G], a POSITA would have had a reasonable expectation of successfully implementing McCowan's microphone-attenuation factor for the noise reference signal. *Supra* § VII.D.1.h ([1G]). It would have been routine for a POSITA to apply this same microphone-attenuation factor for the main signal.

k. Claim 15: “The microphone array of claim 9, wherein the second virtual microphone comprises the first microphone signal subtracted from the second microphone signal.”

125. In my opinion, Kanamori renders obvious this additional feature. Kanamori discloses that the second virtual microphone comprises the second microphone signal subtracted from the first microphone signal. *Supra* § VII.D.1.f ([1E]).

FIG. 17A



Ex. 1005, Fig. 17A (annotated).

126. A POSITA would have known that the sign of the output of signal subtracting section 104 (either positive or negative) has no impact on m2's linear responses to sounds at various distances. *Supra* § VII.B. For example, equations (5a)-(5c) from Section VII.B shows that the pressure at a virtual microphone is modeled using the absolute value, so changing the sign of the differenced quantities will not change the linear response of the virtual microphones. Accordingly, switching the positive and negative inputs to signal subtracting section 104 would not affect either the sensitivity or the directivity of m2. In my

opinion, switching the positive and negative inputs to signal subtracting section 104 would have been obvious to try because there are only two options for which channel enters the positive input and which channel enters the negative input. In my opinion, because both arrangements produce equivalent linear responses and switching the positive and negative inputs requires only a simple modification to Figure 17A's structure, it would have been obvious to construct m2 by subtracting the first microphone signal from the second microphone signal. Moreover, Kanamori discloses various embodiments where the terminals change, further motivating a POSITA to try the alternative arrangement and providing a reasonable expectation of success in achieving this alternative arrangement. *See* Ex. 1005, Figs. 17A, 18A (switching the positive and negative inputs to signal subtracting section 103).

**I. Claim 16: “The microphone array of claim 15,
wherein the first microphone signal is delayed.”**

127. In my opinion, Kanamori discloses, or Kanamori and McCowan render obvious, this additional feature. *Supra* § VII.D.2.g (claim 11).

- m. **Claim 17: “The microphone array of claim 16, wherein the delay is raised to a power that is proportional to a time difference between arrival of the speech at the first virtual microphone and arrival of the speech at the second virtual microphone.”**

128. In my opinion, Kanamori discloses, or Kanamori and McCowan render obvious, this additional feature. *Supra* § VII.D.2.h (claim 12).

- n. **Claim 18: “The microphone array of claim 16, wherein the power is proportional to a sampling frequency multiplied by a quantity equal to a third distance subtracted from a fourth distance, the third distance being between the first physical microphone and the speech source and the fourth distance being between the second physical microphone and the speech source.”**

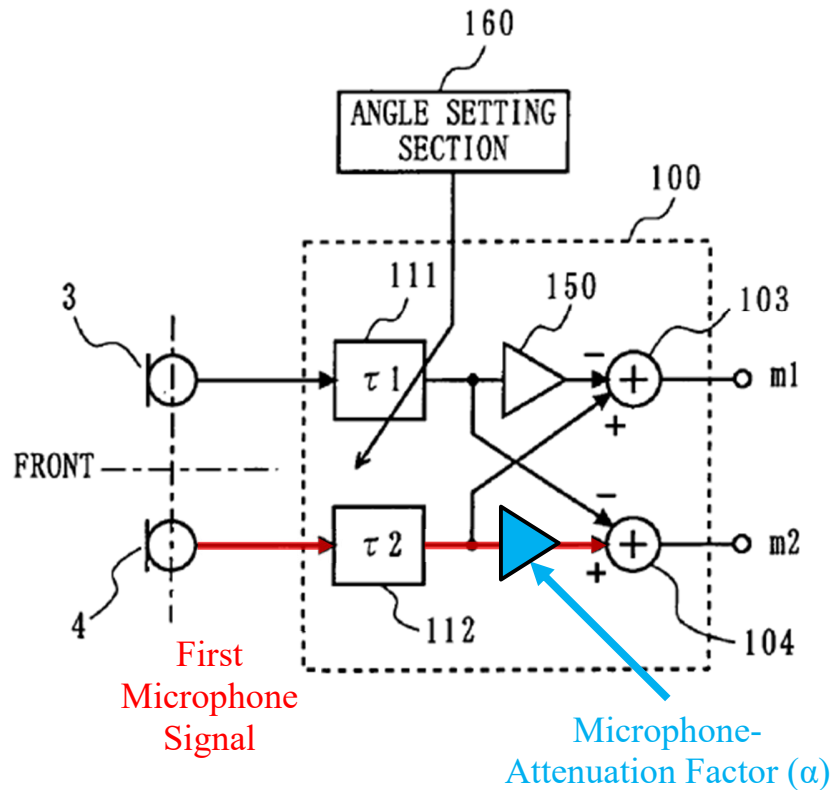
129. In my opinion, Kanamori discloses, or Kanamori and McCowan render obvious, this additional feature. *Supra* § VII.D.2.i (claim 13).

- o. **Claim 19: “The microphone array of claim 18, wherein the first microphone signal is multiplied by a ratio, wherein the ratio is a ratio of the third distance to the fourth distance.”**

130. In my opinion, Kanamori and McCowan render obvious this additional feature. As explained for [1G], it would have been obvious to modify Kanamori to implement McCowan’s microphone-attenuation factor, which corresponds to the distance from a reference microphone to the speech signal divided by the distance from the i^{th} microphone to the speech signal. *Supra* § VII.D.1.h; Ex. 1006, 90. Implementing this microphone-attenuation factor into

Kanamori's Figure 17A, it would have been obvious to multiply the **first microphone signal** by the **microphone-attenuation factor**, which with reference to Figure 17A is the distance from microphone unit 4 to the speech source (third distance) divided by the distance from microphone unit 3 to the speech source (fourth distance). *Supra* § VII.D.1.h ([1G]).

F I G. 1 7 A

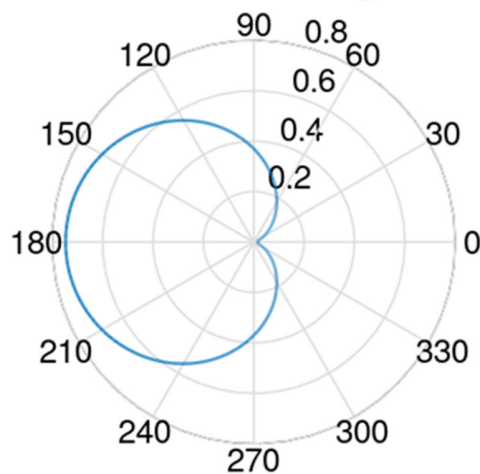


Ex. 1005, Fig. 17A (annotated).

- p. **Claim 20: “The microphone array of claim 4, wherein the single null is located at a distance from at least one of the first physical microphone and the second physical microphone where the source of the speech is expected to be.”**

131. In my opinion, Kanamori and McCowan render obvious this additional feature. As explained for [1G] and claim 4, it would have been obvious to modify Kanamori’s noise reference signal m2 to include a single null oriented at 0°, which is the direction toward a source of the speech in the plot of m2’s linear response depicted below. *Supra* §§ VII.D.1.h ([1G]), VII.D.2.c.i (claim 4).

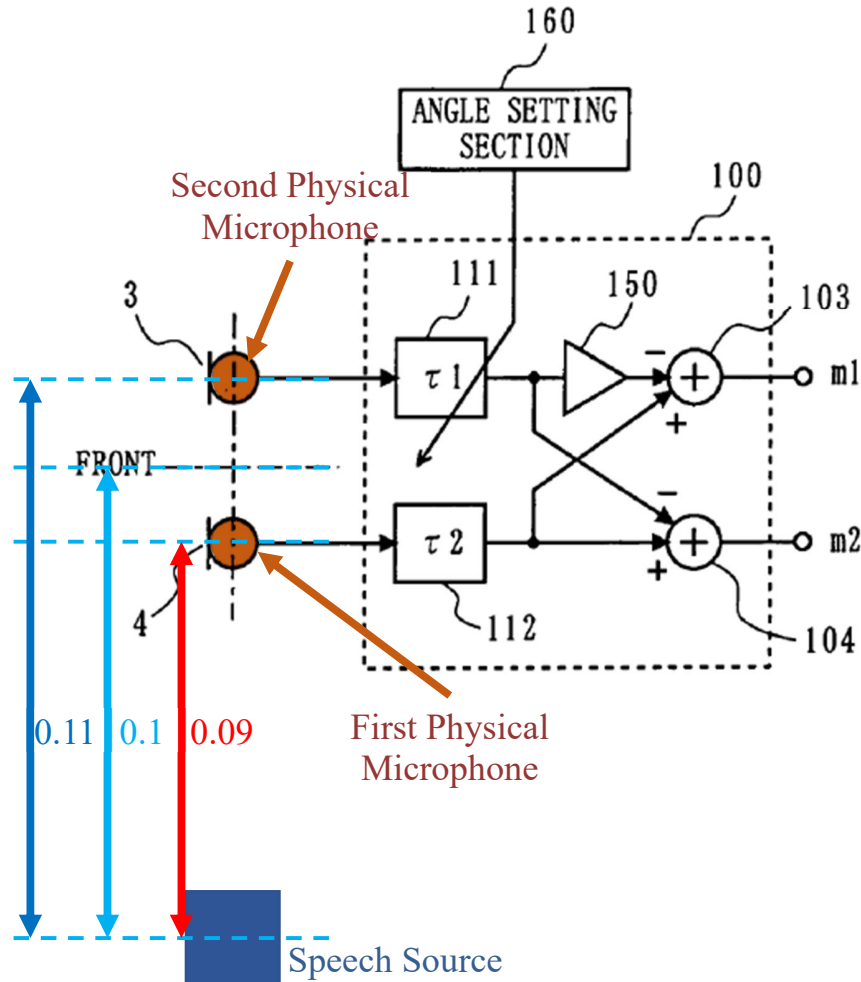
m2, Modified Kanamori Fig. 17a at 0.1m



In the above plot of m2’s linear response to speech, the single null is located at a distance of 0.1 m from where the source of the speech is expected to be and the midpoint of the array. *Supra* § VII.D.1.h ([1G]). In my opinion, a POSITA would understand that this single null is also located a distance from microphone unit 4 (first physical microphone) and microphone unit 3 (second physical microphone).

For example, if the microphones are spaced 2 cm apart, Ex. 1005, ¶ [0151], for a speech source expected at 0° and 0.1 m from the midpoint of the array, the null is located 0.09 m from microphone unit 4 and 0.11 m from microphone unit 3.

F I G . 1 7 A

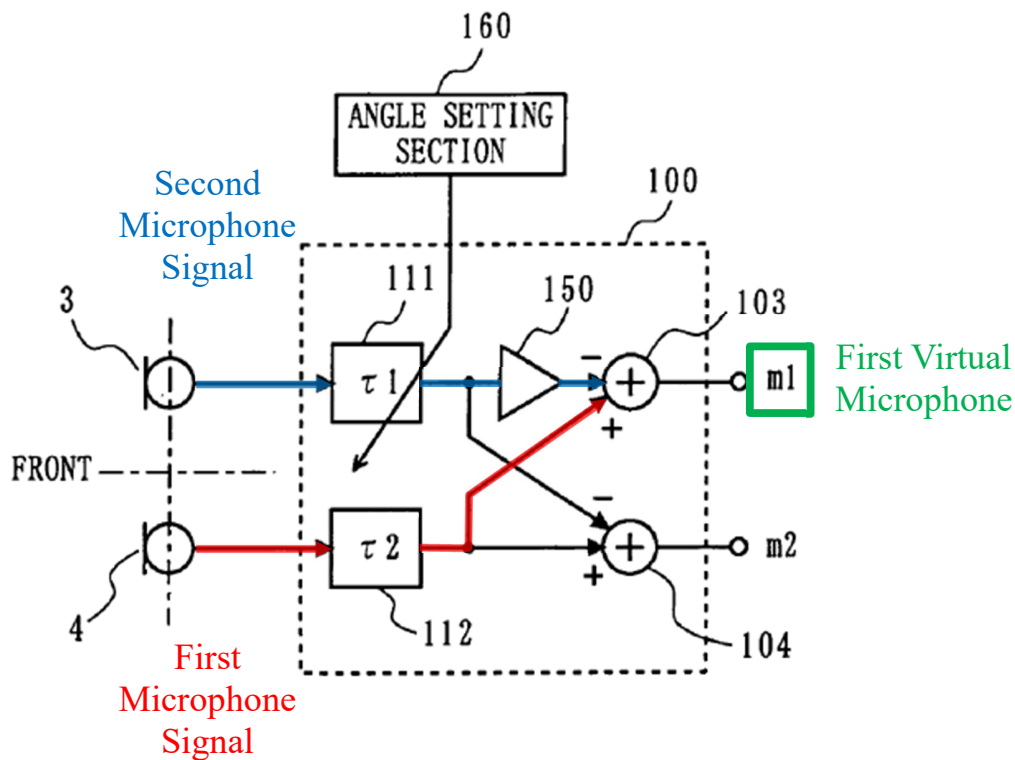


Ex. 1005, Fig. 17A (annotated).

- q. **Claim 21: “The microphone array of claim 1, wherein the first virtual microphone comprises the second microphone signal subtracted from a delayed version of the first microphone signal.”**

132. In my opinion, Kanamori discloses this additional feature. As explained for [1A], **main signal m1** constitutes a first virtual microphone that comprises the “**signal output from the third microphone unit 3**” (second microphone signal) subtracted from a delayed version of the “**signal output from the fourth microphone unit 4**” (first microphone signal). Ex. 1005, ¶¶ [0166], [0174]; *Supra* § VII.D.1.b.

FIG. 17A

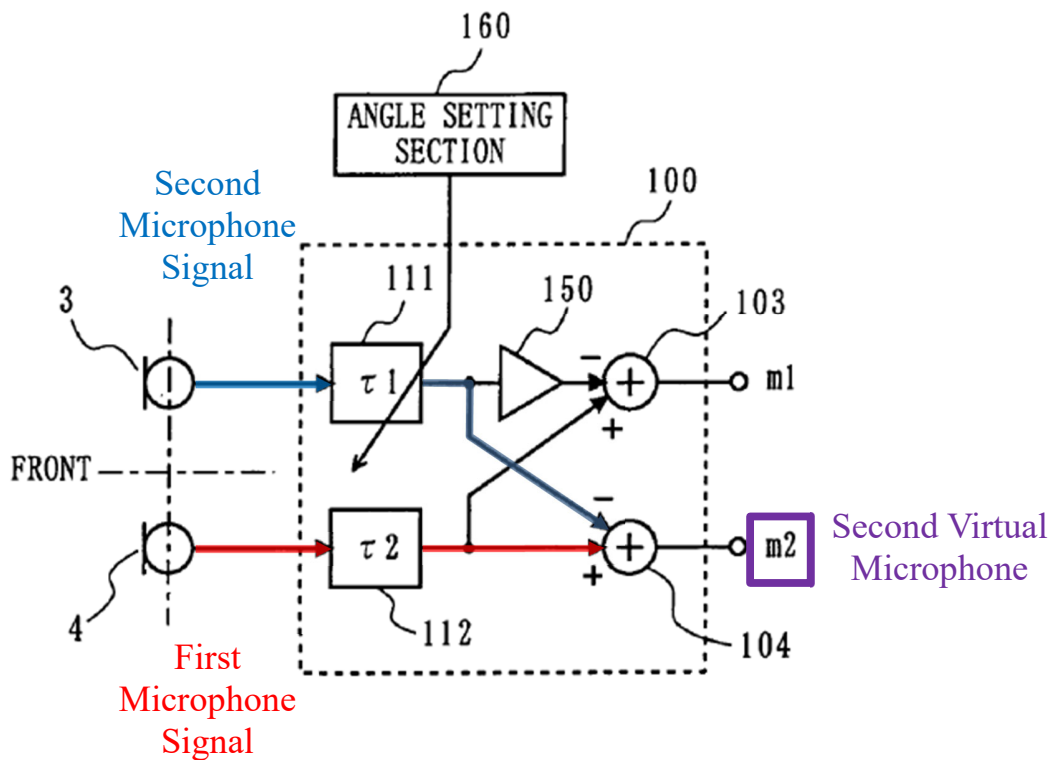


Ex. 1005, Fig. 17A (annotated).

- r. **Claim 22: “The microphone array of claim 21, wherein the second virtual microphone comprises a delayed version of the first microphone signal subtracted from the second microphone signal.”**

133. In my opinion, Kanamori renders obvious this additional feature. As explained for [1E], **noise reference signal m2** constitutes a second virtual microphone that comprises the “**signal output from the third microphone unit 3**” (second microphone signal) subtracted from a delayed version of the “**signal output from the fourth microphone unit 4**” (first microphone signal). Ex. 1005, ¶¶ [0166], [0174]; *Supra* § VII.D.1.f.

F I G. 1 7 A



Ex. 1005, Fig. 17A (annotated).

134. As explained for claim 15, in my opinion, it would have been obvious to switch the positive and negative inputs to signal subtracting section 104 such that the second virtual microphone comprises a delayed version of the first microphone signal subtracted from the second microphone signal. *Supra* § VII.D.2.k (claim 15).

3. Independent Claim 29

a. [29P]: “A device comprising:”

135. In my opinion, Kanamori discloses this feature. Kanamori discloses a “microphone device” (e.g., an audio recorder or video recorder) that includes two microphone units spaced a certain distance apart. Ex. 1005, ¶¶ [0077], [0151], [0201], Figs. 14-21, 23. A microphone device is a device.

b. [29A]: “a first physical microphone generating a first microphone signal;”

136. In my opinion, Kanamori discloses this feature for the reasons explained for [1D]. *Supra* § VII.D.1.e.

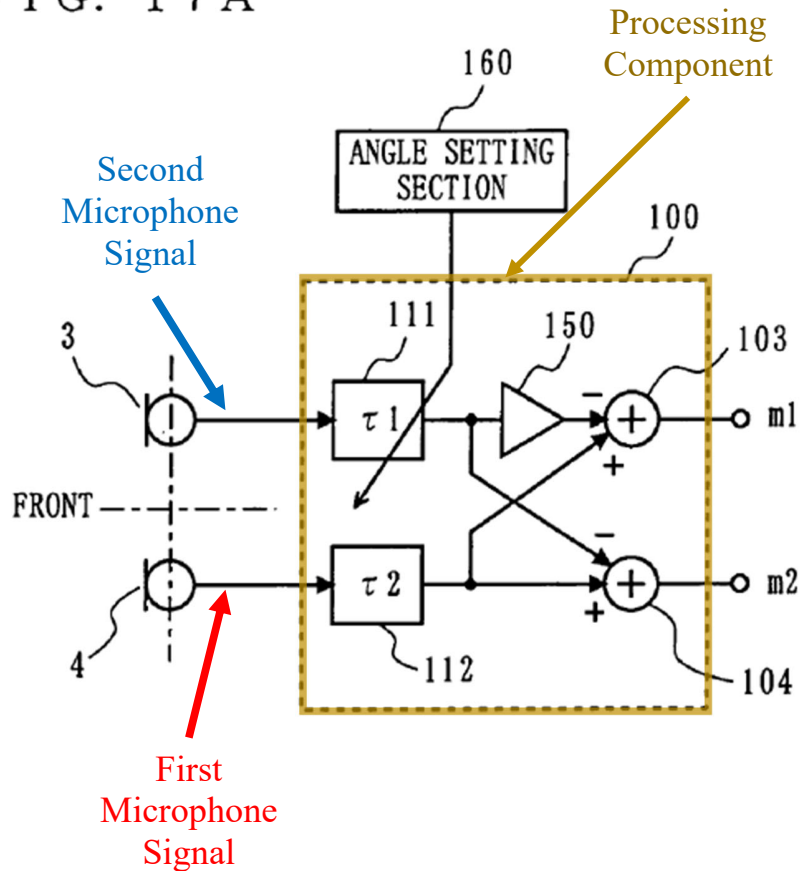
c. [29B]: “a second physical microphone generating a second microphone signal; and”

137. In my opinion, Kanamori discloses this feature for the reasons explained for [1D]. *Supra* § VII.D.1.e.

d. [29C]: “a processing component coupled to the first microphone signal and the second microphone signal,”

138. In my opinion, Kanamori discloses or renders obvious this feature. Kanamori discloses that a “**directivity combining section 100**” is coupled to the **first microphone signal** and **second microphone signal** and is employed to obtain a main signal m1 and a noise reference signal m2. Ex. 1005, ¶¶ [0152], [0173]. In my opinion, a POSITA would have understood that the **directivity combining section** is a processing component because Kanamori discloses that the “directivity combining process” allows for high flexibility in designing directivity and a variable characteristic using a digital process. Ex. 1005, ¶ [0151]. In my opinion, a POSITA would have understood that implementing a variable characteristic using a digital process requires a processing component to carry out the digital process. Fig. 5 in Kanamori further shows a finite impulse response or FIR (digital) filter, which a POSITA would understand is implemented on a processing component. Ex. 1005, ¶ [0108], Fig. 5.

FIG. 17A



Ex. 1005, Fig. 17A (annotated).

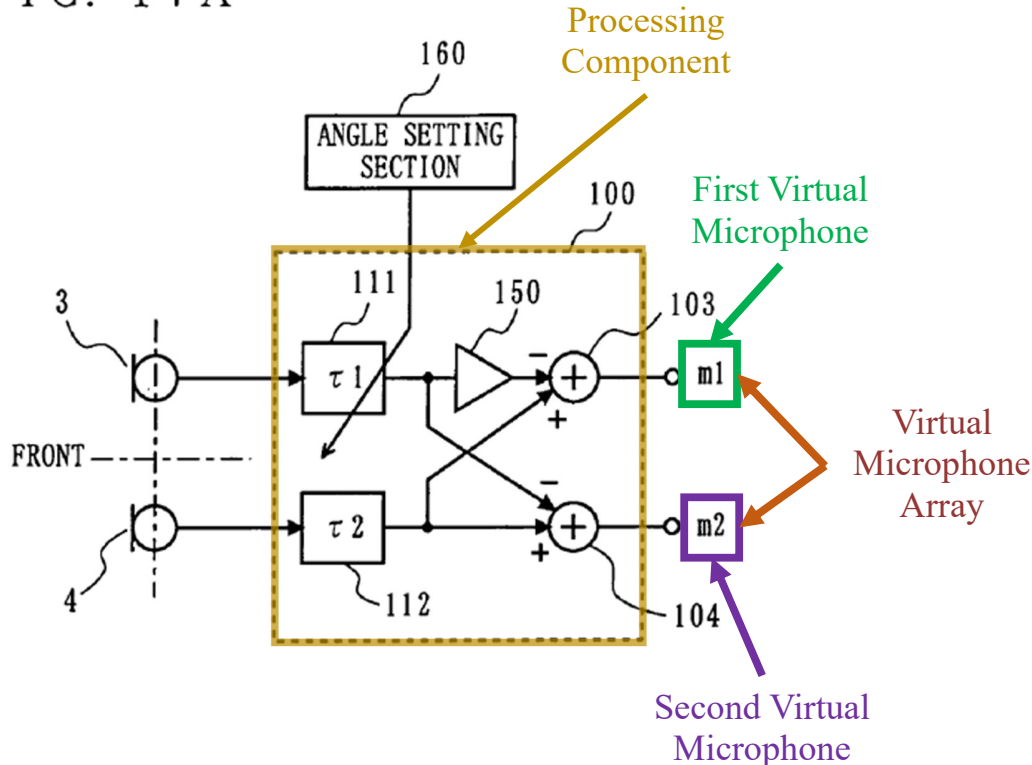
139. To the extent it is argued that it is not disclosed, in my opinion, it would have been obvious to have a processing component coupled to the first microphone signal and the second microphone signal. Kanamori, for example, discloses a signal delaying section 111 that can have a value anywhere from 0 to $2d/c$, which affects the orientation of the directivity pattern formed by m1 and m2. Ex. 1005, ¶¶ [0175]-[0176]. It would have been obvious to use a processor to make the delaying section 111 adjustable, making the microphone device better suited to

handling noise at a wider range of angles. Moreover, processors were conventionally used to perform signal processing involving microphone arrays, including using a processor to form virtual microphones. *E.g.*, Ex. 1013 (Elko '844), 11:43-63, 15:7-13.

e. **[29D]: “the processing component generating a virtual microphone array comprising a first virtual microphone and a second virtual microphone,”**

140. In my opinion, Kanamori discloses or renders obvious this feature. As explained for [29C], Kanamori's **directivity combining section 100** discloses or renders obvious a processing component. *Supra* § VII.D.3.d. This **directivity combining section** performs a “directivity combining process” where the device “obtain[s] directivity by combining signals output from” “a plurality of microphone units having non-directivity.” Ex. 1005, ¶ [0151]. This directivity combining process forms **main signal m1** (first virtual microphone) and **noise reference signal m2** (second virtual microphone). Ex. 1005, ¶¶ [0020], [0025], [0172]. **Main signal m1** and **noise reference signal m2** form a **virtual microphone array**.

FIG. 17A



Ex. 1005, Fig. 17A (annotated).

- f. [29E]: “the first virtual microphone having a first linear response to speech and a first linear response to noise,”

141. In my opinion, Kanamori discloses this feature for the reasons explained for [1B]. *Supra* § VII.D.1.c.

- g. [29F]: “the first linear response to speech being substantially similar across a plurality of frequencies for a speech source at an angle relative to an axis of the microphone array,”

142. In my opinion, Kanamori discloses this feature for the reasons explained for [1C]. *Supra* § VII.D.1.d.

- h. [29G]: “the second virtual microphone having a second linear response to speech and a second linear response to noise,”**

143. In my opinion, Kanamori discloses this feature for the reasons explained for [1F]. *Supra* § VII.D.1.g.

- i. [29H]: “the second linear response to noise being substantially similar to the first linear response to noise, and the second linear response to speech being substantially dissimilar to the first linear response to speech,”**

144. In my opinion, Kanamori and McCowan render obvious this feature for the reasons explained for [1G]. *Supra* § VII.D.1.h.

- j. [29I]: “wherein at least one of the first linear response to noise and the second linear response to noise is non-zero in a direction towards a source of noise”**

145. In my opinion, Kanamori discloses this feature for the reasons explained for [1H]. *Supra* § VII.D.1.i.

- k. [29J]: “wherein the first virtual microphone comprises the second microphone signal subtracted from a delayed version of the first microphone signal,**

146. In my opinion, Kanamori discloses this feature for the reasons explained in claim 21. *Supra* § VII.D.2.q.

- l. [29K]: “wherein the second virtual microphone comprises a delayed version of the first microphone signal subtracted from the second microphone signal.”**

147. In my opinion, Kanamori renders obvious this feature for the reasons explained in claim 22. *Supra* § VII.D.2.r.

4. Dependent Claims 30-40

- a. Claim 30: “The device of claim 29, wherein the first linear response to speech is devoid of a null, wherein the speech is human speech.”**

148. In my opinion, Kanamori discloses this additional feature for the reasons explained in claim 3. *Supra* § VII.D.2.b.

- b. Claim 31: “The device of claim 30, wherein the second linear response to speech of the second virtual microphone includes a single null oriented in a direction toward a source of the speech.”**

149. In my opinion, Kanamori and McCowan render obvious this additional feature for the reasons explained in claim 4. *Supra* § VII.D.2.c.i.

- c. Claim 32: “The device of claim 31, wherein the single null is a region of the second linear response to speech having a measured response level that is lower than the measured response level of any other region of the second linear response to speech.”**

150. In my opinion, Kanamori and McCowan render obvious this additional feature for the reasons explained in claim 5. *Supra* § VII.D.2.c.ii.

- d. **Claim 33: “The device of claim 31, wherein the second linear response to speech includes a primary lobe oriented in a direction away from the source of the speech.”**

151. In my opinion, Kanamori and McCowan render obvious this additional feature for the reasons explained in claim 6. *Supra* § VII.D.2.c.iii.

- e. **Claim 34: “The device of claim 33, wherein the primary lobe is a region of the second linear response to speech having a measured response level that is greater than the measured response level of any other region of the second linear response to speech.”**

152. In my opinion, Kanamori and McCowan render obvious this additional feature for the reasons explained in claim 7. *Supra* § VII.D.2.c.iv.

- f. **Claim 35: “The device of claim 31, wherein the first physical microphone and the second physical microphone are positioned along an axis and separated by a first distance.”**

153. In my opinion, Kanamori discloses this additional feature for the reasons explained in claim 8. *Supra* § VII.D.2.d.

- g. **Claim 36: “The device of claim 35, wherein a midpoint of the axis is a second distance from the speech source that generates the speech, wherein the speech source is located in a direction defined by an angle relative to the midpoint.”**

154. In my opinion, Kanamori discloses this additional feature for the reasons explained in claim 9. *Supra* § VII.D.2.e.

- h. Claim 37: “The device of claim 36, wherein one or more of the first microphone signal and the second microphone signal is delayed.”**

155. In my opinion, Kanamori discloses this additional feature for the reasons explained in claim 11. *Supra* § VII.D.2.g.

- i. Claim 38: “The device of claim 37, wherein the delay is raised to a power that is proportional to a time difference between arrival of the speech at the first virtual microphone and arrival of the speech at the second virtual microphone.”**

156. In my opinion, Kanamori discloses, or Kanamori and McCowan render obvious, this additional feature for the reasons explained in claim 12. *Supra* § VII.D.2.h.

- j. Claim 39: “The device of claim 38, wherein the power is proportional to a sampling frequency multiplied by a quantity equal to a third distance subtracted from a fourth distance, the third distance being between the first physical microphone and the speech source and the fourth distance being between the second physical microphone and the speech source.”**

157. In my opinion, Kanamori discloses or Kanamori and McCowan render obvious this additional feature for the reasons explained in claim 18. *Supra* § VII.D.2.n.

k. Claim 40: “The device of claim 36, wherein one or more of the first microphone signal and the second microphone signal is multiplied by a gain factor.”

158. In my opinion, Kanamori discloses that the second microphone signal is multiplied by a gain factor. Kanamori discloses that “signal amplifying section 150” in Figure 17A “performs constant multiplication” of the “signal output from the third microphone unit 3” (the second microphone signal) after it has been delayed at signal delaying section 111. Ex. 1005, ¶¶ [0166], [0174]. The gain factor is “set at approximately 0.85.” Ex. 1005, ¶¶ [0169], [0174].

* * *

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

160. I declare that all statements made herein of my knowledge are true and that all statements made on information and belief are believed to be true; and further, that these statements were made with knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S.C. § 1001.

161. I declare under penalty of perjury under the laws of the United States of America that the foregoing is true and correct.

Executed on this eighth day of June 2022.

By: 
Jeffrey S. Vipperman