

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

AMAZON.COM, INC.,
Petitioner,

v.

JAWBONE INNOVATIONS, LLC,
Patent Owner.

IPR2023-00253
U.S. Patent No. 8,019,091

DECLARATION OF RICHARD M. STERN, Ph.D.
IN SUPPORT OF PETITION FOR *INTER PARTES* REVIEW
OF CLAIMS 1-20 OF U.S. PATENT NO. 8,019,091

TABLE OF CONTENTS

I.	BACKGROUND -----	1
A.	Experience and Qualifications-----	1
B.	Materials Considered -----	3
II.	APPLICABLE LEGAL STANDARDS -----	4
A.	Claim Construction -----	5
B.	Obviousness -----	6
III.	PERSON OF ORDINARY SKILL IN THE ART -----	10
IV.	TECHNOLOGY BACKGROUND -----	12
A.	Using Adaptive Methods to Generate Transfer Functions and Remove Noise -----	12
B.	Tissue Vibration Sensors for Speech Detection Were Known -----	14
V.	THE '091 PATENT -----	15
A.	Summary of the '091 Patent -----	15
B.	The Priority Date of the '091 Patent -----	17
VI.	CLAIMS 1-20 OF THE '091 PATENT WOULD HAVE BEEN OBVIOUS -----	18
A.	Claims 1-8, 11-14, 16, and 18-20 Would Have Been Obvious in View of Ikeda and Puthuff -----	18
1.	Overview of Ikeda -----	18
2.	Overview of Puthuff -----	23
3.	Claim 11 -----	24
a.	11: Preamble -----	24

b.	11[a]: Two-Microphone Receiver -----	25
c.	11[b]: Processor that Generates First and Second Transfer Functions -----	27
i.	11[b][1]: Generating a First Transfer Function When Voicing Activity is Absent -----	30
ii.	11[b][2]: Generating a Second Transfer Function When Voicing Activity is Present -----	37
d.	11[c]: Noise Removal -----	40
e.	11[d]: Vibration Sensor -----	42
i.	Motivation to Combine Ikeda and Puthuff -----	44
4.	Claim 12 -----	49
5.	Claim 13 -----	49
6.	Claim 14 -----	50
7.	Claim 16 -----	50
8.	Claim 18 -----	51
a.	18: Preamble -----	51
b.	18[a]: Processor Coupled Among a Receiver and Sensor -----	51
c.	18[b]: Voicing Sensor -----	52
d.	18[c]: Processor That Generates Transfer Functions -----	52
i.	18[c][1]: First Transfer Function -----	52
ii.	18[c][2]: Second Transfer Function -----	53

e.	18[d]: Noise Removal-----	53
9.	Claim 19-----	53
10.	Claim 20-----	55
11.	Claim 1 -----	55
a.	1: Preamble-----	55
b.	1[a]: Receiving Acoustic Signals-----	56
c.	1[b]: Sensing Vibrations-----	56
d.	1[c]: Generating a VAD Signal-----	56
e.	1[d]: Generating at Least Two Transfer Functions-----	58
f.	1[e]: Removing Noise-----	60
12.	Claim 2 -----	60
a.	2[a]: Generating One Transfer Function-----	60
b.	2[b]: Removing Noise-----	61
13.	Claim 3 -----	61
14.	Claim 4 -----	62
15.	Claim 5 -----	63
16.	Claim 6 -----	64
17.	Claim 7 -----	64
18.	Claim 8 -----	64
B.	Claims 1-8, 11-14, 16, and 18-20 Would Have Been Obvious in View of Ikeda and Alcivar-----	65
1.	Claim 11 -----	65
a.	11: Preamble -----	65

b.	11[a]: Receiver-----	65
c.	11[b]: Processor that Generates Transfer Functions-----	65
i.	11[b][1]: First Transfer Function -----	66
ii.	11[b][2]: Second Transfer Function -----	66
d.	11[c]: Noise Removal-----	66
e.	11[d]: Vibration Sensor-----	67
i.	Motivation to Combine Ikeda and Alcivar-----	68
2.	Claim 12-----	71
3.	Claim 13-----	72
4.	Claim 14-----	72
5.	Claim 16-----	73
6.	Claim 18-----	73
a.	18: Preamble -----	73
b.	18[a]: Processor Coupled Among a Receiver and Sensor -----	73
c.	18[b]: Voicing Sensor-----	74
d.	18[c]: Processor That Generates Transfer Functions-----	74
i.	18[c][1]: First Transfer Function-----	74
ii.	18[c][2]: Second Transfer Function -----	75
e.	18[d]: Noise Removal-----	75
7.	Claim 19-----	75
8.	Claim 20-----	76

9.	Claim 1 -----	76
a.	1: Preamble -----	76
b.	1[a]: Receiving Acoustic Signals -----	77
c.	1[b]: Sensing Vibrations -----	77
d.	1[c]: Generating a VAD Signal -----	77
e.	1[d]: Generating at Least Two Transfer Functions -----	78
f.	1[e]: Removing Noise -----	79
10.	Claim 2 -----	79
11.	Claim 3 -----	80
12.	Claim 4 -----	81
13.	Claim 5 -----	81
14.	Claim 6 -----	81
15.	Claim 7 -----	82
16.	Claim 8 -----	82
C.	Claims 3, 9, and 15 Would Have Been Obvious in View of Ikeda and Either of Puthuff or Alcivar Further in View of Hussain -----	82
1.	Claim 3 -----	84
2.	Claims 9 and 15 -----	86
a.	9[a] and 15[a]: Dividing Acoustic Data into Sub-bands -----	86
b.	9[b] and 15[b]: Generating a Sub-band Transfer Function -----	87
c.	9[c] and 15[c]: Removing Acoustic Noise -----	88

d.	9[d] and 15[d]: Combining Sub-band Signals -----	88
3.	Motivation to Combine -----	89
D.	Claims 10 and 17 Would Have Been Obvious in View of Ikeda, Puthuff or Alcivar, and Sasaki -----	91
1.	Claims 10 and 17 -----	91
2.	Motivation to Combine -----	92
E.	It Would Have Been Obvious to Generate Transfer Functions Using a Processor in View of the References Discussed Above and Further in View of Bartlett -----	94
VII.	SECONDARY CONSIDERATIONS OF NONOBVIOUSNESS -----	97
VIII.	CONCLUSION -----	97

I, Richard M. Stern, Ph.D., do hereby declare:

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

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

I. BACKGROUND

A. Experience and Qualifications

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

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

5. I am a fellow of the Institute of Electrical and Electronics Engineers (IEEE), the Acoustical Society of America, and the International Speech Communication Association (ISCA). I was the ISCA 2008-2009 Distinguished

Lecturer, a recipient of the Allen Newell Award for Research Excellence in 1992, and I served as the General Chair of Interspeech 2006. Interspeech is the world's largest technical conference focused on speech processing and application.

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

7. I have actively worked on the theory and application of systems using microphone arrays over a period of decades (e.g., Stern et al., 2008; Stern and Menon, 2020), and my research group has developed several array-based algorithms to improve speech recognition accuracy in difficult acoustical environments (e.g., Seltzer et al., 2004; Stern et al., 2007; Kim et al., 2009; Moghimi and Stern, 2019). I have also actively worked on the theory and application of systems to accomplish voice activity detection for more than twenty years. Most recently, my students developed the system that provided best performance worldwide in the 2021 Fearless Steps Phase-03 Speech Activity Detection (SAD) Challenge, which was organized and run by the University of Texas, Dallas (Vuong et al., 2021). My relevant publications, including those cited above, are available on Carnegie Mellon's web site at <http://www.cs.cmu.edu/afs/cs/user/robust/www/papers.html>.

8. I understand a copy of my current curriculum vitae, which lists my publications for the last ten years, is being submitted as Exhibit 1014.

B. Materials Considered

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

Exhibit No.	Description
1001	U.S. Patent No. 8,019,091 (“the ’091 patent”)
1003	U.S. Patent No. 5,978,824 (“Ikeda”)
1004	International Patent Publication No. WO2000/021194 (“Puthuff”)
1005	U.S. Patent No. 3,746,789 (“Alcivar”)
1006	U.S. Patent No. 5,473,684 (“Bartlett”)
1007	Amir Hussain et al., <i>A New Metric for Selecting Sub-Band Processing in Adaptive Speech Enhancement Systems</i> , Proc. 5 th Eur. Conf. on Speech Comm’n and Tech. (Eurospeech ’97) 2611-14 (“Hussain”)
1008	U.S. Patent No. 5,471,538 (“Sasaki”)
1009	U.S. Provisional Patent Application No. 60/219,297
1010	Bernard Widrow et al., <i>Adaptive Noise Cancelling: Principles and Applications</i> , 63 Proc. IEEE 12 (1975) (“Widrow 1975”)
1011	Bernard Widrow et al., <i>Adaptive Signal Processing</i> (1985) (“Widrow 1985”)

Exhibit No.	Description
1012	U.S. Patent No. 6,349,197 (“Oestreich”)
1013	Excerpts from the ’091 patent’s file history

10. I have also relied on my education, training, and experience, and my knowledge of pertinent literature in the field of the ’091 patent.

II. APPLICABLE LEGAL STANDARDS

11. I have been asked to provide my opinion as to whether the claims of the ’091 patent would have been obvious to a person of ordinary skill in the art at the time of the alleged invention, in view of the prior art.

12. I am an electrical engineer by training and profession. The opinions I am expressing in this report involve the application of my training and technical knowledge and experience to the evaluation of certain prior art with respect to the ’091 patent.

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

A. Claim Construction

14. It is my understanding that in determining whether a patent claim is obvious in view of the prior art, the Patent Office construes the claim by giving the claim terms their plain and ordinary meaning, as they would have been understood by a person of ordinary skill in the art at the time of the invention in view of the intrinsic record (patent specification and file history). For the purposes of this review, and to the extent necessary, I have interpreted each claim term in accordance with its plain and ordinary meaning as it would have been understood by a person of ordinary skill in the art at the time of the invention, in view of the intrinsic record. I have been instructed that the time of the invention is July 19, 2000, which I understand to be the earliest claimed priority date of the '091 patent.

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

16. I understand that sources extrinsic to a patent and its prosecution history (such as dictionary definitions and technical publications) may also be used to help

interpret the claim language, but that such extrinsic sources cannot be used to contradict the unambiguous meaning of the claim language that is evident from the intrinsic evidence.

17. Unless expressly stated herein, I have applied the plain and ordinary meaning of the claim terms, which I understand is the meaning that a person of ordinary skill in the art would have given to terms in July 2000 based on a review of the intrinsic evidence.

B. Obviousness

18. It is my understanding that a claim is “obvious” if the claimed subject matter as a whole would have been obvious to a person of ordinary skill in the art at the time of the alleged invention. I understand that an obviousness analysis involves a number of considerations. I understand that the following factors must be evaluated to determine whether a claim would have been obvious: (i) the scope and content of the prior art; (ii) the differences, if any, between each claim of the '091 patent and the prior art; (iii) the level of ordinary skill in the art in July 2000; and (iv) additional considerations, if any, that indicate that the invention was obvious or not obvious. I understand that these “additional considerations” are often referred to as “secondary considerations” of non-obviousness or obviousness.

19. I also understand that the frame of reference when evaluating obviousness is what a hypothetical person of ordinary skill in the pertinent art would

have known in July 2000. I understand that the hypothetical person of ordinary skill is presumed to have knowledge of all pertinent prior art references.

20. I understand that a prior art reference may be a pertinent prior art reference (or “analogous art”) if it is in the same field of endeavor as the patent or if it is pertinent to the problem that the inventors were trying to solve. A reference is reasonably pertinent if it logically would have commended itself to an inventor’s attention in considering the problem at hand. If a reference relates to the same problem as the claimed invention, that supports use of the reference as prior art in an obviousness analysis. Here, all of the references relied on in my obviousness analysis below are in the same field of endeavor as the ’091 patent, e.g., signal capture and processing for speech or audio applications. The references are also pertinent to a particular problem the inventor was focused on, e.g., noise reduction.

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

- combining prior art elements according to known methods to yield predictable results;
- simple substitution of one known element for another to obtain predictable results;

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

22. I understand that “secondary considerations” must be considered as part of the obviousness analysis when present. I further understand that the secondary considerations may include: (1) a long-felt but unmet need in the prior art that was satisfied by the claimed invention; (2) the failure of others; (3) skepticism by experts;

(4) commercial success of a product covered by the patent; (5) unexpected results achieved by the claimed invention; (6) industry praise of the claimed invention; (7) deliberate copying of the invention; and (8) teaching away by others. I also understand that evidence of the independent and nearly simultaneous “invention” of the claimed subject matter by others is a secondary consideration supporting an obviousness determination and may support a conclusion that a claimed invention was within the knowledge of a person of ordinary skill as of July 2000. I am not aware of any evidence of secondary considerations that would suggest that the claims of the '091 patent would have been nonobvious in July 2000.

23. I understand that when assessing obviousness, using hindsight is impermissible; that is, what is known today or what was learned from the teachings of the patent should not be considered. The patent should not be used as a road map for selecting and combining items of prior art. Rather, obviousness must be considered from the perspective of a person of ordinary skill at the time the alleged invention was made – July 2000 in this case.

24. I also understand that an obviousness analysis must consider the invention as a whole, as opposed to just a part or element of the invention. I understand this “as a whole” assessment to require showing that one of ordinary skill in the art at the time of invention, confronted by the same problems as the inventor

and with no knowledge of the claimed invention, would have selected the elements from the prior art and combined them in the claimed manner.

25. It is my understanding that something is “inherent in,” and therefore taught by, the prior art, if it necessarily flows from the explicit disclosure of the prior art. I understand that the fact that a certain result or characteristic may be present in the prior art is not sufficient to establish inherency. However, if the result or characteristic is necessarily present based upon the explicit disclosure in the prior art, it is inherent in the prior art and is therefore disclosed.

III. PERSON OF ORDINARY SKILL IN THE ART

26. It is my understanding that when interpreting the claims of the '091 patent and evaluating whether a claim would have been obvious, I must do so based on the perspective of a person of ordinary skill in the art at the relevant priority date. I have been instructed to assume for the purposes of my opinions that the relevant priority date of the '091 patent is July 19, 2000.

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

must have the capability of understanding the scientific and engineering principles applicable to the pertinent art.

28. Based on my review of the specification and claims of the '091 patent, it is my opinion that a person of ordinary skill in the art would have had a minimum of a bachelor's degree in computer engineering, computer science, electrical engineering, mechanical engineering, or a similar field, and approximately three years of industry or academic experience in a field related to acoustics, speech recognition, speech detection, or signal processing. Work experience could substitute for formal education and additional formal education could substitute for work experience.

29. My conclusions below that the claims of the '091 patent would have been obvious would remain the same even if the priority date, field of endeavor, or level of ordinary skill were slightly different.

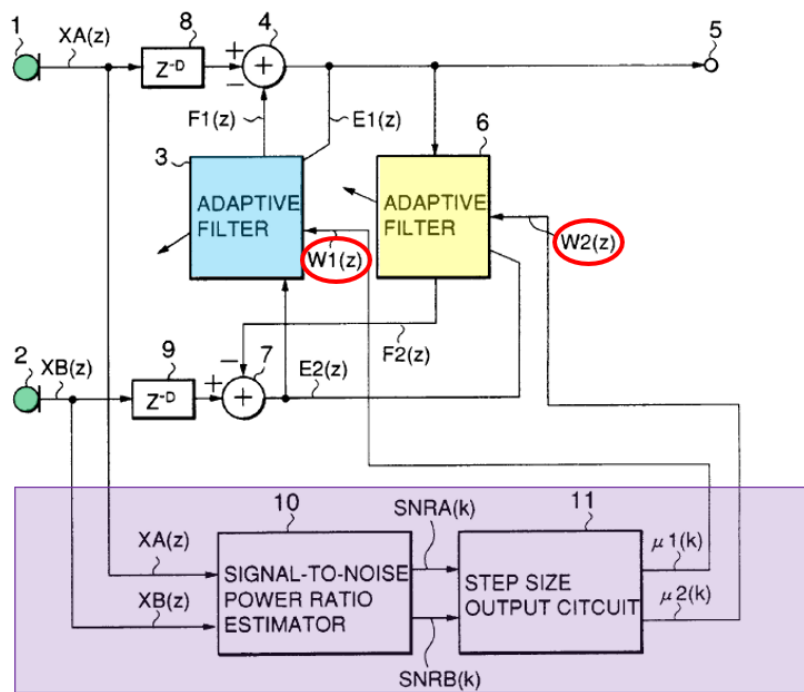
30. I meet the above definition of a person of ordinary skill in the art, and did so as of July 2000. Also, I have worked with persons of ordinary skill in the art through my professional and academic experiences, and I have an understanding of their skill level around July 2000.

IV. TECHNOLOGY BACKGROUND

A. Using Adaptive Methods to Generate Transfer Functions and Remove Noise

31. Two-microphone systems that adaptively remove noise using transfer functions have been used for decades. For example, in 1975, Widrow described a dual-microphone noise-reduction system that used an adaptive filter for generating a “transfer function” to remove noise from acoustic signals such as speech. (Ex. 1010 (Widrow 1975), 1695-1701, 1704-1705, Figs. 3-6, 18-19.) Widrow described similar systems and various adaptive methods in a 1985 textbook. (Ex. 1011 (Widrow 1985).) Widrow 1985 disclosed the general operation of adaptive systems (*id.*, ch. 1), a common adaptive algorithm (*id.*, ch. 6 (describing the “LMS Algorithm”)), and the use of adaptive algorithms to generate transfer functions which can be used to remove noise from acoustic signals (*id.*, ch. 12).

32. Ikeda, as discussed in more detail below, describes a noise canceler that receives acoustic signals from a signal and noise source and uses adaptive filters to generate transfer functions and remove noise from the desired signal. The system includes two microphones and adaptive filters for generating transfer functions. (Ex. 1003 (Ikeda), 4:49-53.) The system uses two adaptive filters (3 and 6) with respective first and second transfer functions $W1(z)$ and $W2(z)$ as shown by Ikeda’s Figure 1:



(*Id.*, Fig. 1, 5:17-23.)¹ Ikeda's first transfer function is calculated when the desired speech signal is absent or negligible. (*Id.*, 3:32-45.) In contrast, Ikeda's second transfer function is calculated when speech is present and noise is ideally absent. (*Id.*, 3:32-45.)

33. After generating the transfer functions, Ikeda's noise canceler uses them to produce an output "from which noise has been cancelled." (Ex. 1003 (Ikeda), 5:62-63; *see also id.* 5:17-62.)

¹ Figures herein have been annotated for clarity.

B. Tissue Vibration Sensors for Speech Detection Were Known.

34. Many non-acoustic methods of determining when speech is present or absent were known in the art. For example, determining periods when voicing activity is present or absent using a sensor that detects tissue vibrations was known for decades before the '091 patent's application was filed. For example, Alcivar, which issued in 1973, disclosed controlling a "voice-activated transmit switch (VOX) for high noise environment voice communication systems" using a "tissue-conduction microphone [] positioned in contact with the user's neck tissue in the vicinity of the larynx." (Ex. 1005 (Alcivar), Abstract.) Alcivar's "tissue conduction microphone" detects voice through "tissue vibrations due to the modulation of the vocal chords." (*Id.*, 4:65-5:54, Abstract.)

35. Voice activity detectors were widely used with microphones to produce a signal with reduced noise. Puthuff, which published in 2000, similarly disclosed a noise suppression system using a "motion transducer [80] comprising a vibration sensor" such as an accelerometer to "adaptively sense[] sound through bone conduction" and therefore detect when a user is speaking. (Ex. 1004 (Puthuff), 5:1-3, 4:13-21; *id.*, 6:6-10.) The vibration sensor is placed against a talker's head (e.g., in a handset or hearing aid) to detect "speech-induced vibrations." (*Id.*, 8:16-19; *id.*, 3:19-23, 2:20-3:5.)

V. THE '091 PATENT

A. Summary of the '091 Patent

36. The '091 patent relates to “systems and methods for detecting and processing a desired signal in the presence of acoustic noise” in devices such as a telephone. (Ex. 1001 (the '091 patent), 1:16-18; *id.*, 2:55-56.) Figure 2 of the '091 patent shows an embodiment with two microphones (102 and 103, both highlighted green) that receive acoustic signals from a signal source 100 (e.g., speech, highlighted yellow) and a noise source 101 (highlighted grey):

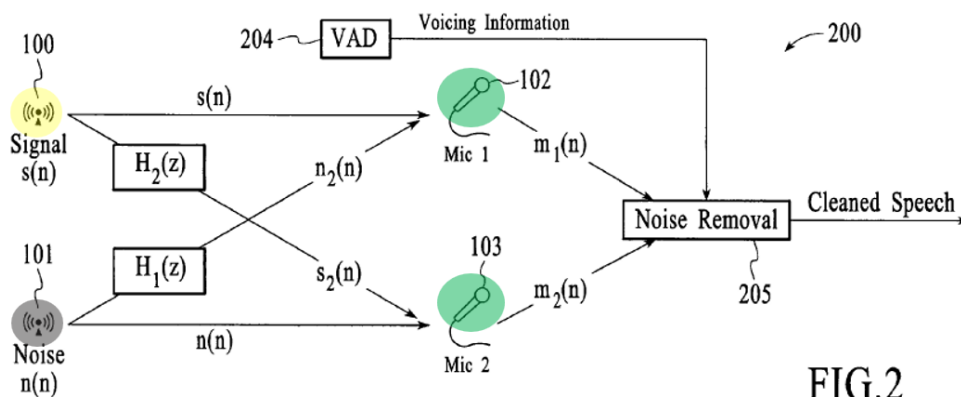


FIG.2

(*Id.*, Fig. 2, 3:20-50.) As was known in the art, a first transfer function, $H_1(z)$, exists between the noise source and Mic 1, and a second transfer function, $H_2(z)$, exists between the speech source and Mic 2. (*Id.*, Fig. 2, 3:53-55.) The transfer functions are approximated at the appropriate time periods, described below, using a VAD that “uses physiological information to determine when a speaker is speaking” through use of a sensor such as “an accelerometer [or] a skin surface microphone in physical

contact with skin of a user.” (*Id.*, 3:39-50.) The approximated transfer functions can be used to remove noise from the speech signal. (*Id.*, 5:28-31.)

37. As in the prior art, the denoising system approximates one transfer function during periods when speech is absent from the acoustic signals and approximates a second transfer function during periods when speech is present in the acoustic signals, as determined by the VAD. (Ex. 1001 (the '091 patent), 4:29-5:6, 7:50-56.) The '091 patent's specification provides the conventional mathematical equations associated with this system. For example, when the VAD indicates that voicing information is absent, the speech signal is assumed to be zero and $H_1(z)$ is approximated as:

$$H_1(z) = \frac{M_{1n}(z)}{M_{2n}(z)}$$

where subscript “n” indicates “that only noise is being received.” (*Id.*, 4:28-45; *id.*, 7:50-53, Eq. 2.) $M_1(z)$ and $M_2(z)$ are the frequency (“z”) domain representations of the acoustic signals received at signal microphone 102 and noise microphone 103, respectively. (*Id.*, 4:5-8, 4:29-51, 7:50-53.)

38. Similarly, when the VAD indicates that voicing information is present, noise is assumed to be negligible or zero and $H_2(z)$ is approximated as:

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

where subscript “s” indicates that only speech is occurring. (Ex. 1001 (the ’091 patent), 4:53-5:5, 7:53-56.)

39. The approximated transfer functions $H_1(z)$ and $H_2(z)$ are then used to “remove the noise from the [speech] signal” resulting in a denoised output signal $S(z)$:

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

(Ex. 1001 (the ’091 patent), 5:13-36.)

40. The system determines whether voice is present or absent using a “voice activity detector” (“VAD”). According to the ’091 patent’s specification, prior art systems “generally” used an acoustic microphone-based VAD, which have “significantly limited” capability in high noise environments or where background noise varies quickly. (Ex. 1001 (the ’091 patent), 1:22-58.) The ’091 patent therefore uses a different type of known VAD, i.e., a VAD that “uses physiological information to determine when a speaker is speaking” through the use of a vibration sensor such as “an accelerometer [or] a skin surface microphone in physical contact with the skin of a user.” (*Id.*, 3:39-50.)

B. The Priority Date of the ’091 Patent

41. I understand that the ’091 patent issued on September 13, 2011, from U.S. Patent Application 10/667,207 (“the ’207 application”), filed September 18,

2003. I also understand that the '207 application claims priority to U.S. Provisional Application No. 60/219,297 (“the '297 application”), which was filed on July 19, 2000, and is a continuation-in-part of U.S. Patent Application No. 09/905,361, filed July 12, 2001. (Ex. 1001 (the '091 patent).) I understand that the '091 patent’s earliest possible priority date is therefore July 19, 2000.

VI. CLAIMS 1-20 OF THE '091 PATENT WOULD HAVE BEEN OBVIOUS.

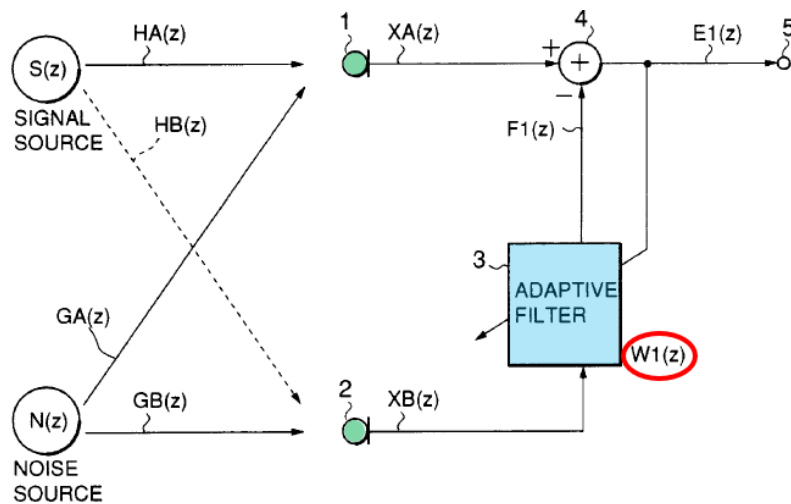
A. Claims 1-8, 11-14, 16, and 18-20 Would Have Been Obvious in View of Ikeda and Puthuff.

42. For at least the reasons I discuss below, claims 1-8, 11-14, 16, and 18-20 would have been obvious in view of Ikeda and Puthuff.

1. Overview of Ikeda

43. Ikeda disclosed a two-microphone noise canceling system that used an adaptive filter. (Ex. 1003 (Ikeda), 1:4-8 (disclosing a “noise canceler for canceling, by use of an adaptive filter, a background noise signal introduced into a speech signal via a microphone, handset or the like”).)

44. Ikeda first describes and shows a “conventional” two-microphone noise cancellation system that implements a “conventional noise cancellation principle” using a primary microphone 1, reference microphone 2, and adaptive filter 3, as shown in Figure 5:



(Ex. 1003 (Ikeda), Fig. 5 (describing “Prior Art”); *id.*, 1:42-43, 3:32-36.) In the conventional noise canceler described in Ikeda, “[a] desired signal $S(z)$ spoken by a speaker (signal source) is input to the main signal microphone 1 adjoining the speaker’s mouth by way of a path having an acoustic transfer characteristic $HA(z)$ ” and “noise $N(z)$ issuing from a noise source is input to the main signal microphone 1 via a path having an acoustic transfer characteristic $GA(z)$.” (*Id.*, 1:46-56.) Similarly, “the noise $N(z)$ is input to a reference signal microphone 2 remote from the speaker by way of a path having an acoustic transfer characteristic $GB(z)$.” (*Id.*, 1:56-59, Fig. 5.) As indicated by the dotted line in Ikeda’s Figure 5, the reference signal microphone 2 can sometimes “contain[] the desired signal component $S(z)$ ” via “acoustic transfer characteristic $HB(z)$.” (*Id.*, 2:46-49, Fig. 5.)

45. The adaptive filter 3 of Ikeda’s conventional noise canceler is used to remove noise that reaches main signal microphone 1. (Ex. 1003 (Ikeda), 2:3-28.)

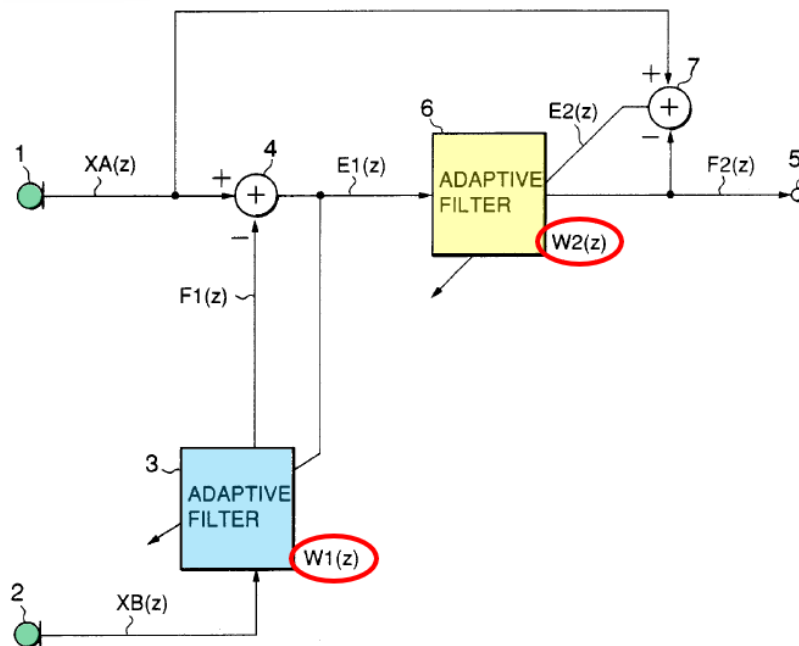
This adaptive filter has a transfer function (also referred to as a “transfer characteristic”) $W1(z)$. (*Id.*, 2:3-5.) Ikeda explains that, in this conventional system, the resulting first microphone signal $XA(z)$ and second microphone signal $XB(z)$ are expressed by the following equations:

$$XA(z) = GA(z)N(z) + HA(z)S(z)$$

$$XB(z) = GB(z)N(z) + HB(z)S(z)$$

(*Id.*, 5:33-38, Eqs. 10-11.)

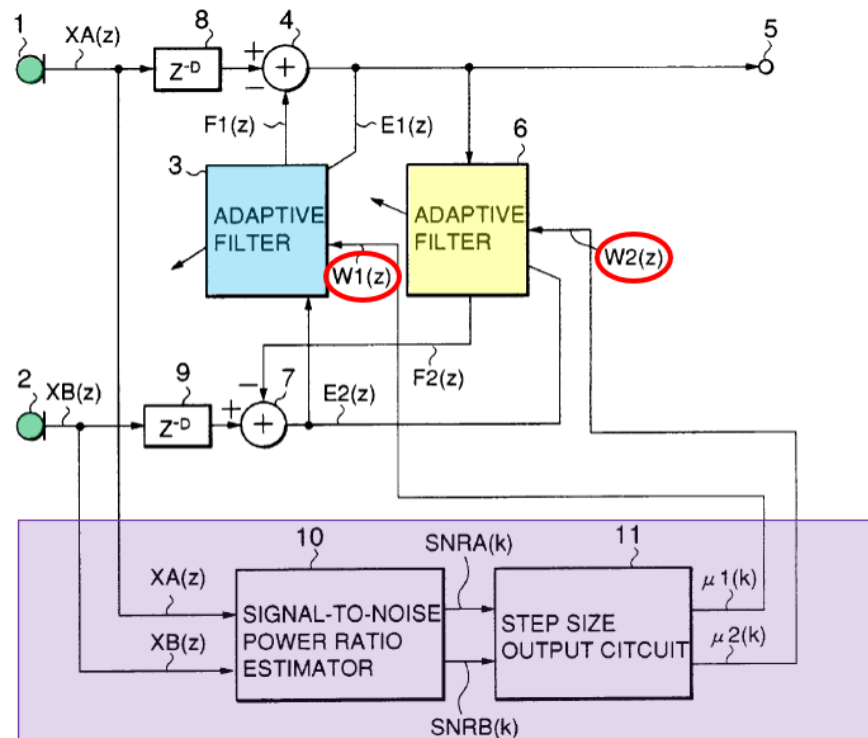
46. Ikeda then discloses that it was known in the art to add a second adaptive filter 6 having a second transfer function $W2(z)$ to correct for distortion that occurs in two-microphone systems, as shown in Figure 6:



(Ex. 1003 (Ikeda), Fig. 6 (describing “Prior Art”).)

47. Ikeda discloses that the first transfer function, $W1(z)$, is generated when voice is absent while the second transfer function, $W2(z)$, is generated when voice is present. (Ex. 1003 (Ikeda), 3:32-45, 2:60-3:10.) As a result, it is “necessary to detect” when the voice signal is present or absent. (*Id.*, 3:32-45.) Ikeda teaches that the system’s ability to cancel noise and correct distortions are not sufficient “unless adequate learning sections are indicated to each adaptive filter,” i.e., that effective noise removal depends on adequate voice detection. (*Id.*, 3:50-54.)

48. Ikeda then describes an improved system, shown in its Figure 1, that still uses two microphones (1, 2), two adaptive filters (3, 6), and two transfer functions ($W1(z)$, $W2(z)$), and further includes a voice activity detector (shown in purple):



(Ex. 1003 (Ikeda), Fig. 1, 4:49-53, 5:17-23.) Ikeda's voice activity detector determines whether voice is present or absent by comparing the signals received at microphones 1 and 2 (green) and controls the filters accordingly. (*Id.* 3:32-45, 5:64-6:5, 2:32-36, 6:16-60.)

49. Ikeda also discloses the mathematical equations upon which its system is based, as I describe in detail below. Ikeda discloses that the first transfer function $W1(z)$, which is generated by adaptive filter 3, can be approximated as

$$W1(z) = \frac{GA(z)}{GB(z)}$$

and the second transfer function $W2(z)$, which is generated by adaptive filter 6, can be approximated as

$$W2(z) = \frac{HB(z)}{HA(z)}$$

(Ex. 1003 (Ikeda), 5:53-56, Eqs. 14, 15.)

50. Ikeda explains that combinations of the transfer functions are used to reduce noise in the system's output signal, $E1(z)$. Specifically, Ikeda teaches that the transfer functions are used to produce an output, $E1(z)$, that "does not contain any noise signal" and can be expressed as

$$E1(z) = XA(z) - W1(z)E2(z)$$

while the noise output $E2(z)$ is expressed as

$$E2(z) = XB(z) - W2(z)E1(z)$$

(Ex. 1003 (Ikeda), 2:43-44; *id.*, 5:17-30, 5:51-63, Eqs. 8-9.) The mathematical representation for $E1(z)$, above, can be rewritten as

$$E1(z) = \frac{XA(z) - XB(z)W1(z)}{1 - W2(z)W1(z)}$$

using simple mathematical steps that I describe in detail below. (¶¶59-83 (claim limitations 11[b]-11[c]), below.)

51. In summary, two-microphone noise reduction systems that use two transfer functions and a voice activity detection algorithm to determine periods of voicing activity were known in the art as well as the mathematical steps for using such transfer functions to reduce noise in the output signal of the system.

2. Overview of Puthuff

52. Puthuff disclosed a system for “suppress[ing] the deleterious effects of high acoustic noise environment” by “adaptively sens[ing] sound through bone conduction.” (Ex. 1004 (Puthuff), 4:13-21.) Puthuff disclosed detecting voice using a microphone 70 and “a motion transducer [80] comprising a vibration sensor” such as an accelerometer. (*Id.*, 5:1-3.) By placing a vibration sensor “against the skin of the talker’s skull” or “at the talker’s throat,” Puthuff’s system detects “speech-induced vibrations” that propagate “to any part of a person’s body supported by bone structure.” (*Id.*, 8:16-19; *id.*, 3:19-23, 2:20-3:5.)

53. Puthuff's system can be controlled using "[c]haracteristics of the loudness" of bone-conducted voice signals. (Ex. 1004 (Puthuff), 6:14-20.) Puthuff's system is "particular[ly] suit[ed] to voice-activated applications, as the bone conduction expedient improves system immunity to false triggering due to noisy environments." (*Id.*, 6:6-8.) One example of such an application is "turning off a radio transmitter in the absence of the user's voice." (*Id.*, 6:8-10.)

3. Claim 11

54. Claim 11 recites a noise removal system with a "receiver" that contains "at least two acoustic microphones," a voicing sensor that detects "human tissue vibration," a processor that generates two transfer functions, and removing noise from an acoustic signal using the two or more transfer functions.²

a. 11: Preamble

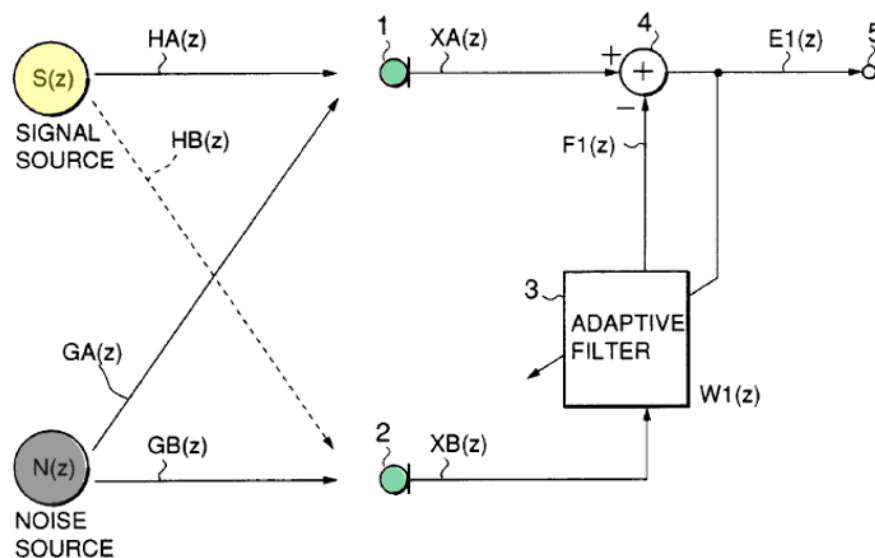
55. The preamble of claim 11 recites "[a] system for removing acoustic noise from the acoustic signals." To the extent the preamble is limiting, Ikeda discloses such a system because it discloses a "noise canceler for canceling, by use of an adaptive filter, a background noise signal introduced into a speech signal input via a microphone, a handset or the like." (Ex. 1003 (Ikeda), 1:4-19.)

² The voicing sensor limitation is addressed last for clarity.

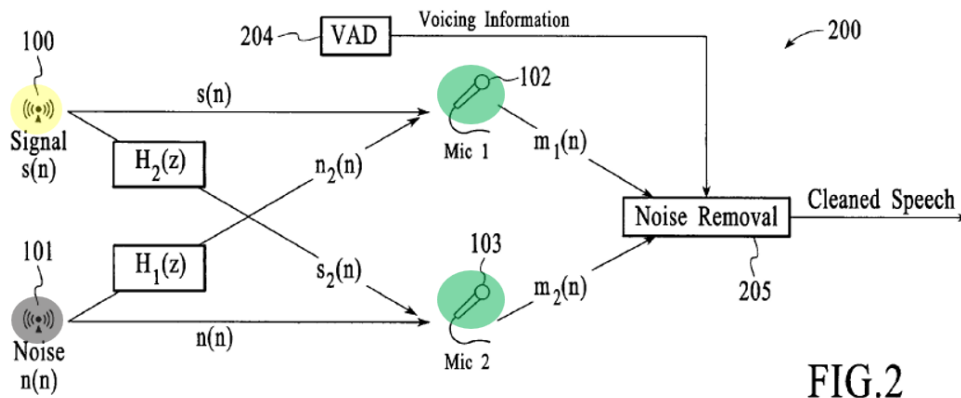
b. 11[a]: Two-Microphone Receiver

56. Claim 11 recites “a receiver that receives at least two acoustic signals via at least two acoustic microphones positioned in a plurality of locations.”

57. Ikeda discloses a noise reduction system comprising a receiver that receives at least two acoustic signals at two or more microphones placed at a plurality of locations. Ikeda discloses a first acoustic signal (yellow) received by a first acoustic microphone 1 (green) at a first location: “[a] desired signal $S(z)$ spoken by a speaker (signal source) is input to the main signal microphone 1 adjoining the speaker’s mouth.” (Ex. 1003 (Ikeda), 1:46-56.) Additionally, a second acoustic signal such as noise (grey) is received by a second acoustic microphone 2 (green) at a second location: “the noise $N(z)$ is input to a reference signal microphone 2 remote from the speaker.” (*Id.*) Each acoustic microphone also receives the other acoustic signal. (*Id.*, Fig. 5.) This is shown in Figure 5:

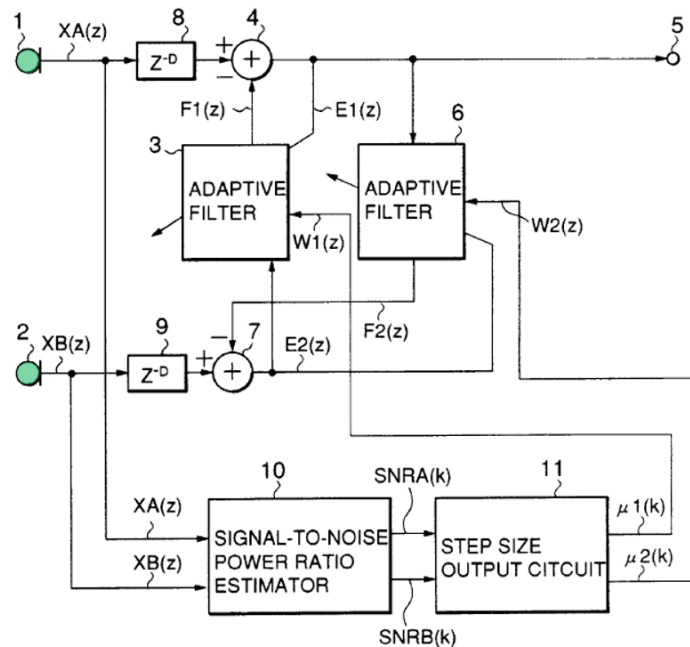


(*Id.*, 1:46-56, Fig. 5.) Ikeda's Figure 5 discloses the same signals (yellow and grey) and microphones (green) as those in the '091 patent:



(Ex. 1001 (the '091 patent), Fig. 2.)

58. Ikeda also describes an improved noise canceler that includes a second adaptive filter. (Ex. 1003 (Ikeda), 4:49-56.) When discussing the improved system, Ikeda explains that the dual-microphone structure remains the same as that of the noise canceler in Figure 5. (*Id.*, 5:31-38.) Ikeda states that the system “includes a first microphone 1 [e.g., a first acoustic microphone] for a main signal [e.g., a first acoustic signal] [and] a second microphone 2 [e.g., a second acoustic microphone] for a reference signal [e.g., a second acoustic signal].” (*Id.*, 4:49-53, Fig. 1.) Ikeda's improved two-microphone system is shown in Figure 1:



(*Id.*, Fig. 1, 4:49-53, 5:31-38.) Ikeda further discloses that the improved noise canceler can be used in a handset. (*Id.*, 1:4-19.) Thus, Ikeda discloses this limitation.

c. 11[b]: Processor that Generates First and Second Transfer Functions

59. Claim 11 next recites “a processor coupled among the receiver and the at least one sensor that generates a plurality of transfer functions, wherein the plurality of transfer functions includes a first transfer function . . . [and] a second transfer function.” Ikeda and/or Puthuff render this limitation obvious. As discussed below, Ikeda discloses generating a plurality of transfer functions. (¶¶65-79 (claim elements 11[b][1]-[2]), below.)

60. A person of ordinary skill in the art would have understood Ikeda’s transfer functions to be generated by a processor coupled to the receiver and the

VAD because that is how systems such as Ikeda's were commonly implemented. For example, Widrow 1985, a textbook cited in Ikeda as disclosing examples of prior art noise cancelers, describes the use of processors that implement adaptive algorithms and generate transfer functions. (Ex. 1011 (Widrow 1985), 4 ("Recent progress in microcircuit design and production has resulted in very compact, economical, and reliable signal processors" resulting in "a very fast-growing field of application for all types of digital signal processing, including adaptive processing. Current applications for adaptive systems are in such fields as communication"); Ex. 1003 (Ikeda), 1:9-19 (citing Widrow 1985); *see also* Ex. 1011 (Widrow 1985), 8, 11, 13, 16, 93, 97, 339, 368, Figs. 1.2(a)-(d), 1.5(a)-(d).)

61. Alternatively, it would have been obvious to a person of ordinary skill in the art to use a processor coupled to the receiver and VAD to process the signals and generate transfer functions for several reasons. For example, it was well known in the art to use processors for that purpose. In my own work in robust speech recognition I used signal processing to remove the effects of unknown additive noise and unknown linear filtering to improve the accuracy of speech recognition systems in difficult environments well before the earliest possible priority date of the '091 patent. One example of such signal processing is shown in my 1996 book chapter, "Signal Processing for Robust Speech Recognition." (Richard M. Stern et al., *Signal Processing for Robust Speech Recognition*, in *Speech Recognition* 351-78 (Lee and

Soong eds., 1996).) These approaches included the use of algorithms based on the direct estimation of transfer functions that are part of the acoustic transmission, as well as algorithms based on the use of microphone arrays.

62. The '091 patent's priority application admits processors were well known, stating that performing the "calculation of the transfer functions H1 and H2 can be accomplished by the use of any techniques used by those skilled in the art, including but not limited to adaptive techniques, recursive techniques, and simpler ones." (Ex. 1009 ('297 application), 5.) A person of ordinary skill in the art would have understood that these "calculations" would be performed by a processor and that the known "techniques" would therefore involve a processor coupled to the microphones and the VAD.

63. Thus, using a processor "coupled among the receiver and the at least one sensor that generates a plurality of transfer functions" as recited represents nothing more than the simple addition of one known element (a processor as was known in the art) to another known element (Ikeda's microphones and VAD) to obtain the predictable results of improved voice detection in noisy environments. Moreover, the combination represents the use of a known technique (processing signals and generating transfer functions using a processor) to improve a similar device and method (Ikeda's voice detection system) in the same way. The combination also applies a known technique (processing signals and generating

transfer functions using a processor) to a known device and method (Ikeda's) that is ready for improvement and yields predictable results.

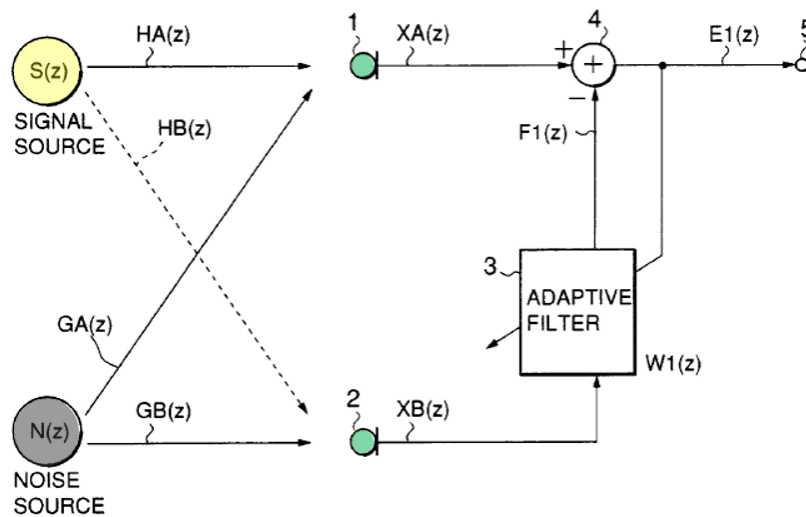
64. Alternatively, this limitation would have been obvious in view of Puthuff. Puthuff discloses the use of "available sound processing circuitry and off-the-shelf technology" such as a processor for processing microphone and accelerometer input, and explains that "digital signal processing" can be used. (Ex. 1004 (Puthuff), 6:2-5, 13:23-26; *id.*, 8:14-16, 11:8-16. Puthuff also claims a device that includes "a processing circuit for generating the electrical output signal from" signals received from microphone and vibration sensors. (*Id.*, claim 1.) A person of ordinary skill in the art would have been motivated to use a processor (as described in Puthuff) to generate the transfer functions in Ikeda for the reasons discussed above (e.g., doing so was widely known and merely reflected the use of a known technique to improve a similar device and method in the same way). Accordingly, this limitation was well known in the art and would have been obvious to a person of ordinary skill in the art in view of Ikeda and/or Puthuff.

i. 11[b][1]: Generating a First Transfer Function When Voicing Activity is Absent

65. Claim 11 recites generating "a first transfer function representative of a ratio of energy of acoustic signals received using at least two different acoustic microphones . . . wherein the first transfer function is generated in response to a

determination that voicing activity is absent from the acoustic signals for a period of time.” Ikeda discloses this limitation.

66. Ikeda’s prior art noise canceler receives a desired (e.g., speech) signal $S(z)$ (yellow) and a noise signal $N(z)$ (grey) at two microphones (green) as shown in Figure 5:



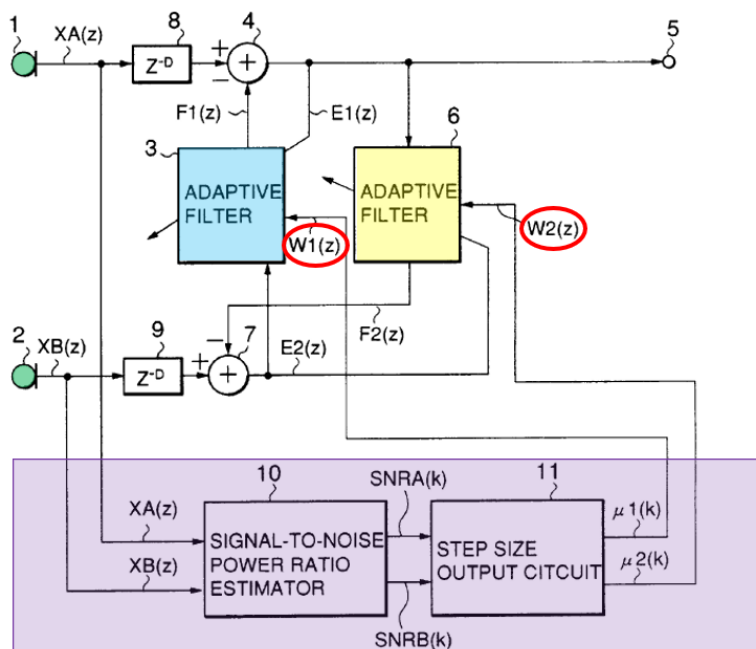
(Ex. 1003 (Ikeda), Fig. 5; ¶¶56-58 (claim element 11[a]), above).) Figure 5 shows the acoustic path between a desired speech signal $S(z)$ and primary microphone 1 which has “an acoustic transfer characteristic $HA(z)$.” (Ex. 1003 (Ikeda), 1:42-50; Fig. 5.) Figure 5 also shows the acoustic path between the desired speech signal and reference microphone 2, indicated by a dotted line, which has an acoustic transfer characteristic $HB(z)$. (*Id.*, 2:43-49, Fig. 5.) The acoustic paths from noise source $N(z)$ to the primary and reference microphones are also shown in Figure 5. (*Id.*, Fig. 5.) The acoustic path between the noise source $N(z)$ and primary microphone 1 has

69. Ikeda discloses generating a first transfer function, $W1(z)$, that is ideally approximated as

$$W1(z) = \frac{GA(z)}{GB(z)}$$

(Ex. 1003 (Ikeda), 5:17-63, Eq. 14; *id.*, 1:54-2:7, 5:17-22.) Thus, Ikeda discloses generating a first transfer function.

70. Ikeda further discloses generating transfer function $W1(z)$ in response to a determination that voicing activity is absent from the acoustic signals for a period of time. Specifically, Ikeda discloses that most closely approximating the transfer function $W1(z)$ requires calculating $W1(z)$ when “the desired signal component $S(z)$ is absent or negligibly small.” (Ex. 1003 (Ikeda), 3:32-36; *id.*, 2:45-51.) Ikeda discloses that one way to control when $W1(z)$ is calculated is by determining from “outside,” e.g., using a sensor other than the physical microphones, whether $S(z)$ is absent or negligibly small. (*Id.*, 3:32-61, 9:11-17.) However, instead of using an “outside” sensor, Ikeda’s system uses a “signal-to-noise power ratio estimator” and a “step size output circuit,” as shown in Figure 1 (purple):



(*Id.*, 4:49-56, Fig. 1.) The “signal-to-noise power ratio estimator” receives microphone signals $XA(z)$ and $XB(z)$ from main microphone 1 and microphone 2, respectively, as inputs and outputs “the estimated signal-to-noise power ratio $SNRA(k)$ of the main signal and the estimated signal-to-noise power ratio $SNRB(k)$ of the reference signal.” (*Id.*, 8:6-12, Figs. 1, 2.)

71. The “step size output circuit” receives $SNRA(k)$ and $SNRB(k)$ as inputs and outputs control signals that control “the step size[] $\mu_1(k)$ of the adaptive filter 3” and “the step size $\mu_2(k)$ of the adaptive filter 6.” (Ex. 1003 (Ikeda), 8:13-63.) The “step size [$\mu_1(k)$] is reduced when the signal-to-noise power ratio [$SNRA(k)$] is great, or it is increased when the ratio is small” while the opposite holds true for $\mu_2(k)$. (*Id.*, 8:13-63, Figs. 3-4.) Accordingly, “the step size output circuit 11 controls the step size to be fed to the adaptive filter[s].” (*Id.*, 8:64-9:3.) Moreover,

“an arrangement may be made such that the step sizes $\mu_1(k)$ and $\mu_2(k)$ assigned to the adaptive filters 3 and 6, respectively, are compared, and the smaller one of them is set to be zero in order to interrupt the learning function of the filter whose step size is determined to be zero.” (*Id.*, 9:4-10.)

72. Through this system, Ikeda updates $W_1(z)$ by a relatively larger amount when “the desired signal of the main signal obstructing the estimation is smaller than the noise signal to be estimated,” which includes when the desired signal is absent entirely such as when voicing activity is absent. (Ex. 1003 (Ikeda), 5:64-6:5, 6:16-60.) In contrast, when the desired signal is “greater than the noise signal” the amount that $W_1(z)$ updates is decreased, which can include freezing the filter by setting the update amount “to be zero in order to interrupt the learning function of the filter.” (*Id.*, 6:2-5, 9:4-10; *id.*, 6:16-60.) Thus, when voicing activity is absent (e.g., when $SNRA(k)$ and $SNRB(k)$ are small or zero), step size $\mu_1(k)$ “is increased” while $\mu_2(k)$ “is decreased.” (*Id.*, 8:36-63; *id.*, 7:54-63, 8:13-35.) After comparing “step sizes $\mu_1(k)$ and $\mu_2(k)$ assigned to the adaptive filters 3 and 6, respectively” and setting the smaller $\mu_2(k)$ to zero, adaptive filter 3 calculates $W_1(z)$. (*Id.*, 9:4-10.) Thus, Ikeda discloses generating the transfer characteristic $W_1(z)$ when voicing activity is absent from the acoustic signals.

73. Ikeda further discloses calculating transfer function $W_1(z)$ as a ratio of energy of the acoustic signals received at primary microphone 1 ($X_A(z)$) and

reference microphone 2 ($X_B(z)$) when voicing activity is absent. When voicing activity is absent, $S(z)$ is approximately equal to zero (*see* Ex. 1001 (the '091 patent), 4:52-56) and $X_A(z)$ and $X_B(z)$ simplify as follows:

$$X_A(z) = G_A(z)N(z) + H_A(z) \cdot 0$$

$$X_B(z) = G_B(z)N(z) + H_B(z) \cdot 0$$

and therefore, by eliminating the terms equal to zero,

$$X_A(z) = G_A(z)N(z)$$

$$X_B(z) = G_B(z)N(z)$$

Solving for $G_A(z)$ and $G_B(z)$ by dividing both sides by $N(z)$ gives:

$$G_A(z) = \frac{X_A(z)}{N(z)}$$

$$G_B(z) = \frac{X_B(z)}{N(z)}$$

Accordingly, Ikeda's transfer function $W1(z)$ is approximated as follows when voicing activity is absent:

$$W1(z) = \frac{G_A(z)}{G_B(z)} = \frac{X_A(z) / N(z)}{X_B(z) / N(z)} = \frac{X_A(z)}{X_B(z)}$$

Thus, Ikeda discloses or at least renders obvious that, when voicing activity is absent, the transfer function $W1(z)$ is representative of a ratio of energy of acoustic signals received at microphones 1 ($X_A(z)$) and 2 ($X_B(z)$).

74. Ikeda’s disclosure of this limitation is confirmed by comparing Ikeda’s equations to those in the ’091 patent:

	’091 Specification	Ikeda
Primary microphone input	$M_1(z)$	$XA(z)$
Reference microphone input	$M_2(z)$	$XB(z)$
First transfer function ³	$H_1(z) = \frac{M_1(z)}{M_2(z)}$	$W1(z) = \frac{XA(z)}{XB(z)}$

(Compare Ex. 1001 (the ’091 patent), 4:3-8, 4:40-45 with Ex. 1003 (Ikeda), 5:6-9.)

ii. 11[b][2]: Generating a Second Transfer Function When Voicing Activity is Present

75. Claim 11 recites generating “a second transfer function representative of the acoustic signals, wherein the second transfer function is generated in response to a determination that voicing activity is present in the acoustic signals for the period of time.” Ikeda discloses this limitation.

76. As discussed above, Ikeda’s noise canceler receives a noise signal $N(z)$ and a desired (speech) signal $S(z)$ at two microphones. ¶¶65-74 (claim element

³ The subscript “n” in $M_{1n}(z)$ and $M_{2n}(z)$ has been removed for simplicity. It is merely a conventional notation to show that only noise is present when transfer function $H_1(z)$ is calculated. (Ex. 1001 (the ’091 patent), 4:37-39, 6:1-4.)

11[b][1]), above.) Ikeda further discloses generating a second transfer function, $W2(z)$ that is ideally approximated as

$$W2(z) = \frac{HB(z)}{HA(z)}$$

(Ex. 1003 (Ikeda), 5:18-63, Eq. 15.) Thus, Ikeda discloses generating a second transfer function.

77. Ikeda further discloses generating transfer function $W2(z)$ when voicing activity is present. Specifically, Ikeda discloses that most closely approximating the transfer function $W2(z)$ requires calculating $W2(z)$ when “signal component $S(z)$ is present and the noise component $N(z)$ is absent or negligibly small.” (Ex. 1003 (Ikeda), 3:36-40; *see id.*, 2:45-51.) As discussed above, Ikeda determines when this occurs using the “signal-to-noise power ratio estimator” and a “step size output circuit.” (*Id.*, 5:64-6:15; ¶¶65-74 (claim element 11[b][1]), above.)

78. Ikeda discloses that calculating $W2(z)$ when voicing activity is present results in $W2(z)$ being representative of the acoustic signals. When voicing activity is present but noise is absent or negligible, $N(z)$ is approximately equal to zero, and $XA(z)$ and $XB(z)$ simplify as follows:

$$XA(z) = GA(z) \cdot 0 + HA(z)S(z)$$

$$XB(z) = GB(z) \cdot 0 + HB(z)S(z)$$

and therefore, by eliminating the terms equal to zero,

$$XA(z) = HA(z)S(z)$$

$$XB(z) = HB(z)S(z)$$

Solving for HA(z) and HB(z) by dividing both sides by S(z) results in:

$$HA(z) = \frac{XA(z)}{S(z)}$$

$$HB(z) = \frac{XB(z)}{S(z)}$$

Accordingly, Ikeda's transfer function W2(z) is approximated as follows when voicing activity is present and noise is negligible:

$$W2(z) = \frac{HB(z)}{HA(z)} = \frac{XB(z) / S(z)}{XA(z) / S(z)} = \frac{XB(z)}{XA(z)}$$

Thus, when voicing activity is present and noise is negligible, the transfer function W2(z) is representative of the acoustic signals received at microphones 1 (XA(z)) and 2 (XB(z)).

79. Ikeda's disclosure of this limitation is confirmed by comparing Ikeda's equations to the similar equations provided in support of this limitation in the '091 patent:

	'091 Specification	Ikeda
Primary microphone input	$M_1(z)$	$XA(z)$
Reference microphone input	$M_2(z)$	$XB(z)$
Second transfer function ⁴	$H_2(z) = \frac{M_2(z)}{M_1(z)}$	$W2(z) = \frac{XB(z)}{XA(z)}$

(Compare Ex. 1001 (the '091 patent), 4:3-8, 5:1-5 with Ex. 1003 (Ikeda), 5:6-9.)

d. 11[c]: Noise Removal

80. Claim 11 recites “wherein acoustic noise is removed from the acoustic signals using the first transfer function and at least one combination of the first transfer function and the second transfer function to produce the denoised acoustic data stream.” Ikeda discloses this limitation.

81. Ikeda discloses generating a first transfer function $W1(z)$ and a second transfer function $W2(z)$. (¶¶65-79 (claim elements 11[b][1]-[2]), above.) Ikeda uses these two transfer functions to remove noise from the acoustic signals resulting in denoised output signal $E1(z)$. (Ex. 1003 (Ikeda), 4:52-53, 5:12-15, Fig. 1.) The denoised output signal $E1(z)$ is represented by:

$$E1(z) = XA(z) - W1(z)E2(z)$$

where

⁴ The subscript “s” in $M_{1s}(z)$ and $M_{2s}(z)$ has been removed for simplicity as explained above. (¶74, n.3 (claim element 11[b][1]), above.)

$$E2(z) = XB(z) - W2(z)E1(z)$$

(*Id.*, 5:17-30, Eqs. 8-9.) These equations can be rearranged as follows. First, the value of $E2(z)$ from the second equation can be substituted into the first equation:

$$E1(z) = XA(z) - W1(z) \cdot (XB(z) - W2(z)E1(z))$$

then, multiplying out the right-hand term,

$$E1(z) = XA(z) - XB(z)W1(z) + W1(z)W2(z)E1(z)$$

and subtracting the right-hand term from both sides,

$$E1(z) - W1(z)W2(z)E1(z) = XA(z) - XB(z)W1(z)$$

factoring $E1(z)$ out of the left side of the equation,

$$E1(z) \cdot (1 - W1(z)W2(z)) = XA(z) - XB(z)W1(z)$$

and finally, solving for $E1(z)$ results in:

$$E1(z) = \frac{XA(z) - XB(z)W1(z)}{1 - W1(z)W2(z)}$$

Thus, Ikeda discloses or renders obvious using the first transfer function ($W1(z)$, in the numerator of the above fraction) and a combination of the first transfer function and the second transfer function ($W1(z)W2(z)$, in the denominator), to produce the denoised output.

82. Ikeda's disclosure of this limitation is confirmed by comparing Ikeda's equations to those in the '091 patent:

	'091 Specification	Ikeda
Primary microphone input	$M_1(z)$	$XA(z)$
Reference microphone input	$M_2(z)$	$XB(z)$
First transfer function	$H_1(z) = \frac{M_1(z)}{M_2(z)}$	$W1(z) = \frac{XA(z)}{XB(z)}$
Second transfer function	$H_2(z) = \frac{M_2(z)}{M_1(z)}$	$W2(z) = \frac{XB(z)}{XA(z)}$
Output after removing noise using the first transfer function and a combination of the first and second transfer functions	$S(z)$ $= \frac{M_1(z) - M_2(z)H_1(z)}{1 - H_2(z)H_1(z)}$	$E1(z)$ $= \frac{XA(z) - XB(z)W1(z)}{1 - W2(z)W1(z)}$

(Compare Ex. 1001 (the '091 patent), 4:3-8; 4:40-45, 5:1-5, 5:22-26, Eq. 3 with Ex. 1003 (Ikeda), 5:6-9.)

83. Thus, Ikeda discloses or at least renders obvious this limitation.

e. 11[d]: Vibration Sensor

84. Claim 11 recites “at least one sensor that receives human tissue vibration information associated with human voicing activity of a user.” Puthuff discloses this limitation.

85. Puthuff discloses a system for detecting and transmitting speech using an acoustic microphone and a “vibration sensor” such as an “accelerometer.” (Ex. 1004 (Puthuff), 5:1-9.) The acoustic microphone may be “any conventional acoustic sound transducer such as, e.g., a standard microphone.” (*Id.*, 9:1-6.) The accelerometer “is a transducer device which senses the vibratory displacement of the

bones in proximity to the ear as imparted primarily by vibrations during speech.” (*Id.*; *id.*, 8:12-16.) The device is “held against the skin of the talker’s skull and may take the form of an ear level pickup worn behind the ear, or a telephone handset held against the ear during telephonic communication.” (*Id.*, 8:16-19.)

86. Puthuff explains that an acoustic microphone is better for picking up a talker’s voice than using bone conduction in “an acoustically quiet environment,” but that bone conduction becomes “a better alternative” when acoustic noise levels are high because the quality of acoustic microphone detection “degrade[s].” (Ex. 1004 (Puthuff), 4:19-24.) Puthuff’s system therefore uses both methods of voice detection. “At low ambient noise levels, the microphone signal is permitted to dominate as the output signal” (i.e., “sound detection is effected primarily through a conventional microphone”). (*Id.*, 10:13-14, 5:4-7.) In other words, Puthuff’s system prioritizes the output of its acoustic microphone signal over the output of its accelerometer in environments with low noise levels. However, at “high ambient noise levels,” the system prioritizes the output of “the bone conduction component of the audio signal [i.e., the accelerometer’s output], comprising primarily the talker’s speech.” (*Id.* 10:10-25.) In this way, the system “exploits the phenomenon of bone conduction to suppress the deleterious effects of high acoustic noise environments by “providing a device which adaptively senses sound through bone conduction when acoustic noise levels are high.” (*Id.*, 4:13-16.) To accomplish this,

Puthuff uses a transition circuit that fades out the acoustic microphone signals and fades in the “bone vibration sensor signal” as ambient noise levels increase. (*Id.*, 5:7-9; *id.*, 10:5-11:5.)

87. Puthuff therefore discloses a “sensor that receives human tissue vibration information associated with human voicing activity of a user” as recited in this claim.

i. Motivation to Combine Ikeda and Puthuff

88. A person of ordinary skill in the art would have immediately recognized that one could combine the teachings of Ikeda and Puthuff and use Puthuff’s vibration sensor either (a) in place of, or (b) in addition to, Ikeda’s VAD (consisting of the signal-to-noise power ratio estimator and, optionally, the step size output circuit) to more accurately determine periods when voicing activity is absent or present. Each of the two proposed combinations are addressed below.

89. Puthuff discloses that its system functions as a VAD to indicate whether or not voicing activity is present. For example, Puthuff teaches that the incorporation of a vibration sensor “improves system immunity to false triggering due to noisy environments” and determines “the absence of the user’s voice” to conserve power. (Ex. 1004 (Puthuff), 6:6-10.) One of ordinary skill in the art would have been motivated to use a vibration sensor such as Puthuff’s accelerometer as a

VAD in Ikeda's noise canceler instead of Ikeda's signal-to-noise power ratio estimator for at least the following reasons:

90. First, Ikeda and Puthuff address the same problem—accurately detecting voice in noisy environments. (Ex. 1003 (Ikeda), 3:32-61; Ex. 1004 (Puthuff), Abstract, 3:1-5, 3:22-34.)

91. Second, Ikeda discloses the need for accurately determining periods when a desired signal (e.g., voicing activity) is present and periods when a desired signal is absent for an adaptive noise canceler to function optimally. (Ex. 1003 (Ikeda), 3:32-54.) Thus, person of ordinary skill in the art would have been motivated to look to other voice activity detection methods. Puthuff explains that “vibration transducers and even accelerometers” had been used for decades to detect voicing activity and transmit signals. (Ex. 1004 (Puthuff), 3:19-24.) Puthuff also teaches that its vibration sensor reduces false triggering in noisy environments, and that the system is capable of determining “the absence of the user's voice.” (*Id.*, 6:6-10.) Thus, a person of ordinary skill in the art would have been motivated to incorporate a vibration sensor, as in Puthuff, to provide improved voice activity detection.

92. Third, Puthuff discloses that voice signals can be detected with an accelerometer using “off-the-shelf components” with “reduced production costs and power consumption requirements.” (Ex. 1004 (Puthuff), 6:2-10, 12:17-13:2.) Thus,

a person of ordinary skill in the art would have been motivated to use an accelerometer sensor in Ikeda's system as an inexpensive, easy-to-find component for detecting voice activity.

93. Fourth, the use of vibration sensors such as accelerometers as noise-resistant sensors for detecting voicing activity was well known for decades, as Puthuff recognizes. (Ex. 1004 (Puthuff), 2:20-24, 3:19-25.) Thus, using a vibration sensor (as in Puthuff) as a VAD in Ikeda's system would represent nothing more than the simple addition of one known element (a vibration sensor used to detect voicing activity as in Puthuff) to another known element (a system requiring voice activity detection as in Ikeda) to obtain the predictable results of improved voice activity detection, particularly in noisy environments. Moreover, the combination represents the use of a known technique (detecting voicing activity with a vibration sensor as in Puthuff) to improve a similar device and method (Ikeda's noise canceling system that relies on voice activity detection) in the same way. The combination also applies a known technique (detecting voicing activity with a vibration sensor as in Puthuff) to a known device and method (Ikeda's system) that is ready for improvement and yields predictable results (a system with improved voice detection, particularly in noisy environments).

94. Fifth, Puthuff expressly teaches that its vibration sensor is useful for "dual microphone" systems. (*Id.*, 6:11-23.) Thus, a person of ordinary skill in the

art would have been motivated to combine Puthuff's teachings with Ikeda's dual microphone system, and would have had a reasonable expectation of success in doing so.

95. A person of ordinary skill in the art would have further expected success in modifying Ikeda's noise canceler to use an accelerometer or other vibration sensor as a VAD because the signal processing of such an adaptive noise canceler does not depend on the type of VAD that is used as long as it is accurate. Similarly, a vibration sensing VAD would detect voice signals in a predictable manner regardless of how such detected signals were used after being detected.

96. Combining Ikeda and Puthuff also would have been obvious for a second, independent reason. A person of ordinary skill in the art alternatively would have been motivated to incorporate Puthuff's vibration sensor and transition circuit into Ikeda such that the output of Ikeda's microphone 1, $XA(z)$, was used in Puthuff's system as the output of the microphone 20. In this combination, Ikeda's "main signal" $XA(z)$ would be input into Puthuff's equalizer 22, amplified, faded, and combined with the signal from the accelerometer 30 at the adder circuit 38. (See Ex. 1004 (Puthuff), Fig. 1.) The output signal 40 from Puthuff's system would be input into Ikeda's signal-to-noise power ratio generator in place of the original microphone-only signal $XA(z)$.

97. A person of ordinary skill in the art would have immediately recognized that Puthuff's vibration sensor and transition circuit could be used so that the signal sent to Ikeda's signal-to-noise power ratio estimator as the "main signal" would be more accurate under varying conditions as follows: during periods of low noise, the signal transmitted to the signal-to-noise power ratio estimator would be predominantly the signal from the first microphone 1. However, during periods of high ambient noise, the signal transmitted to the signal-to-noise power ratio estimator would be predominantly the "bone conduction component of the audio signal, comprising primarily the talker's speech." (Ex. 1004 (Puthuff), 10:10-13.) A POSITA would have been motivated to modify Ikeda in this manner for several reasons, including the first through third and fifth reasons provided above. (¶¶90-92, 94, above.) Furthermore, such a modification would have improved the accuracy of Ikeda's signal-to-noise power ratio generator because it would ensure that the "main signal" being provided to the signal-to-noise power ratio generator (which in Ikeda's system is only the output of microphone 1, $XA(z)$) was the most accurate representation of the speech signal (the "output signal 40" in Puthuff). Such a modification would be particularly advantageous in high-noise environments for the many reasons Puthuff explains.

98. A person of ordinary skill in the art would have expected success in combining Ikeda's noise canceler with an accelerometer or other vibration sensor

for detecting voice signals, and to transmit to the output signal 40 from Puthuff's system to Ikeda's signal-to-noise power ratio estimator, because the signal-to-noise power ratio estimator would simply receive a more accurate and better representation of the user's voice in high noise environments. As discussed above, such a modification would require nothing more than off-the-shelf components. (§92, above.)

4. Claim 12

99. Claim 12 depends from claim 11 and further recites "wherein the sensor includes a mechanical sensor in contact with the skin." Puthuff discloses this limitation because it discloses that its accelerometer voicing sensor (which is a mechanical sensor) can be "held up against the skin of the talker's skull" or "against the ear during telephonic communication." (Ex. 1004 (Puthuff), 8:16-19; *id.*, Abstract (explaining that the vibration sensor may be an accelerometer, velocity sensor, or displacement sensor), 5:1-4, 19:6-20:4; §§84-98 (claim element 11[d]), above.)

5. Claim 13

100. Claim 13 depends from claim 11 and further recites "wherein the sensor includes at least one of an accelerometer, a skin surface microphone in physical contact with skin of a user, a human tissue vibration detector, a radio frequency (RF) vibration detector, and a laser vibration detector." Puthuff discloses this limitation

because it discloses that the vibration sensor may be an accelerometer, velocity sensor, or displacement sensor. (Ex. 1004 (Puthuff), 5:1-4, 8:16-19, 19:6-20:4; ¶¶84-98 (claim element 11[d]), ¶99 (claim 12), above.)

6. Claim 14

101. Claim 14 depends from claim 11 and further recites “wherein the human tissue is at least one of on a surface of a head, near the surface of the head, on a surface of a neck, near the surface of the neck, on a surface of a chest, and near the surface of the chest.” Puthuff discloses this limitation because it discloses sensing tissue vibrations behind the ear (e.g., on or near a surface of the head) or in a face mask. (Ex. 1004 (Puthuff), 8:16-19.) Alternatively, it would have been obvious to one of ordinary skill in the art to position the vibration sensor on or near the head, neck, or chest because Ikeda teaches that vibration sensors are effective when placed on the “throat” or “any part of a person’s body supported by bone structure.” (*Id.*, 2:20-3:5, 3:19-23; ¶¶84-98 (claim element 11[d]), above.)

7. Claim 16

102. Claim 16 depends from claim 11 and further recites that the receiver “includes a plurality of independently located microphones.” Ikeda discloses this limitation because it discloses two independently located microphones. (Ex. 1003 (Ikeda), Figs. 1, 5; ¶¶56-58 (claim element 11[a]), above.)

8. Claim 18

103. Independent claim 18 is directed to a “signal processing system” and recites several components addressed in claim 11, including a processor that generates transfer functions, a receiver, microphones, and a voicing sensor. Ikeda and Puthuff disclose and render obvious each of these limitations as discussed above. (¶¶54-98 (claim 11), above.)

a. 18: Preamble

104. Claim 18’s preamble recites “[a] signal processing system coupled among a user and an electronic device, wherein the signal processing system includes a denoising subsystem for removing acoustic noise from acoustic signals.” To the extent the preamble is limiting, Ikeda discloses it because Ikeda discloses a “noise canceler for canceling, by use of an adaptive filter, a background noise signal introduced into a speech signal input via a microphone, a handset or the like” that comprises an electronic device and is intended for use by a user in, for example, a hearing aid or telephone handset. (Ex. 1003 (Ikeda), 8:16-20, 33:10-13, 1:4-2:6; ¶55 (claim 11 preamble), above.)

b. 18[a]: Processor Coupled Among a Receiver and Sensor

105. Claim 18 recites “a processor coupled among a receiver and at least one sensor, wherein the receiver is coupled to receive the acoustic signals via at least two

microphones.” Ikeda discloses and/or renders obvious this limitation for the reasons discussed above. (¶¶56-79 (claim elements 11[a]-11[b]), above.)

c. 18[b]: Voicing Sensor

106. Claim 18 recites “wherein the at least one sensor detects human tissue vibration associated with human voicing activity of a user.” Puthuff discloses this limitation, and a person of ordinary skill in the art would have been motivated to combine Ikeda and Puthuff for the reasons discussed above. (¶¶84-98 (claim element 11[d]), above.)

d. 18[c]: Processor That Generates Transfer Functions

107. Claim 18 recites “wherein the processor generates a plurality of transfer functions.” Ikeda and Puthuff disclose and render obvious this limitation for the reasons discussed above. (¶¶59-79 (claim element 11[b]), above.)

i. 18[c][1]: First Transfer Function

108. Claim 18 recites “wherein a first transfer function representative of a ratio of acoustic energy received by the two microphones is generated in response to a determination that voicing activity is absent from the acoustic signals for a specified period of time.” Ikeda discloses and/or renders obvious this limitation. (¶¶65-74 (claim element 11[b][1]), above.)

ii. 18[c][2]: Second Transfer Function

109. Claim 18 recites “wherein a second transfer function representative of the acoustic signals is generated in response to a determination that voicing activity is present in the acoustic signals for a specified period of time.” Ikeda discloses and/or renders obvious this limitation. (¶¶75-79 (claim element 11[b][2]), above.)

e. 18[d]: Noise Removal

110. Claim 18 recites “wherein acoustic noise is removed from the acoustic signals using the first transfer function and at least one combination of the first transfer function and the second transfer function to produce a denoised acoustic data stream.” Ikeda discloses and/or renders obvious this limitation. (¶¶80-83 (claim element 11[c]), above.)

9. Claim 19

111. Claim 19 depends from claim 18 and further recites “wherein the at least one electronic device includes at least one of cellular telephones, personal digital assistants, portable communication devices, computers, video cameras, digital cameras, and telematics systems.” Ikeda and Puthuff disclose and/or render obvious this limitation.

112. Ikeda discloses a system for canceling noise in signals received by “a microphone, a handset or the like.” (Ex. 1003 (Ikeda), 1:4-19.) Ikeda also explains that “a background noise signal introduced into a speech signal input via, e.g., a

microphone or handset is a critical problem when it comes to narrow band speech coder[s], speech recognition device[s] and so forth.” (*Id.*, 1:9-13.) A person of ordinary skill in the art would have understood that, at the time, a “narrow band speech coder” was a common component in cellular telephones because it was “generally known” to convert signals using a “speech coder” for the “transmission of speech to or from a mobile station.” (Ex. 1012 (Oestreich), 1:9-2:7.) Furthermore, a person of ordinary skill in the art would have understood that a “handset” was a common way to refer to a telephone microphone, including a cellular telephone microphone. Accordingly, a person of ordinary skill would have understood Ikeda to disclose, or at least suggest, the use of its system in at least a cellular telephone. Moreover, a person of ordinary skill in the art would have understood that a “handset” and/or a “microphone” was a common component of a “portable communication device” and thus would have understood Ikeda to disclose that its system could be used in such a device. Ikeda therefore discloses and/or renders obvious this limitation.

113. Puthuff discloses that its voicing sensor “may be part of a stand alone ear level communication transmitter” or “implemented in a telephone system or other communication systems.” (Ex. 1004 (Puthuff), Title, Abstract.) Puthuff further discloses embodiments in a “telephone handset” or in a “face mask worn with, for example, a firefighter’s helmet.” (*Id.*, 8:16-20.) Accordingly, Puthuff

discloses this limitation and it would have been obvious to implement Ikeda's system, modified to include a vibration sensor, as taught by Puthuff, in a cellular telephone or portable communication device.

10. Claim 20

114. Claim 20 depends from claim 18 and further recites “wherein the human tissue is at least one of on a surface of a head, near the surface of the head, on a surface of a neck, near the surface of the neck, on a surface of a chest, and near the surface of the chest.” Puthuff discloses this limitation for the reasons discussed above. (¶101 (claim 14), above.)

11. Claim 1

115. Independent claim 1 is a method claim that recites several functions addressed in claim 11, including receiving acoustic signals, receiving a voice activity signal, generating transfer functions, and removing noise. These limitations are disclosed by the prior art for the reasons discussed above. (¶¶54-98 (claim 11), above.)

a. 1: Preamble

116. Claim 1 recites “[a] method for removing noise from acoustic signals.” To the extent the preamble is limiting, Ikeda discloses such a method because it discloses the use of a “noise canceler.” (Ex. 1003 (Ikeda), 1:4-19; ¶55 (claim 11 preamble), above.)

b. 1[a]: Receiving Acoustic Signals

117. Claim 1 recites “receiving at least two acoustic signals using at least two acoustic microphones positioned in a plurality of locations.” Ikeda discloses this limitation for the reasons discussed above. (¶¶56-58 (claim element 11[a]), above.)

c. 1[b]: Sensing Vibrations

118. Claim 1 recites “receiving a voice activity signal that includes information on vibration of human tissue associated with human voicing activity of a user.” Puthuff discloses this limitation, and a person of ordinary skill in the art would have been motivated to combine its disclosure with Ikeda, for the reasons discussed above. (¶¶84-98 (claim element 11[d]), above.)

d. 1[c]: Generating a VAD Signal

119. Claim 1 recites “generating a voice activity detection (VAD) signal using the voice activity signal.” Ikeda and Puthuff disclose this limitation and/or render it obvious.

120. Ikeda discloses this limitation. As discussed above, Ikeda discloses a noise canceler that includes a “signal-to-noise power ratio estimator” and “step size output circuit” that process signals from physical microphones, determine the presence or absence of a desired signal (e.g., a voice signal) and, based on that determination, control the generation of transfer functions by the adaptive filters.

(¶¶65-79 (claim elements 11[b][1]-[2]), above.) A person of ordinary skill in the art would have understood that the signals output by the signal-to-noise power ratio estimator and/or the step size output circuit comprise a “voice detection (VAD) signal” because they contain information about whether a desired signal (e.g., a voice signal) is present, which is used to control Ikeda’s adaptive filters.

121. Additionally, Ikeda discloses that prior, conventional systems were configured to “detect the [] sections where the desired signal component $S(z)$ is absent (or little) and the sections where the noise signal component $N(z)$ is absent (or little) and to command the adaptive filters to perform lea[r]ning [e.g., calculate a transfer function] in such sections.” (Ex. 1003 (Ikeda), 3:40-45.) A person of ordinary skill in the art would have understood that “command[ing] the adaptive filters” to calculate a transfer function at the appropriate time requires generating a VAD signal to control when the noise canceler’s adaptive filters generate transfer functions. (*Id.*, 3:50-54.) Accordingly, to the extent Puthuff’s voice activity signal is used as an input into Ikeda’s VAD as discussed above (¶¶84-98 (claim element 11[d]), above), Ikeda’s VAD generates a voice activity detection signal using Puthuff’s voice activity signal.

122. Puthuff also discloses this limitation. Puthuff discloses that its vibration sensor is capable of detecting the presence of a user’s voice. (¶¶84-98 (claim element 11[d]), above.) Puthuff further discloses that its system is capable of

transmitting a signal to “turn[] off a radio transmitter in the absence of [a] user’s voice” based on what is detected by its vibration sensor. (Ex. 1004 (Puthuff), 6:11-20.) A person of ordinary skill in the art would have understood that such a signal is a VAD signal because it contains information regarding whether Puthuff’s vibration sensor has detected voice activity. Puthuff therefore discloses this limitation to the extent that Puthuff’s VAD would replace Ikeda’s VAD.

e. 1[d]: Generating at Least Two Transfer Functions

123. Claim 1 recites “generating at least two transfer functions representative of a ratio of energy of the acoustic signal received using at least two different acoustic microphones of the at least two acoustic microphones when the VAD indicates that user voicing activity is absent, wherein the at least two transfer functions comprise a first transfer function and a second transfer function.” Ikeda discloses this limitation.

124. I understand the patent owner has asserted that claim 1 has the same scope as claims 11 and 18. Based on that contention, Ikeda discloses and/or renders obvious this limitation for the same reasons discussed above for claim 11. (¶¶65-79 (claim elements 11[b][1]-[2]), above.)

125. Ikeda also discloses this limitation as written. Ikeda discloses that
for the adaptive filter 3 [which applies the first transfer
function discussed above] to estimate a noise signal

contained in the main signal accurately, it is necessary to increase the amount of updating of the filter coefficient when the desired signal of the main signal obstructing the estimation is smaller than the noise signal to be estimated. Conversely, when the desired signal of the main signal is greater than the noise signal, it is necessary to *reduce* [e.g., not eliminate] the above amount because the signal obstructing the estimation is greater than the noise signal.

On the other hand, for the adaptive filter 6 [which applies the second transfer function discussed above] to estimate the desired signal of the reference signal accurately, it is necessary to increase the amount of updating of the filter coefficient when the noise signal contained in the reference signal obstructing the estimation is smaller than the desired signal. Conversely, when the noise signal of the reference signal is greater than the desired signal, it is necessary to *reduce* [e.g., not eliminate] the above amount because the signal obstructing the estimation is greater than the desired signal.

(Ex. 1003 (Ikeda), 5:64-6:15 (emphasis added); *cf. id.*, 9:4-10 (discussing an *alternative* embodiment in which only one transfer function updates at a time).) Thus, Ikeda discloses that both the first and second transfer functions may update *both* when “the noise signal . . . is smaller than the desired signal” (e.g., when voicing activity is present) *and* when “the noise signal . . . is greater than the desired signal” (e.g., when voicing activity is not present). (*Id.*) Accordingly, Ikeda discloses that both the first and second transfer functions may be generated (e.g., updated) when voicing activity is absent.

126. As discussed above, Ikeda discloses that its VAD signal is used to indicate when voicing activity is present or absent and to control the generation of Ikeda's transfer functions. (¶¶65-79 (claim elements 11[b][1]-[2]), above.) A person of ordinary skill in the art would also have understood that, depending on how Puthuff's vibration sensor is combined with Ikeda's system (¶¶84-98 (claim element 11[d]), above), Puthuff's vibration sensor may likewise provide a VAD signal that Ikeda's system would use in lieu of Ikeda's VAD signal for the same purpose of controlling the generation of Ikeda's transfer functions. Ikeda and Puthuff therefore disclose this limitation.

f. 1[e]: Removing Noise

127. Claim 1 recites "removing acoustic noise from at least one of the acoustic signals by applying the first transfer function and at least one combination of the first transfer function and the second transfer function to the acoustic signals and generating denoised acoustic signals." Ikeda discloses this limitation as discussed above. (¶¶80-83 (claim element 11[c]), above.)

12. Claim 2

a. 2[a]: Generating One Transfer Function

128. Claim 2 depends from claim 1 and further recites "wherein removing noise further comprises generating one transfer function of the at least two transfer functions to be representative of a ratio of energy of the acoustic signal received

when the VAD indicates that user voice activity is present.” As discussed above, Ikeda discloses this limitation because it discloses that its transfer functions, at least one of which is representative of a ratio of energy of the acoustic signals received, may both be generated (e.g., updated) when the VAD indicates that user voice activity is present. (¶¶65-74, 84-98 (claim elements 11[b][1], 11[d]), above.)

b. 2[b]: Removing Noise

129. Claim 2 also recites “removing noise from the acoustic signals using at least one combination of the at least two transfer functions to generate the denoised acoustic signals.” Ikeda discloses this limitation for the reasons discussed above. (¶¶80-83 (claim element 11[c]), above.)

13. Claim 3

130. Claim 3 depends from claim 1 and further recites “wherein the acoustic signals include at least one reflection of at least one associated noise source signal and at least one reflection of at least one acoustic source signal.” Ikeda discloses the use of its system to cancel “a background noise signal introduced into a speech signal.” (Ex. 1003 (Ikeda), 1:4-8.) A person of ordinary skill in the art would have understood that any real-world application involving acoustic signals (e.g., speech signals and background noise signals) would take place in echoic or reverberant environments which inherently create at least one reflection of both the background noise signal (noise source signal) and the speech signal (acoustic source signal).

(*See, e.g.*, Ex. 1007 (Hussain), 2611.) Thus, Ikeda inherently discloses, or at least renders obvious, this limitation. Alternatively, this would have been obvious in view of Puthuff's disclosure that such systems can be used in a face mask worn with a firefighter's helmet, because a person of ordinary skill in the art would have understood that such masks reflect the noise and acoustic source signals. (Ex. 1004 (Puthuff), 8:16-20.)

14. Claim 4

131. Claim 4 depends from claim 1 and further recites “wherein generating the at least two transfer functions comprises recalculating the at least two transfer functions during at least one prespecified interval.” For the reasons discussed above, Ikeda discloses this limitation because it discloses recalculating transfer function $W1(z)$ at least during the prespecified interval when voice activity is absent and recalculating transfer function $W2(z)$ at least during the prespecified interval during which voice activity is present. (¶¶65-79 (claim elements 11[b][1]-[2]), above.) Furthermore, it would have been obvious to a person of ordinary skill in the art in view of their own knowledge to recalculate the transfer functions $W1(z)$ and $W2(z)$ at regular (e.g., prespecified) intervals during periods when voicing activity is absent and/or present, for example, to match the system's sample rate and/or to ensure that the transfer functions continue to closely approximate their ideal values, thus ensuring accurate noise cancellation.

15. Claim 5

132. Claim 5 depends from claim 1 and further recites “wherein generating the at least two transfer functions comprises use of at least one technique selected from a group consisting of adaptive techniques and recursive techniques.”

133. Ikeda discloses this limitation. Ikeda discloses the use of adaptive filters for calculating transfer functions with the use of an “LMS algorithm” which is a “typical” adaptive algorithm for updating transfer function coefficients. (Ex. 1003 (Ikeda), 1:31-41, 2:64-65 (“the filter 6 performs the above learning based on, e.g., the LIM scheme.”), 6:19-61 (explaining “[a] method of updating the [adaptive filter] coefficient[], assuming the LIM scheme as a learning algorithm and the adaptive filter 3 by way of example.”), Figs. 1, 5 (showing the use of adaptive filters).) Thus, Ikeda discloses this limitation.

134. Even if Ikeda had not disclosed this limitation, it would have been obvious. PO admitted that techniques, including adaptive and recursive techniques, for generating transfer functions were known in the art. (Ex. 1009 (’297 application), 5 (“The calculation of the transfer functions H1 and H2 can be accomplished by the use of any techniques used by those skilled in the art, including but not limited to adaptive techniques, recursive techniques, and simpler ones.”).) Accordingly, the additional limitation of this claim does not add any novel subject

matter and claim 5 is thus obvious for the same reasons as claim 1. (§§65-79 (claim elements 11[b][1]-[2]), above.)

16. Claim 6

135. Claim 6 depends from claim 1 and further recites “wherein information on the vibration of human tissue is provided by a sensor in contact with the skin.” Puthuff discloses this limitation for the reasons discussed above. (§§84-98 (claim element 11[d]), above.)

17. Claim 7

136. Claim 7 depends from claim 1 and further recites “wherein information on the vibration of human tissue is provided via at least one sensor selected from among at least one of an accelerometer, a skin surface microphone in physical contact with skin of a user, a human tissue vibration detector, a radio frequency (RF) vibration detector, and a laser vibration detector.” Puthuff discloses this limitation for the reasons discussed above. (§100 (claim 13), above.)

18. Claim 8

137. Claim 8 depends from claim 1 and further recites “wherein the human tissue is at least one of on a surface of a head, near the surface of the head, on a surface of a neck, near the surface of the neck, on a surface of a chest, and near the surface of the chest.” Puthuff discloses this limitation for the reasons discussed above. (§101 (claim 14), above.)

B. Claims 1-8, 11-14, 16, and 18-20 Would Have Been Obvious in View of Ikeda and Alcivar.

138. Ikeda and Alcivar render claims 1-8, 11-14, 16, and 18-20 obvious.

1. Claim 11

a. 11: Preamble

139. The preamble of claim 11 recites “[a] system for removing acoustic noise from the acoustic signals.” To the extent the preamble is limiting, Ikeda discloses it for the reasons discussed above. (¶¶55 (Ikeda and Puthuff, claim 11 preamble), above.)

b. 11[a]: Receiver

140. Claim 11 recites “a receiver that receives at least two acoustic signals via at least two acoustic microphones positioned in a plurality of locations.” Ikeda discloses this limitation as discussed above. (¶¶56-58 (Ikeda and Puthuff, claim element 11[a]), above.)

c. 11[b]: Processor that Generates Transfer Functions

141. Claim 11 next recites “a processor coupled among the receiver and the at least one sensor that generates a plurality of transfer functions, wherein the plurality of transfer functions includes a first transfer function . . . [and] a second transfer function.” Ikeda renders this limitation obvious as discussed above. (¶¶59-64 (Ikeda and Puthuff, claim element 11[b]), above.)

i. 11[b][1]: First Transfer Function

142. Claim 11 recites generating “a first transfer function representative of a ratio of energy of acoustic signals received using at least two different acoustic microphones . . . wherein the first transfer function is generated in response to a determination that voicing activity is absent from the acoustic signals for a period of time.” Ikeda discloses this limitation as discussed above. (¶¶65-74 (Ikeda and Puthuff, claim element 11[b][1]), above.)

ii. 11[b][2]: Second Transfer Function

143. Claim 11 recites generating “a second transfer function representative of the acoustic signals, wherein the second transfer function is generated in response to a determination that voicing activity is present in the acoustic signals for the period of time.” Ikeda discloses this limitation as discussed above. (¶¶75-79 (Ikeda and Puthuff, claim element 11[b][2]), above.)

d. 11[c]: Noise Removal

144. Claim 11 recites “wherein acoustic noise is removed from the acoustic signals using the first transfer function and at least one combination of the first transfer function and the second transfer function to produce the denoised acoustic data stream.” Ikeda discloses this limitation. (¶¶80-83 (Ikeda and Puthuff, claim element 11[c]), above.)

e. 11[d]: Vibration Sensor

145. Claim 11 recites “at least one sensor that receives human tissue vibration information associated with human voicing activity of a user.” Alcivar discloses this limitation.

146. Alcivar discloses a “tissue-conduction microphone [] positioned in contact with the user’s neck tissue in the vicinity of the larynx” for use in voice communication systems to detect voicing activity and control a voice-activated switch. (Ex. 1005 (Alcivar), Abstract; *id.*, 1:5-8, 1:21-28, 1:31-33, 4:65-5:12.) Alcivar’s tissue-conduction microphone detects “tissue vibrations due to modulation of the vocal chords (speech).” (*Id.*, 4:65-5:6.) Alcivar’s tissue-conduction microphone is therefore a sensor that receives human tissue vibration associated with human voicing activity.

147. Alcivar discloses that its system uses a tissue-conduction microphone to detect voicing activity and “generate a transmitter keying signal” to control a voice-activated transmit/receive switch. (Ex. 1005 (Alcivar), 4:65-68.) The tissue-conduction microphone is attached to a system so that, when the user “speaks, the throat microphone [] detects the speech signal which after amplification and detection turns OFF the receiver power switch [] and then turns ON the transmitter power switch.” (*Id.*, 8:4-7.) “The transmitter is turned OFF and the receiver is turned ON again as soon as the diver stops speaking.” (*Id.*, 8:11-16.) Thus,

Alcivar's tissue-conduction microphone is used to determine when a voice signal is present or absent, and can be used to determine when a voice signal is likely greater or smaller than a noise signal (e.g., Alcivar's tissue-conduction microphone and system is a VAD). Accordingly, a person of ordinary skill in the art would have understood that Alcivar's tissue-conduction microphone could be used instead of Ikeda's signal-to-noise power ratio estimator, and optionally instead of Ikeda's step size output circuit, to control the generation of Ikeda's transfer functions. (¶¶59-79 (Ikeda and Puthuff, claim element 11[b]), above.)

i. Motivation to Combine Ikeda and Alcivar

148. A person of ordinary skill in the art would have been motivated to substitute Alcivar's tissue-conduction microphone and VAD for Ikeda's acoustic microphone-based VAD for several reasons.

149. First, Ikeda and Alcivar address the same problem—accurately detecting voicing activity in noisy environments. (Ex. 1003 (Ikeda), 3:32-61; Ex. 1005 (Alcivar), Abstract, 1:5-8, 1:25-27, 1:55-62.)

150. Second, Alcivar suggests the combination. Ikeda discloses the need for accurately determining periods when a desired signal (e.g., voicing activity) is present and periods when a desired signal is absent for an adaptive noise canceler to function optimally. (Ex. 1003 (Ikeda), 3:32-54.) A person of ordinary skill in the

art would have been motivated to use a tissue-conduction microphone or other vibration sensor, as disclosed in Alcivar, instead of Ikeda's method of detecting voicing activity because, as Alcivar explains, such a system can "discriminate against unwanted noises on the basis of amplitude alone, without having to resort to more elaborate signal processing techniques." (Ex. 1005 (Alcivar), 5:8-12.) A person of ordinary skill in the art would have understood from this disclosure in Alcivar that a tissue-conduction microphone (e.g., vibration-based) VAD would be a cheaper, simpler option instead of Ikeda's acoustic microphone-based VAD for determining periods when voicing activity is present or absent in high-noise environments.

151. Third, Ikeda expresses the need for accuracy when determining periods of voicing activity and periods when voicing activity is absent in adaptive noise cancelers. (Ex. 1003 (Ikeda), 3:32-54.) Alcivar discloses that its tissue-conduction microphone is "particularly suited for high noise environment communication systems" and in contrast to acoustic microphone-based systems that are "highly susceptible" to "non-speech noises. (Ex. 1005 (Alcivar), 1:25-27, 1:5-20; *id.*, Abstract.) Thus, a POSITA would have been motivated to use Alcivar's vibration-based VAD with Ikeda's system as a sensor for detecting voice signals to determine the presence or absence of voicing activity because Alcivar's tissue-conduction

microphone would improve the accuracy of Ikeda's system in high-noise environments.

152. Fourth, the use of vibration sensors to detect the presence and absence of voicing activity was well known. Thus, this combination represents nothing more than the simple addition of one known element (a tissue-conduction microphone and VAD as in Alcivar) to another known element (an acoustic microphone array and adaptive noise canceler as in Ikeda) to obtain the predictable results of improved voice activity detection in noisy environments. Moreover, the combination represents the use of a known technique (the incorporation of a tissue-conduction microphone and VAD as in Alcivar) to improve a similar device and method (Ikeda's system) in the same way (improved performance in noisy environments). Finally, the combination applies a known technique (the incorporation of a tissue-conduction microphone and VAD as in Alcivar) to a known device and method (Ikeda's system) that is ready for improvement and yields predictable results (a system with improved voice detection).

153. Alcivar discloses that its system "can be used in other communication systems[.]" (Ex. 1005 (Alcivar), 1:55-62; *id.*, 1:5-8, 1:34-36.) Thus, a person of ordinary skill in the art would have reasonably expected success in combining Alcivar's tissue-conduction microphone and VAD with Ikeda's system.

154. A person of ordinary skill in the art would also have expected success in combining Ikeda's noise canceler with a tissue-conduction, vibration-based microphone and VAD for detecting voice signals because the signal processing of such an adaptive noise canceler does not depend on the type of voice signal detector that is used as long as the sensor accurately detects periods of voicing activity and periods when voicing activity is absent. Similarly, a tissue-conduction microphone or other vibration sensor would detect voice signals in a predictable manner regardless of how such detected signals were used after being detected.

2. Claim 12

155. Claim 12 depends from claim 11 and further recites "wherein the sensor includes a mechanical sensor in contact with the skin." Alcivar discloses the additional limitation of claim 12 because, in addition to the reasons discussed above, it discloses a tissue-conduction microphone placed "in good contact with the [user's] neck tissue" or "in the vicinity of the larynx." (Ex. 1005 (Alcivar), 5:16-17, 4:67-5:8; ¶¶145-154 (Ikeda and Alcivar claim element 11[d]), above.) A person of ordinary skill in the art would have understood that Alcivar's tissue-conduction microphone is a mechanical sensor at least because it mechanically detects vibrations in the user's skin (e.g., through physical contact with the skin).

3. Claim 13

156. Claim 13 depends from claim 11 and further recites “wherein the sensor includes at least one of an accelerometer, a skin surface microphone in physical contact with skin of a user, a human tissue vibration detector, a radio frequency (RF) vibration detector, and a laser vibration detector.” Alcivar discloses this limitation because a person of ordinary skill in the art would have understood that Alcivar’s throat-conduction microphone is a skin-surface microphone in physical contact with skin of a user and/or human tissue vibration detector. (¶¶145-154 (Ikeda and Alcivar claim element 11[d]), above.)

4. Claim 14

157. Claim 14 depends from claim 11 and further recites “wherein the human tissue is at least one of on a surface of a head, near the surface of the head, on a surface of a neck, near the surface of the neck, on a surface of a chest, and near the surface of the chest.” Alcivar discloses this limitation because it discloses sensing tissue vibrations on the surface and near the surface of the neck as discussed above. (¶¶84-98 (Ikeda and Alcivar claim element 11[d]), 100 (Ikeda and Alcivar claim 13), above.) Additionally, Alcivar discloses this limitation because Alcivar discloses sensing vibrations using a tissue-conduction microphone “placed against the chest cavity” of the user. (Ex. 1005 (Alcivar), 5:6-8.)

5. Claim 16

158. Claim 16 depends from claim 11 and further recites that the receiver “includes a plurality of independently located microphones.” Ikeda discloses the additional limitation of claim 16 for the same reasons discussed above. (¶¶56-58 (Ikeda and Puthuff claim element 11[a]), above.)

6. Claim 18

159. Independent claim 18 recites virtually identical limitations to those addressed above for claim 11, including a processor that generates first and second transfer functions, a receiver, microphones, and a voicing sensor, as well as removing acoustic noise from the acoustic signals.

a. 18: Preamble

160. Claim 18’s preamble recites “[a] signal processing system coupled among a user and an electronic device, wherein the signal processing system includes a denoising subsystem for removing acoustic noise from acoustic signals.” To the extent the preamble is limiting, Ikeda discloses it for the reasons discussed above. (¶55 (Ikeda and Puthuff claim 11 preamble), above.)

b. 18[a]: Processor Coupled Among a Receiver and Sensor

161. Claim 18 recites “a processor coupled among a receiver and at least one sensor, wherein the receiver is coupled to receive the acoustic signals via at least two

microphones.” Ikeda discloses and/or renders obvious this limitation for the reasons discussed above. (¶¶56-79 (Ikeda and Puthuff claim elements 11[a] and 11[b]), above.)

c. 18[b]: Voicing Sensor

162. Claim 18 recites “wherein the at least one sensor detects human tissue vibration associated with human voicing activity of a user.” Alcivar discloses this limitation, and a person of ordinary skill in the art would have been motivated to combine Ikeda and Alcivar for the reasons discussed above. (¶¶145-154 (Ikeda and Alcivar claim element 11[d]), above.)

d. 18[c]: Processor That Generates Transfer Functions

163. Claim 18 recites “wherein the processor generates a plurality of transfer functions.” Ikeda discloses and/or renders obvious this limitation for the reasons discussed above. (¶¶59-79 (Ikeda and Puthuff claim element 11[b]), above.)

i. 18[c][1]: First Transfer Function

164. Claim 18 recites “wherein a first transfer function representative of a ratio of acoustic energy received by the two microphones is generated in response to a determination that voicing activity is absent from the acoustic signals for a specified period of time.” Ikeda discloses and/or renders obvious this limitation for

the reasons discussed above. (¶¶65-74 (Ikeda and Puthuff claim element 11[b][1]), above.)

ii. 18[c][2]: Second Transfer Function

165. Claim 18 recites “wherein a second transfer function representative of the acoustic signals is generated in response to a determination that voicing activity is present in the acoustic signals for a specified period of time.” Ikeda discloses and/or renders obvious this limitation for the reasons discussed above. (¶¶75-79 (Ikeda and Puthuff claim element 11[b][2]), above.)

e. 18[d]: Noise Removal

166. Claim 18 recites “wherein acoustic noise is removed from the acoustic signals using the first transfer function and at least one combination of the first transfer function and the second transfer function to produce a denoised acoustic data stream.” Ikeda discloses and/or renders obvious this limitation. (¶¶80-83 (Ikeda and Puthuff claim element 11[c]), above.)

7. Claim 19

167. Claim 19 depends from claim 18 and further recites “wherein the at least one electronic device includes at least one of cellular telephones, personal digital assistants, portable communication devices, computers, video cameras, digital cameras, and telematics systems.” Ikeda discloses and/or renders obvious the

additional limitation of claim 19 for the reasons discussed above. (¶¶111-112 (Ikeda and Puthuff claim 19), above.)

8. Claim 20

168. Claim 20 depends from claim 18 and further recites “wherein the human tissue is at least one of on a surface of a head, near the surface of the head, on a surface of a neck, near the surface of the neck, on a surface of a chest, and near the surface of the chest.” Alcivar discloses the additional limitation of claim 20 for the reasons discussed above. (¶157 (Ikeda and Alcivar claim 14), above.)

9. Claim 1

169. Independent claim 1 is a method claim that recites several functions addressed in claim 11, including receiving acoustic signals, receiving a voice activity signal, and removing noise. These limitations are disclosed by the prior art for the reasons discussed above. (¶¶139-154 (Ikeda and Alcivar, claim 11), above.)

a. 1: Preamble

170. Claim 1 recites “[a] method for removing noise from acoustic signals.” To the extent the preamble is limiting, Ikeda discloses such a method because it discloses the use of a “noise canceler.” (Ex. 1003 (Ikeda), 1:4-19; ¶55 (Ikeda and Puthuff claim 11 preamble), above.)

b. 1[a]: Receiving Acoustic Signals

171. Claim 1 recites “receiving at least two acoustic signals using at least two acoustic microphones positioned in a plurality of locations.” Ikeda discloses this limitation for the reasons discussed above. (¶¶56-58 (Ikeda and Puthuff claim element 11[a]), above.)

c. 1[b]: Sensing Vibrations

172. Claim 1 recites “receiving a voice activity signal that includes information on vibration of human tissue associated with human voicing activity of a user.” Alcivar discloses this limitation, and a person of ordinary skill in the art would have been motivated to combine its disclosure with Ikeda, for the reasons discussed above. (¶¶145-154 (Ikeda and Alcivar claim element 11[d]), above.)

d. 1[c]: Generating a VAD Signal

173. Claim 1 recites “generating a voice activity detection (VAD) signal using the voice activity signal.” Alcivar discloses and/or renders obvious this limitation.

174. As discussed above, Alcivar discloses that its tissue-conducting microphone forms part of a VAD system. (¶¶145-154 (Ikeda and Alcivar claim element 11[d]), above.) Alcivar’s system receives voice activity signals using the tissue-conducting microphone and uses them to generate a “transmitter keying signal” (e.g., a VAD signal). (Ex. 1005 (Alcivar), 4:65-67.) Alcivar discloses that

the output of its tissue-conducting microphone is further processed, including using a “Schmitt trigger circuit 68,” which “provide[s] sharply defined ON/OFF levels and some degree of immunity to low-level random noise inputs.” (*Id.*, 5:55-67.) The signal is then sent to “audio detector 62.” (*Id.*, 6:1-11.) A person of ordinary skill in the art would have understood that the output of Alcivar’s audio detector is a VAD signal at least because it provides a “sharply defined” signal that denotes when the user is or is not speaking. (*Id.*, 5:55-67; *id.*, 8:3-16.)

175. Alcivar therefore discloses and/or renders obvious this limitation. Moreover, one of ordinary skill in the art would have been motivated to combine Ikeda and Alcivar for the reasons discussed above. (¶¶145-154 (Ikeda and Alcivar, claim element 11[d]), above.)

e. 1[d]: Generating at Least Two Transfer Functions

176. Claim 1 recites “generating at least two transfer functions representative of a ratio of energy of the acoustic signal received using at least two different acoustic microphones of the at least two acoustic microphones when the VAD indicates that user voicing activity is absent, wherein the at least two transfer functions comprise a first transfer function and a second transfer function.” Ikeda discloses this limitation for the reasons discussed above. (¶¶65-79 (Ikeda and Puthuff claim element 11[b][1]-11[b][2]), above.)

177. As discussed above, Ikeda discloses that a VAD signal is used to indicate when voicing activity is present or absent and to control the generation of Ikeda's transfer functions. (¶¶65-79 (Ikeda and Puthuff claim elements 11[b][1] and 11[b][2]), above.) A person of ordinary skill in the art would have understood that, when combining Alcivar's VAD with Ikeda's system (¶¶145-154 (Ikeda and Alcivar claim element 11[d]), above), Alcivar's VAD would likewise provide a VAD signal that Ikeda's system would use in lieu of Ikeda's VAD signal for the same purpose of controlling the generation of Ikeda's transfer functions. Ikeda and Alcivar therefore disclose this limitation.

f. 1[e]: Removing Noise

178. Claim 1 recites "removing acoustic noise from at least one of the acoustic signals by applying the first transfer function and at least one combination of the first transfer function and the second transfer function to the acoustic signals and generating denoised acoustic signals." Ikeda discloses this limitation as discussed above. (¶¶80-83 (Ikeda and Puthuff claim element 11[c]), above.)

10. Claim 2

179. Claim 2 depends from claim 1 and further recites "wherein removing noise further comprises generating one transfer function of the at least two transfer functions to be representative of a ratio of energy of the acoustic signal received when the VAD indicates that user voice activity is present." Ikeda discloses this

limitation for the reasons discussed above. (§§128 (Ikeda and Puthuff claim element 2[a]), 65-79 (Ikeda and Puthuff claim element 11[b][1]-11[b][2]), above.) Claim 2 also recites “removing noise from the acoustic signals using at least one combination of the at least two transfer functions to generate the denoised acoustic signals.” Ikeda discloses this additional limitation of claim 2 for the same reasons discussed above. (§§129 (Ikeda and Puthuff claim element 2[b]), 80-83 (Ikeda and Puthuff claim element 11[c]), 144 (Ikeda and Alcivar claim element 11[c]).)

11. Claim 3

180. Claim 3 depends from claim 1 and further recites “wherein the acoustic signals include at least one reflection of at least one associated noise source signal and at least one reflection of at least one acoustic source signal.” The additional limitation of claim 3 would have been obvious to a person of ordinary skill in the art in view of Ikeda for the reasons discussed above. (§130 (Ikeda and Puthuff claim 3), above.)

181. Additionally, Alcivar discloses that in the context of an underwater communication system, in which a diver is wearing a mask with communication capabilities, the diver’s mask may have “sharp, well-defined resonances” below a certain acoustic frequency, which causes acoustic microphones to pick up reverberations below that frequency. (Ex. 1005 (Alcivar), 2:9-12.) Alcivar discloses methods for reducing that reverberation. (*Id.*, 2:9-22.) A person of ordinary skill in

the art would have understood that Ikeda's system would also have worked to reduce the reverberations created by Alcivar's diving mask. Thus, Alcivar discloses and/or renders obvious this limitation.

12. Claim 4

182. Claim 4 depends from claim 1 and further recites "wherein generating the at least two transfer functions comprises recalculating the at least two transfer functions during at least one prespecified interval." Ikeda discloses the additional limitation of claim 4 for the reasons discussed above. (¶131 (Ikeda and Puthuff claim 4), above.)

13. Claim 5

183. Claim 5 depends from claim 1 and further recites "wherein generating the at least two transfer functions comprises use of at least one technique selected from a group consisting of adaptive techniques and recursive techniques." Ikeda discloses and/or renders obvious the additional limitation of claim 5. (¶¶132-134 (Ikeda and Puthuff claim 5), above.)

14. Claim 6

184. Claim 6 depends from claim 1 and further recites "wherein information on the vibration of human tissue is provided by a sensor in contact with the skin." Alcivar discloses this limitation for the same reasons discussed above. (¶155 (Ikeda and Alcivar claim 12), above.)

15. Claim 7

185. Claim 7 depends from claim 1 and further recites “wherein information on the vibration of human tissue is provided via at least one sensor selected from among at least one of an accelerometer, a skin surface microphone in physical contact with skin of a user, a human tissue vibration detector, a radio frequency (RF) vibration detector, and a laser vibration detector.” Alcivar discloses the additional limitation of claim 7 for the reasons discussed above. (¶156 (Ikeda and Alcivar claim 13), above.)

16. Claim 8

186. Claim 8 depends from claim 1 and further recites “wherein the human tissue is at least one of on a surface of a head, near the surface of the head, on a surface of a neck, near the surface of the neck, on a surface of a chest, and near the surface of the chest.” Alcivar discloses the additional limitation of claim 8 for the reasons discussed above. (¶157 (Ikeda and Alcivar claim 14), above.)

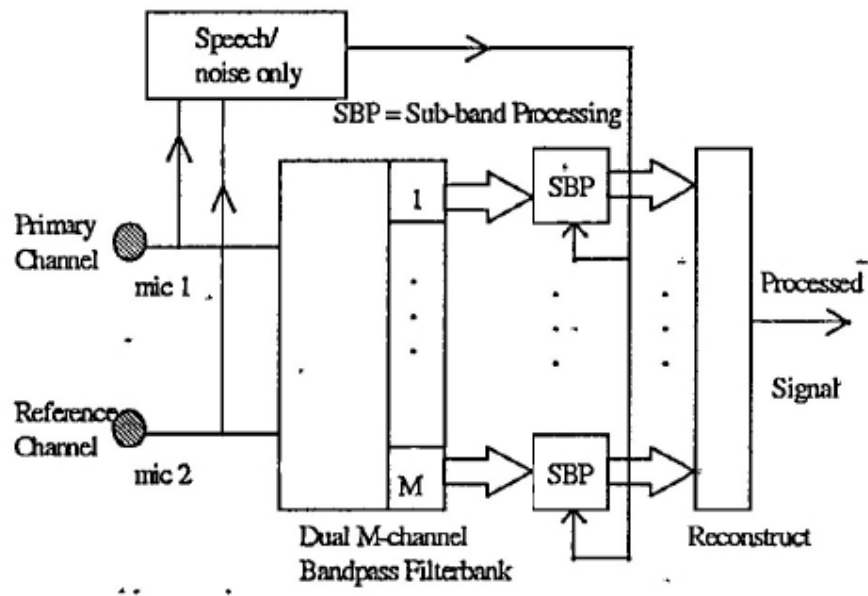
C. Claims 3, 9, and 15 Would Have Been Obvious in View of Ikeda and Either of Puthuff or Alcivar Further in View of Hussain.

187. The combination of either Ikeda and Puthuff or Ikeda and Alcivar, further in view of Hussain, renders claims 3, 9, and 15 obvious. Each of these claims depend from a claim discussed above; the limitations of each parent claim are

obvious for the reasons discussed above. (¶¶54-98, 115-127 (Ikeda and Puthuff claims 1 and 11), 139-154, 169-178 (Ikeda and Alcivar claims 1 and 11), above.)

188. Hussain discloses an “adaptive noise cancellation scheme” that improves performance in reverberant environments by processing acoustic signals in separate sub-bands to produce a denoised signal. (Ex. 1007 (Hussain), 2611-2613, Fig. 1.)

189. Like Ikeda, Hussain uses a dual microphone array. (*Id.*, 2611-2613, Fig. 1.) A filter bank separates signals received by the microphones into individual sub-bands for processing. (*Id.*) Hussain then uses “an adaptive filter” to determine a “differential acoustic-path transfer function” for individual sub-bands. (*Id.*) When voicing activity is detected, the transfer function is used “to process the noisy speech signal” and output a denoised speech signal. (*Id.*) Hussain applies sub-band processing (“SBP”) to each sub-band, reconstructs the processed signals, and outputs the denoised processed signal, as shown below.



(Id., Fig. 1.)

1. Claim 3

190. Claim 3 depends from claim 1 and further recites “wherein the acoustic signals include at least one reflection of at least one associated noise source signal and at least one reflection of at least one acoustic source signal.” The additional limitation of claim 3 would have been obvious for the reasons discussed above. (¶¶130 (Ikeda and Puthuff claim 3), 180-181 (Ikeda and Alcivar claim 3), above.)

191. Hussain also discloses this limitation. Hussain describes “results of comparative experiments with simulated anechoic data extended to include simulated reverberant data.” (Ex. 1007 (Hussain), 2611.) The signal processing method disclosed in Hussain addresses the well-known problem of noise in, for example, “hands-free telephones [and] portable phones” which includes “the

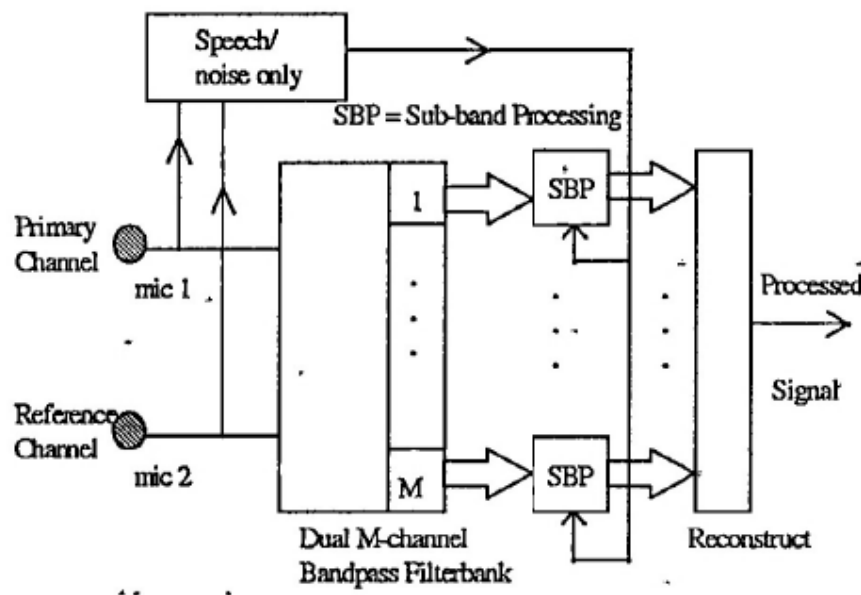
presence of multiple noise sources [and] reverberant environments.” (*Id.*) Hussain’s sub-band processing (described below) improves noise cancellation “in reverberant environments” where “[m]ulti-band processing has been found to be important in combating reverberation effects.” (*Id.*)

192. Thus, Hussain discloses an environment matching the additional limitation of claim 3 and discloses the use of devices like Ikeda’s (as modified by Puthuff or Alcivar) in such an environment. It also explains why a person of ordinary skill in the art would have been motivated to use a device like Ikeda’s in such a reverberant environment (e.g., because such reverberant environments were where people used hands-free telephones and portable phones). Finally, Hussain discloses that a person of ordinary skill in the art would have reasonably expected success in such an environment when using a system like Ikeda’s as modified by Hussain’s teachings. Additionally, one of ordinary skill in the art would have been motivated to combine Hussain with Ikeda, and would have reasonably expected success in doing so, for the reasons discussed below. (¶¶199-203 (Hussain motivation to combine), below.)

2. Claims 9 and 15

a. 9[a] and 15[a]: Dividing Acoustic Data into Sub-bands

193. Claims 9 and 15 depend from claims 1 and 11, respectively, and further recite “dividing acoustic data of the acoustic signals into a plurality of subbands.” Hussain discloses “[a] multi-microphone adaptive speech enhancement system employing diverse subband processing” in which acoustic signals are dividing into “a set of sub-bands provided by a filter bank as shown in Figure 1”:



(Ex. 1007 (Hussain), 2611, Fig. 1.) Thus, Hussain discloses this limitation.

194. Additionally, Puthuff discloses that acoustic signals can be “split into more than two bands” using digital signal processing so that only noisy sub-bands are affected by signal processing. (Ex. 1004 (Puthuff), 16:4-15.) With this implementation, “[t]he digital signal processing . . . permits only the affected band,

or bands, to responsively fade over to the bone conduction signal, without fading other, unaffected bands, thus maintaining the microphone path signal for the band signals where the microphone signal is the cleaner signal.” (*Id.*, 16:21-25.) Thus, Puthuff also discloses this limitation.

b. 9[b] and 15[b]: Generating a Sub-band Transfer Function

195. Claims 9 and 15 further recite “generating a subband transfer function representative of the ratio of acoustic energies received in each microphone in each subband.” Ikeda discloses generating transfer functions representative of a ratio of acoustic energies received in each microphone, as discussed above. (¶¶65-79 (Ikeda and Puthuff claim elements 11[b][1]-[2]).)

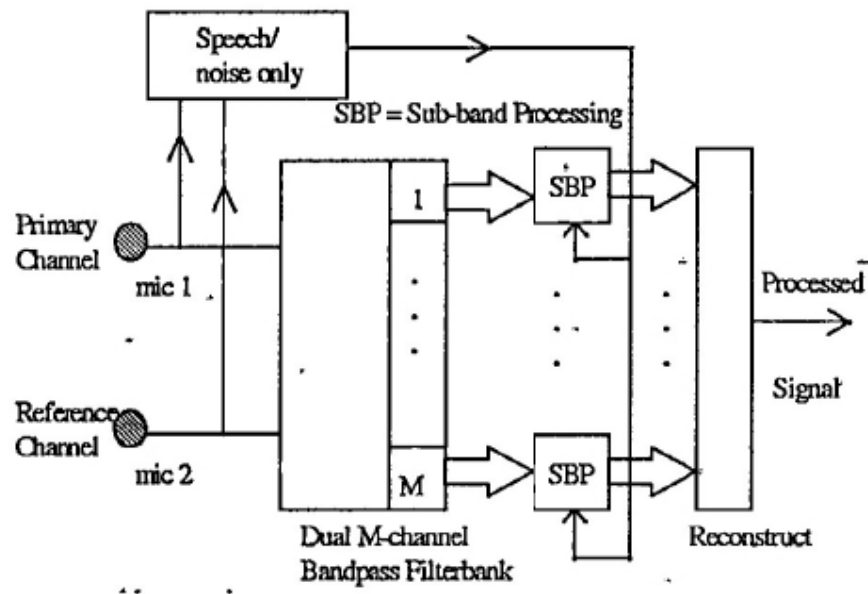
196. Additionally, Hussain discloses a “Multi-Microphone sub-band Adaptive (MMSBA) speech enhancement system” that identifies “a differential acoustic path transfer function during a noise only period in intermittent speech” “within a set of sub-bands provided by a filter bank.” (Ex. 1007 (Hussain), 2611.) Accordingly, Hussain also discloses this limitation. Furthermore, it would have been obvious to generate Ikeda’s transfer functions using Hussain’s sub-band processing methods. Thus, this limitation would have been obvious in view of Ikeda and Hussain.

c. 9[c] and 15[c]: Removing Acoustic Noise

197. Claims 9 and 15 further recite “removing acoustic noise from each of the plurality of subbands using the subband transfer function, wherein a plurality of denoised acoustic subband signals are generated.” Hussain discloses removing acoustic noise from sub-bands using sub-band transfer functions through an “Intermittent coherent noise canceller” function that “models the differential acoustic path transfer function between the microphones” and applies the transfer function to result in an “output E [that is] a noise reduced, filtered version of the sub-band speech signal.” (Ex. 1007 (Hussain), 2612.) Thus, Hussain discloses this limitation.

d. 9[d] and 15[d]: Combining Sub-band Signals

198. Claims 9 and 15 further recite “combining the plurality of denoised acoustic subband signals to generate the denoised acoustic signal.” Hussain discloses combining the plurality of denoised acoustic sub-band signals output by sub-band processing circuits, indicated as “SBP” blocks in Figure 1, by combining them in a “reconstruct” circuit resulting in a single denoised “processed signal”:



(Ex. 1007 (Hussain), Fig. 1.) Thus, Hussain discloses this limitation.

3. Motivation to Combine

199. A person of ordinary skill in the art would have been motivated to combine the teachings of Ikeda (including as modified by Puthuff or Alcivar) and Hussain for several reasons.

200. First, Ikeda and Hussain both address the same problem—removing noise from a desired speech signal using adaptive methods. (Ex. 1003 (Ikeda), 3:32-60, 4:6-25; Ex. 1007 (Hussain), 2611.)

201. Second, Hussain suggests the combination. Hussain discloses that operating on individual sub-bands “combat[s] reverberation effects” and produces “more than 6dB SNR improvements over conventional full-band methods in reverberant environments.” (Ex. 1007 (Hussain), 2611.) Thus, a person of ordinary

skill in the art would have been motivated to implement Ikeda's adaptive noise cancellation methods on individual sub-bands as in Hussain for improved performance.

202. Third, Hussain discloses sub-band processing's improved performance over full-band noise cancellation systems that "model acoustic path transfer functions." (Ex. 1007 (Hussain), 2611.) Thus, combining Hussain and Ikeda to perform Ikeda's noise cancellation on individual sub-bands would have been nothing more than combining prior art elements (Ikeda's transfer function-based noise cancellation approach and Hussain's sub-band processing) according to known methods to yield the predictable results of improved noise cancellation performance in reverberant environments. Moreover, the combination represents the use of a known technique (sub-band processing as in Hussain) to improve a similar device and method (Ikeda's system) in the same way. Finally, the combination applies a known technique (sub-band processing as in Hussain) to a known device and method (Ikeda's system) that is ready for improvement and yields predictable results (improved noise cancellation performance in reverberant environments).

203. A person of ordinary skill in the art would have had a reasonable expectation of success in combining Ikeda and Hussain because they relate to similar systems that perform adaptive noise cancellation. Moreover, performing adaptive

noise cancellation on individual sub-bands was a well-understood technique that would perform predictably in an adaptive noise reduction system such as Ikeda's.

D. Claims 10 and 17 Would Have Been Obvious in View of Ikeda, Puthuff or Alcivar, and Sasaki.

204. The combination of either Ikeda and Puthuff or Ikeda and Alcivar, further in view of Sasaki, renders claims 10 and 17 obvious. Each of these claims depend from a claim discussed above; the limitations of each parent claim are obvious for the reasons discussed above. (¶¶54-98, 115-127 (Ikeda and Puthuff claims 1 and 11), 139-154, 169-178 (Ikeda and Alcivar claims 1 and 11), above.)

1. Claims 10 and 17

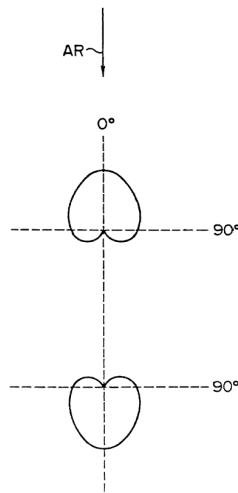
205. Claims 10 and 17 depend from claims 1 and 11, respectively, and further recite “wherein the at least two acoustic microphones comprise a first directional acoustic microphone and a second directional acoustic microphone, wherein the first directional acoustic microphone and the second directional acoustic microphone selectively attenuate the acoustic signals based on the direction of arrival.” Sasaki discloses this limitation.

206. Sasaki describes a system with a microphone array in which “main input microphone 11 is unidirectional and placed with its most sensible side in the front direction” while “reference input microphone [21] is also unidirectional and is

placed with its most sensible side in the rear direction.” (Ex. 1008 (Sasaki), 6:6-20.)

Figure 6 shows the two microphones’ sensitivity patterns:

F I G. 6



(*Id.*, Fig. 6.) As shown, “the reference input microphone 21 has a low sensitivity in the desired sound arrival direction and a high sensitivity in the rear direction or noise arrival direction” while the main input microphone has a high sensitivity in the direction of the desired sound. (*Id.*, 6:18-20.) Thus, the two microphones are directional acoustic microphones that selectively attenuate acoustic signals based on the direction of arrival. Accordingly, Sasaki discloses this limitation.

2. Motivation to Combine

207. A person of ordinary skill in the art would have been motivated to combine the teachings of Ikeda (including as modified by Puthuff or Alcivar) and Sasaki, and to use directional microphones in Ikeda’s system, for several reasons.

208. First, Ikeda and Sasaki both address the same problem—removing noise from a desired speech signal using adaptive methods. (Ex. 1003 (Ikeda), 3:32-60, 4:6-25; Ex. 1008 (Sasaki), 3:36-54 (describing an “adaptive noise reduction system”), 4:17-28.)

209. Second, because Ikeda requires a “first microphone 1 for a main signal [and] a second microphone 2 for a reference signal,” but does not disclose a specific type of microphone, a person of ordinary skill in the art would have been motivated to look to Sasaki’s teachings on the use of directional microphones so that Ikeda’s first microphone receives the main signal while the second microphone does not, improving the accuracy of Ikeda’s system. (Ex. 1003 (Ikeda), 4:49-53.)

210. Third, directional microphones were well known and both Ikeda and Sasaki disclose a primary microphone for detecting a desired signal and a reference microphone for detecting primarily noise. (*Id.*, 4:49-53; Ex. 1008 (Sasaki), Abstract, 6:6-20.) For example, I have found in my own work as early as 1993 that the use of directional microphones can substantially improve speech recognition accuracy in noisy environments. (*E.g.*, Thomas M. Sullivan and Richard M. Stern, *Multi-Microphone Correlation-Based Processing for Robust Speech Recognition*, Proc. of the ICASSP ’93 at 91-94.) In the years since I have enjoyed continued success with a variety of multi-microphone techniques. Thus, this combination represents nothing more than the simple addition of one known element (directional

microphones as in Sasaki) to another known element (an acoustic microphone array as in Ikeda) to obtain the predictable results of improved voice activity detection in noisy environments. Moreover, the combination represents the use of a known technique (using directional microphones as in Sasaki) to improve a similar device and method (Ikeda's) in the same way. Finally, the combination applies a known technique (using directional microphones as in Sasaki) to a known device and method (Ikeda's) that is ready for improvement and yields predictable results (a system with improved voice activity detection).

211. A person of ordinary skill in the art would have had a reasonable expectation of success in combining Ikeda and Sasaki because directional microphones were well known in the art and using them to selectively attenuate a desired signal and a reference signal based on direction of arrival would have yielded a predictable result.

E. It Would Have Been Obvious to Generate Transfer Functions Using a Processor in View of the References Discussed Above and Further in View of Bartlett.

212. Each limitation of claims 11-20 is disclosed and/or rendered obvious for the reasons discussed above. (¶¶54-114, 139-168 (claims 11-20), above.) However, it would also have been obvious to use a processor coupled among the

microphones and voicing sensor to generate transfer functions as recited in independent claims 11 and 18 in view of Bartlett.

213. Bartlett discloses a “noise-canceling” system that contains two microphones (30 and 32), an adaptive filter, and a digital signal processor (“DSP”). (Ex. 1006 (Bartlett), 6:66-7:8.) Bartlett performs signal processing (e.g., generating a plurality of transfer functions) through an “adaptive algorithm 120 [which] is implemented by, for example, a digital signal processor.” (*Id.*, 6:66-67.) Accordingly, Bartlett discloses this limitation and/or renders this limitation obvious.

214. A person of ordinary skill in the art would have been motivated to combine the teachings of Ikeda (and Puthuff or Alcivar) with Bartlett, and to use a processor coupled with the acoustic receiver (and voicing sensor) to generate Ikeda’s transfer functions, for several reasons.

215. First, Ikeda and Bartlett address the same problems—accurately detecting speech and removing noise from speech signals using adaptive methods. (Ex. 1003 (Ikeda), Abstract, 1:4-8, 3:56-61; Ex. 1006 (Bartlett), 1:5-10, 1:41-53, 2:14-24, 2:55-63, 4:9-13, 6:60-7:17.)

216. Second, Ikeda discloses general adaptive filters used to generate transfer functions, and Bartlett discloses that an adaptive filter is “exemplarily implemented by a digital signal processor.” (Ex. 1006 (Bartlett), 7:7-8.) Thus,

Bartlett's disclosure would motivate a person of ordinary skill in the art to use a processor for performing signal processing in Ikeda's system.

217. Third, processors were well known to provide fast, robust ways to perform signal processing. Thus, the use of a processor to generate transfer functions in the place of other means would have been nothing more than the simple substitution of one known element (a processor as in Bartlett) for another (another method of generating transfer functions) to obtain predictable results. Moreover, the combination represents the use of a known technique (using a processor to generate a transfer function) to improve a similar device and method (Ikeda's system) in the same way. The combination also applies a known technique (using a processor to generate a transfer function) to a known device and method (Ikeda's system) that is ready for improvement and yields predictable results.

218. Fourth, noise cancellation systems were commonly used with electronic devices that included processors. A person of ordinary skill in the art would have been motivated to generate transfer functions and perform noise reduction with such an already-included processor for a simplified, cheaper, and more robust design.

219. A person of ordinary skill in the art would have had a reasonable expectation of success in combining Ikeda and Bartlett because they relate to similar systems (dual-microphone systems that implement adaptive techniques to remove

noise from a speech signal) and because processors were conventional components for performing signal processing. Using a processor to generate transfer functions would simply be a conventional application of a processor that would yield predictable results.

VII. SECONDARY CONSIDERATIONS OF NONOBVIOUSNESS

220. I am not aware of any secondary considerations of nonobviousness. If Patent Owner identifies any alleged evidence of secondary considerations in the future, I reserve the right to respond to that information.

VIII. CONCLUSION

221. For the foregoing reasons, it is my opinion that claims 1-20 of the '091 patent would have been obvious to a person of ordinary skill in the art at the time of the alleged invention in view of the prior art discussed above.

222. I reserve the right to supplement my opinions in the future to address or respond to any issues that the Patent Owner may raise, as well as new information including, but not limited to, any claim constructions advanced by the Patent Owner or adopted by the Board in the Institution Decision, and respond to any alleged secondary considerations as they become available to me.

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

Executed on November 21, 2022 at Pittsburgh, Pennsylvania.



Richard M. Stern, Ph.D.