

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

META PLATFORMS, INC.,

Petitioner,

v.

JAWBONE INNOVATIONS, LLC,

Patent Owner.

Patent No. 8,467,543

Filing Date: March 27, 2003

Issue Date: June 18, 2013

Inventor: Gregory C. Burnett, Nicolas J. Petit,
Alexander M. Asseily, and Andrew E. Einaudi

Title: MICROPHONE AND VOICE ACTIVITY DETECTION (VAD)
CONFIGURATIONS FOR USE WITH COMMUNICATION SYSTEMS

DECLARATION OF AKBAR M. SAYEED, PH.D.

Case No. IPR2023-01291

TABLE OF CONTENTS

	<u>Page(s)</u>
I. INTRODUCTION	1
A. Background and Qualifications	1
II. COMPENSATION	6
III. MATERIALS CONSIDERED	6
IV. LEGAL PRINCIPLES	7
V. LEVEL OF SKILL IN THE ART	8
VI. THE CLAIMED INVENTION OF THE '543 PATENT	9
VII. CLAIM CONSTRUCTION	13
A. “voice detection subsystem,” “voice activity signals,” and “human voicing activity”	13
B. “generating a noise waveform estimate associated with noise of the acoustic signals”	15
VIII. THE ALLEGED PRIOR ART	21
A. Andrea (U.S. Patent No. 5,732,143)	21
B. Hussain (<i>A New Metric for Selecting Sub-band Processing in Adaptive Speech Enhancement Systems</i>)	23
C. Holzrichter (U.S. Patent No. 6,006,175)	26
D. Stevens (<i>A Miniature Accelerometer for Detecting Glottal Waveforms and Nasalization</i>)	28
E. Alcivar (U.S. Patent No. 3,746,789)	28
IX. PETITIONER’S COMBINATIONS DO NOT RENDER ANY CLAIM OBVIOUS.....	29

A.	GROUND 1: The Combination of Andrea, Hussain, and Holzrichter Does Not Render Obvious Claims 1-2 and 5-26	29
1.	Meta’s Combination Does Not Render Obvious “wherein the denoising method includes generating a noise waveform estimate associated with noise of the acoustic signals,” [1H]	29
i.	The Combination Does Not Render the Limitation Obvious Under the Proper Construction	29
ii.	The Combination Does Not Render the Limitation Obvious Under the Claim Language	32
2.	Meta’s Combination Does Not Disclose or Render Obvious “subtracting the noise waveform estimate from the acoustic signal when the acoustic signal includes speech and noise,” [1H]	34
3.	Meta’s Combination Does Not Disclose or Render Obvious “wherein the voice detection subsystem is configured to receive the voice activity signals using a sensor independent from the microphone array and to output the control signals generated from the voice activity signals to the denoising system” [1I]	37
i.	Andrea Does Not Teach This Limitation	37
ii.	Holzrichter Does Not Teach or Render Obvious This Limitation	38
B.	GROUND 2-3: These Grounds Fail Based on the Failure of Ground 1	42
X.	CONCLUSION.....	42

I, Akbar M. Sayeed, declare as follows:

1. I have been asked by counsel for Patent Owner, Jawbone Innovations, LLC (“Jawbone” or “Patent Owner”), to review U.S. Patent No. 8,467,543 (the “’543 Patent”) entitled MICROPHONE AND VOICE ACTIVITY DETECTION (VAD) CONFIGURATIONS FOR USE WITH COMMUNICATION SYSTEMS, and to provide my technical review, analysis, insights, and opinions regarding the ’543 Patent in view of the prior art cited by Petitioner Meta Platforms, Inc. (“Meta” or “Petitioner”). I submit this Declaration in support of Patent Owner’s Response in this IPR proceeding. I have personal knowledge of the matters stated herein and would be competent to testify to them if required.

2. I have been retained on behalf of Jawbone for the above-captioned *inter partes* review proceeding. I understand that the ’543 Patent is currently assigned to Jawbone.

3. I am over 18 years of age. I have personal knowledge of the facts stated in this Declaration and could testify competently if asked to do so.

I. INTRODUCTION

A. Background and Qualifications

4. I have reviewed and am familiar with the specification of the ’543 Patent. I understand the ’543 Patent has been provided as Exhibit 1001. I will cite to the specification using the following format: ’543 Patent at col.:line.

5. My CV is being submitted simultaneously herewith (Ex. 2004).

6. I received my B.S. degree in Electrical and Computer Engineering (ECE) from the University of Wisconsin-Madison in 1991. I received my M.S. and Ph.D. degrees in ECE from the University of Illinois at Urbana-Champaign in 1993 and 1996, respectively. In my undergraduate and graduate studies, I took a variety of courses generally focusing on the areas of signal processing, communication theory, information theory, electromagnetics and antennas, and statistical techniques in signal processing and communications. My Ph.D. dissertation was entitled “Statistical Time-Frequency Analysis” in which I proposed a new framework for statistical signal processing using time-frequency representations, mathematical tools that extend the powerful theory of Fourier transforms, and Fourier analysis to time-varying signals and systems.

7. I currently work as an Independent Researcher, Engineer, and Technical Consultant, and worked as professor of Electrical and Computer Engineering at the University of Wisconsin-Madison from 1997-2021, where I directed the Wireless Communications and Sensing Laboratory until my retirement on August 1, 2021 to pursue a career as an independent researcher and consultant.

8. My current work as an independent researcher and consultant spans STEM (science, technology, engineering & mathematics) fields through the lens of

information science and technology. My STEM skills and experience cover a broad range of areas, including:

- Wireless technologies for sensing, processing, and communication of information (5G/6G/XG), internet of things, and emerging technologies for untapped spectrum (*e.g.*, millimeter-wave, THz).
- Sensing and acquisition of data in new modalities through technological innovation.
- Machine learning and statistical techniques for extracting useful information from data.
- Quantum information science and technology (computing, sensing, and communication) spanning foundational concepts, algorithms, and platforms.
- Conception, design, and development of new information technologies through basic theory, computational modeling, and hybrid software-hardware prototyping and experimentation.

9. I have written/co-written 200+ papers in leading journals and conferences. I am an inventor/co-inventor of 10 patents.

10. After receiving my Ph.D. in 1996, I spent a year at Rice University as a postdoctoral research fellow where I expanded the scope of my research to the growing field of wireless communications that underpins much of mobile cellular

technology. In particular, the propagation channels encountered in mobile cellular communication are inherently time-varying in nature and that provided a natural connection to my Ph.D. research. After spending a year at Rice University, I accepted a position as an Assistant Professor in the ECE Department at the University of Wisconsin-Madison in 1997. I was granted tenure and promoted to Associate Professor in 2003, and was promoted to Professor in 2008.

11. I founded and directed the Wireless Communication and Sensing Laboratory at UW-Madison and engaged in research, teaching, and innovation in the basic science, engineering, and technology of sensing, processing, and communication of information. In particular, I led a team in the conception, design, development, and successful demonstration of a state-of-the-art prototype for multi-beam MIMO (multiple input multiple output) communication at 28 GHz based on the concept of a lens array that I had pioneered.

12. I co-founded and co-led the NSF-sponsored Research Coordination Network (RCN) on millimeter-wave wireless (2016-2019) for bringing together researchers from academia, industry, and government agencies to address critical technical challenges in the area.

13. From 2017-2019 I served as a Program director in the Electrical, Communications, and Cyber Systems (ECCS) Division of the Directorate of Engineering at the National Science Foundation. At the NSF, I managed existing

programs and initiated new research programs involving sensing, processing, and communication of information in all possible physical modalities, architectures, and technologies, including quantum information science and engineering. I actively contributed to three NSF-wide working groups: Electromagnetic Spectrum Management (ESM) working group, and the working groups of two of the 10 NSF Big Ideas: i) Quantum Leap and ii) Harnessing the Data Revolution. In particular, I co-led the team that developed the NSF solicitation for Quantum Leap Challenge Institutes (Feb 2019) in response to the National Quantum Initiative (NQI) Act signed in December 2018.

14. I have a demonstrated history (25 yrs.) of leading an interdisciplinary research and development group at the forefront of foundational and technological innovations in wireless communication and sensing. Physics-based accurate and computationally tractable system modeling has been a key underlying theme in my work, spanning communication and sensor networks, multipath propagation over highly dynamic environments, wideband MIMO systems, mm-wave and THz wireless, and prototype development (hardware & software).

15. My technical background, experience and skills are directly relevant to the subject matter in this case. The modeling, analysis, and processing of information bearing signals and waves, including transmission and reception through antennas and antenna arrays, and propagation over channels, has been an

underlying theme throughout my career. In particular, I have extensive experience in the conception, modeling, development, and analysis of techniques and algorithms for processing and filtering of signals from an array of sensors or antennas, including beamforming, interference and noise suppression, and adaptive filtering methods. While most of my work has focused on electromagnetic signals, the underlying physics and mathematics are very similar to acoustic signals.

II. COMPENSATION

16. My compensation for time worked on this proceeding is not dependent on any issues related to the '543 Patent, the outcome of this proceeding, or the substance of my opinions. My compensation for time worked on this proceeding is at my customary rate of \$200 per hour. I have no financial interest in, or affiliation with, the Patent Owner or any of the real parties in interest.

III. MATERIALS CONSIDERED

17. In providing my technical review, analysis, insights, and opinions, I have considered the '543 Patent and its prosecution history.

18. I have also considered the Petition filed by the Petitioner in this proceeding and the relevant exhibits relied on by Petitioner, including the expert declaration submitted by David V. Anderson, Ph.D. ("Ex. 1003"), and Dr. Anderson's deposition testimony (Ex. 2001, 2005).

19. I have also considered my own experience and knowledge, as discussed above and described more fully in my CV, in areas including signal processing, electrical engineering, and other related technologies.

IV. LEGAL PRINCIPLES

20. I understand that a patent claim is unpatentable as “obvious” if the subject matter of the claim as a whole would have been obvious to a person of ordinary skill in the art (POSITA) as of the time of the invention at issue.

21. I understand that the use of “the person of ordinary skill” rubric is to prevent one from improperly, in the present day, using hindsight to decide whether a claim is obvious.

22. I understand that the following factors must be evaluated to determine whether the claimed subject matter is obvious: (1) the scope and content of the prior art; (2) the difference or differences, if any, between the scope of the patent claim and the scope of the prior art; and (3) the level of ordinary skill in the art at the time of the invention.

23. I understand that certain secondary considerations, such as commercial success, skepticism of experts, surprise, and copying, may provide evidence of non-obviousness. I further understand that such considerations are often the most probative and determinative of obviousness or non-obviousness.

24. I understand that I must construe a claim in accordance with the ordinary and customary meaning of the language of such claim as understood by one of ordinary skill in the art and the prosecution history pertaining to the patent.

V. LEVEL OF SKILL IN THE ART

25. I understand that I should perform my analysis from the viewpoint of a person of ordinary skill in the art. I understand that this hypothetical person of ordinary skill in the art is considered to have the normal skills of a person in a certain technical field. I understand that factors that may be considered in determining the level of ordinary skill in the art include: (1) the education level of the inventor; (2) the types of problems encountered in the art; (3) the prior art solutions to those problems; (4) rapidity with which innovations are made; (5) the sophistication of the technology; and (6) the education level of active workers in the field.

26. In my opinion, a person of ordinary skill in the art would have a bachelor's degree in computer engineering, computer science, electrical engineering, mechanical engineering, or a similar field, and approximately two years of industry or academic experience in a field related to acoustics, speech recognition, speech detection, or signal processing. Work experience can substitute for formal education and additional formal education can substitute for work experience.

VI. THE CLAIMED INVENTION OF THE '543 PATENT

27. I have reviewed the '543 Patent, entitled MICROPHONE AND VOICE ACTIVITY DETECTION (VAD) CONFIGURATIONS FOR USE WITH COMMUNICATION SYSTEMS. It was issued on June 18, 2013. The '543 Patent was filed on March 27, 2003.

28. The '543 Patent claims systems for performing acoustic noise suppression in a multiple microphone system with the use of a voice activity detector. '543 Patent, Abstract. Embodiments of the system use two microphones and a voice activity detection (VAD) system, which includes both a VAD device and a VAD algorithm. '543 Patent, 5:45-50. The signals from the primary and secondary microphones are filtered into multiple frequency subbands, and each subband is denoised using its own adaptive filter. *Id.*, 16:31-36. The VAD system uses algorithms for processing the voice activity signals from the VAD device and outputs control signals used to adapt the noise suppression system to the received signals. *Id.*, 16:46-54, 19:53-64. The control signals can be used to determine when to adapt the transfer functions used in the denoising process to the presence or absence of speech. *Id.*, 14:51-52, 16:63-67.

29. The '543 Patent describes the embodiment of FIG. 1:

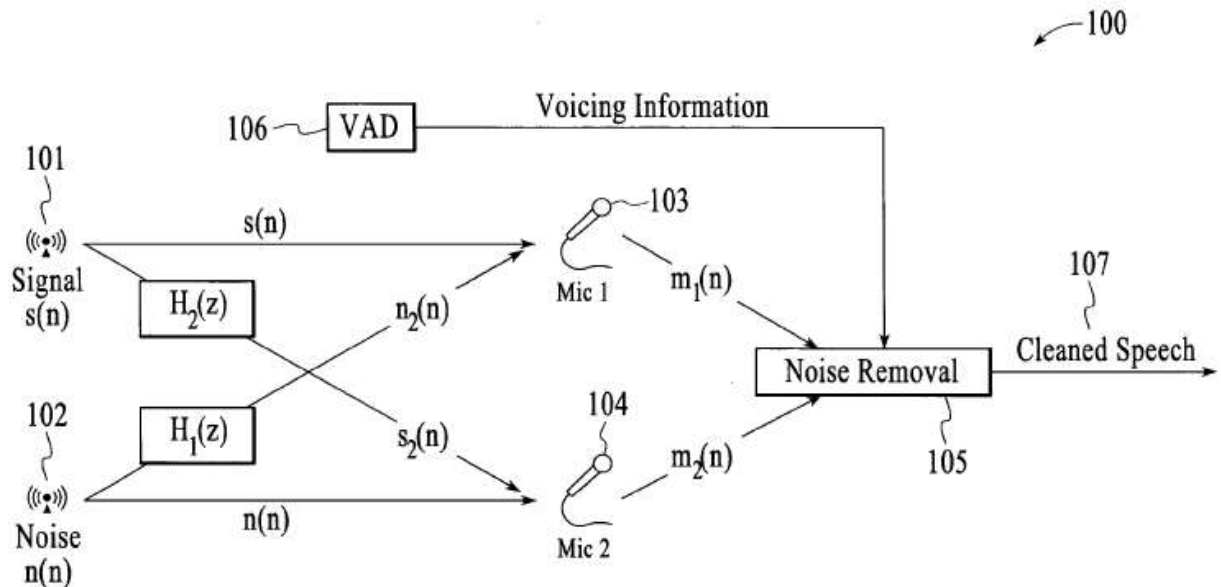


FIG.1

30. In this embodiment, “the signal processing system 105 includes two microphones MIC 1 103 and MIC 2 104 that receive signals or information from at least one speech source 101 and at least one noise source 102.” *Id.*, 14:43-46. “The path $s(n)$ from the speech source 101 to MIC 1 and the path $n(n)$ from the noise source 102 to MIC 2 are considered to be unity,” which means that MIC 1 is presumed to receive the entire speech signal and MIC 2 is presumed to receive the entire noise signal. *Id.*, 14:46-48. $H_1(z)$ is a transfer function that represents the amount of the noise signal received at MIC 1, and $H_2(z)$ is a transfer function that represents the amount of the speech signal that is received at MIC 2. *Id.*, 14:48-50.

31. The '543 Patent teaches that “[t]he acoustic information coming into MIC 1 is denoted by $m_1(n)$. The information coming into MIC 2 is similarly labeled $m_2(n)$. In the z (digital frequency) domain, we can represent them as $M_1(z)$ and $M_2(z)$. Thus,

$$M_1(z) = S(z) + N(z)H_1(z)$$

$$M_2(z) = N(z) + S(z)H_2(z).”$$

Id., 14:52-60.

32. The '543 Patent uses “[a] VAD signal 106, derived in some manner, . . . to control the method of noise removal.” *Id.*, 14:51-52. The '543 Patent teaches that using the VAD allows for simplification of these equations, such that the equation for the sounds signal $S(z)$ can be derived as:

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

Id., 15:60-62.

33. Assuming the situation where “ $H_2(z)$ is quite small,” *i.e.*, that the amount of the speech signal reaching MIC 2 is very low, and assuming that “ $H_1(z)$ is less than unity,” *i.e.*, that less than the entire noise signal reaches MIC 1, the speech signal can be calculated as:

$$S(z) \approx M_1(z) - M_2(z)H_1(z).$$

Id., 16:2-3.

34. An accurate voice activity detection signal is key to the '543 Patent's noise suppression system. *Id.*, 2:7-12. The patent explains that it is crucial to implement "a sufficiently accurate VAD . . . to adapt the coefficients of H_1 (noise only) and H_2 (if needed, when speech is being produced)." *Id.*, 16:55-67. If the VAD is inaccurate and speech is occurring during the calculation of the coefficients of H_1 (noise only), it "would result in the removal of a large part of the speech energy from the signal of interest," resulting in "signal distortion and reduction (de-signaling)." *Id.*

35. The '543 Patent discloses that using multiple subbands in processing the acoustic signals results in significant noise suppression. *Id.*, 16:9-11. The signals from MIC 1 and MIC 2 are filtered into multiple subbands and the resulting data from each subband is sent to its own adaptive filter. *Id.*, 16:31-36. The results from each subband are added back together to form the final denoised signal. *Id.*, 16:38-40.

VII. CLAIM CONSTRUCTION

36. I understand that claim terms are to be given their plain and ordinary meaning as a POSITA would have understood them in light of the specification and prosecution history.

A. “voice detection subsystem,” “voice activity signals,” and “human voicing activity”

37. In my opinion, a POSITA would understand that the “voice” referred to in the terms “voice detection subsystem,” “voice activity signals,” and “human voicing activity” refers to both voiced and unvoiced speech.

38. The specification explains that “[t]he terms ‘speech’ or ‘voice’ as used herein generally refer to voiced, unvoiced, or mixed voiced and unvoiced human speech. *Unvoiced speech or voiced speech is distinguished where necessary.*” *Id.*, 4:34-37 (emphasis added). It was well-known in the art that voiced speech occurs when the vocal cords are vibrating, and unvoiced speech occurs when the vocal cords are not vibrating. For example, Holzrichter explains the differences between voiced and unvoiced speech in similar terms. Ex. 1007, 12:60-65 (“There are two major types of speech, ‘voiced’ where the vocal folds open and close rapidly providing periodic bursts of air into the vocal tract, and ‘unvoiced’ where constrictions in the vocal tract cause air turbulence and associated ‘modified-white’ air flow noise.”),

32:44-51 (“the vocal folds which open, but do not vibrate open and closed, during whispered and unvoiced speech”)

39. The '543 Patent explains that its voice activity detector, or VAD, detects speech which includes both voiced and unvoiced speech: “the term ‘VAD’ is generally defined as a vector or array signal, data, or information that in some manner represents *the occurrence of speech* in the digital or analog domain.” '543 Patent, 4:49-52 (emphasis added). The specification of the '543 Patent does not characterize a VAD or voice detection subsystem as detecting only voiced speech or only unvoiced speech in any embodiment. Thus, “speech” and “voice” in the context of the '543 Patent, in general, and voice activity detection, in particular, include both voiced speech and unvoiced speech.

40. The '543 Patent discloses that its noise suppression system is an improvement over the prior art because it uses an accurate voice activity detection signal in the denoising process. '543 Patent, 2:7-12. The '543 Patent teaches that it is crucial to implement “a sufficiently accurate VAD . . . to adapt the coefficients of H_1 (noise *only*) and H_2 (if needed, when *speech* is being produced). *Id.*, 16:55-67 (emphasis added). If the VAD is inaccurate and speech is occurring during the calculation of the coefficients of H_1 (noise only), it “would result in the removal of a large part of the speech energy from the signal of interest,” resulting in “signal distortion and reduction (de-signaling).” *Id.* Based on these teachings, the “voice

activity” relevant to the claims must include both voiced speech and unvoiced speech. If “voice activity” included only voiced speech, then the noise only period would be corrupted with unvoiced speech, resulting in the “signal distortion and reduction (de-signaling)” that the invention was intended to avoid. The ’543 Patent’s discussion of voice activity detection does not distinguish between voiced and unvoiced speech.

41. In my opinion, a POSITA would understand that the claim term “voice detection subsystem” refers to a subsystem that detects the presence of speech, including voiced and unvoiced speech; “voice activity signals” are signals that indicate speech activity, including voiced and unvoiced speech activity; and “human voicing activity” refers to human speech activity, including voiced speech and unvoiced speech.

B. “generating a noise waveform estimate associated with noise of the acoustic signals”

42. In my opinion, a POSITA would understand “generating a noise waveform estimate associated with noise of the acoustic signals” to mean “calculating a noise waveform estimate based on the noise associated with all the acoustic signals.” This limitation is recited in independent Claims 1 and 26.

43. The claim language, as well as the specification, makes clear that the noise waveform estimate is calculated using all the acoustic signals received by the

microphone array. First, the surrounding language draws a distinction between “the acoustic *signals*” and “the acoustic signal.” For example, the relevant portion of Claim 1 reads:

wherein the denoising method includes generating a noise waveform estimate associated with noise of the acoustic signals and subtracting the noise waveform estimate from the acoustic signal when the acoustic signal includes speech and noise

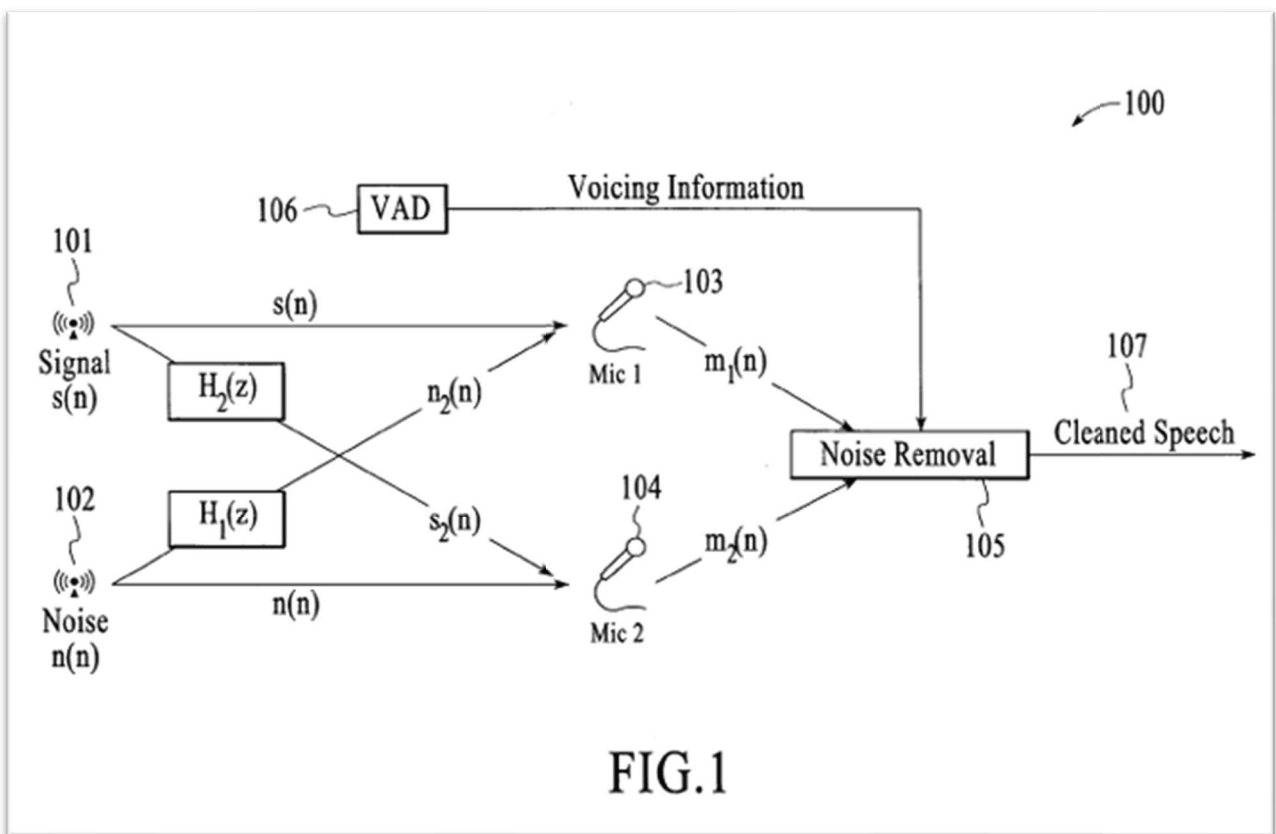
Ex. 1001, cl. 1.

44. In other words, the noise waveform estimate is generated based on all the “acoustic signals,” and that generated waveform estimate is subtracted from a single “acoustic signal.”

45. Moreover, a POSITA would read the remaining claim language and understand that the noise waveform estimate is not the same as one of the acoustic signals as output from a microphone. For example, Claim 1 states that “the microphone array providing acoustic signals of an environment to components of the denoising subsystem.” Ex. 1001, cl. 1. Claim 1 further states that “the denoising method includes . . . subtracting the noise waveform estimate from the acoustic signal.” *Id.* A POSITA would understand that a system cannot subtract an estimate from an acoustic signal, unless “acoustic signal” refers to the electrical representation of the signal.

46. The specification similarly uses “acoustic signals” to include the signals coming *from* the microphones. *E.g.*, FIG. 1 (showing $m_1(n)$ and $m_2(n)$ output from the microphone); 14:52-54. Given the inventor’s decision to recite an additional operation (generating) and a new element (the noise waveform estimate), a POSITA would understand that the noise waveform estimate was both distinct from the acoustic signals and the product of a subsequent operation.

47. FIG. 1 shows this:



48. As shown in FIG. 1, both Mic 1 and Mic 2 each receive both signal and noise. Ex. 1001, FIG. 1; *id.*, 5:29-36; 14:40-63 (“There is always some leakage of noise into MIC 1 and some leakage of signal into MIC 2.”).

49. Thus, as the specification explains, the desired signal can be characterized according to the following equations:

$$\begin{aligned}S(z) &= M_1(z) - N(z)H_1(z) \\N(z) &= M_2(z) - S(z)H_2(z) \\S(z) &= M_1(z) - [M_2(z) - S(z)H_2(z)]H_1(z) \\S(z)[1 - H_2(z)H_1(z)] &= M_1(z) - M_2(z)H_1(z)\end{aligned}$$

Ex. 1001, 14:67-15:63. Solving for S(z) results in:

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

Id., 15:60-63.

50. In practical applications, $H_2(z)H_1(z)$ will be negligible. *Id.*, 15:64-67.

Thus, the signal is:

$$S(z) \approx M_1(z) - M_2(z)H_1(z).$$

Id., 15:64-16:4.

51. $M_2(z)H_1(z)$ is the calculated noise waveform estimate that is subtracted from the acoustic signal $M_1(z)$.

52. $H_1(z)$ is calculated as the $M_1(z)$ signal over the $M_2(z)$ signal when only noise is being received. *Id.*, 14:66-15:19.

$$\begin{aligned} M_{1n}(z) &= M_{2n}(z)H_1(z) \\ H_1(z) &= \frac{M_{1n}(z)}{M_{2n}(z)} \end{aligned}$$

53. Thus, the denoised signal can be calculated as:

$$S(z) \approx M_1(z) - M_2(z)H_1(z).$$

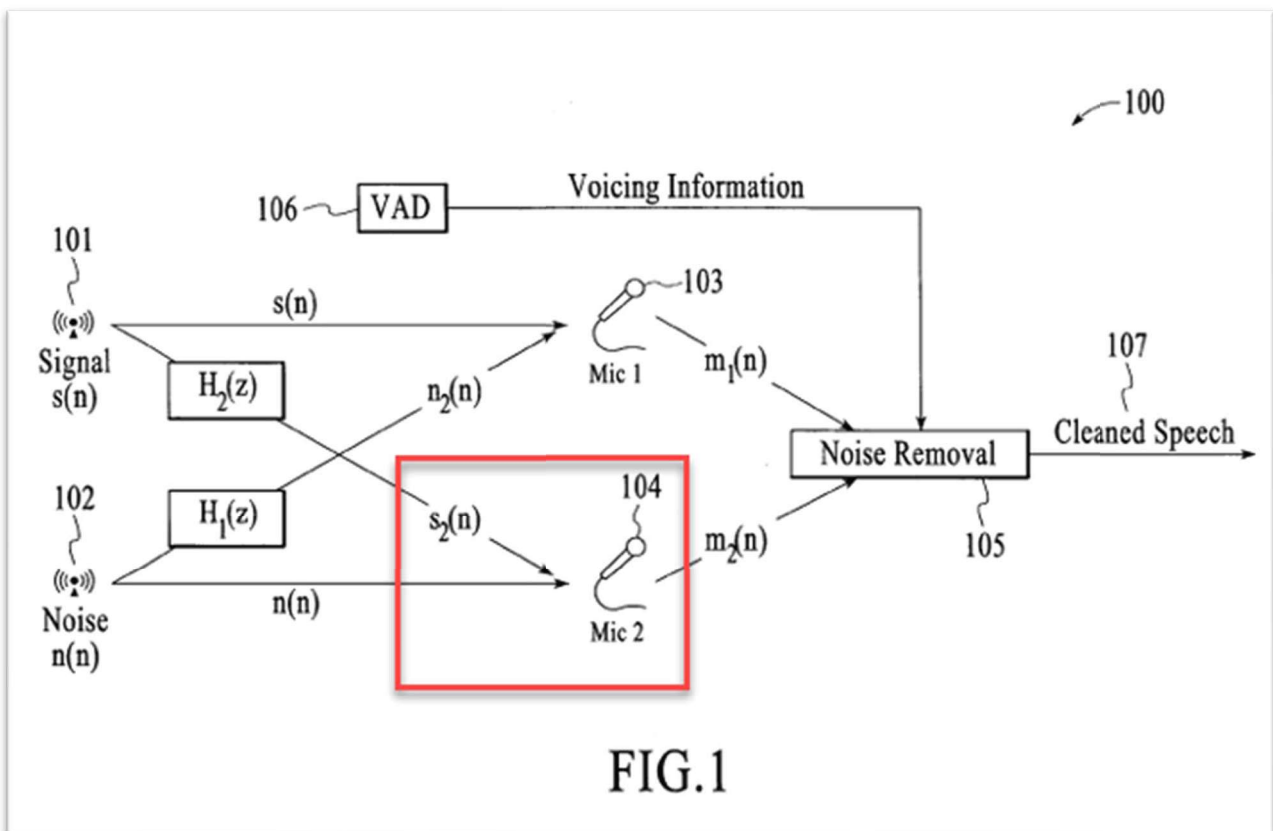
$$S(z) \approx M_1(z) - M_2(z) \frac{M_{1n}(z)}{M_{2n}(z)}$$

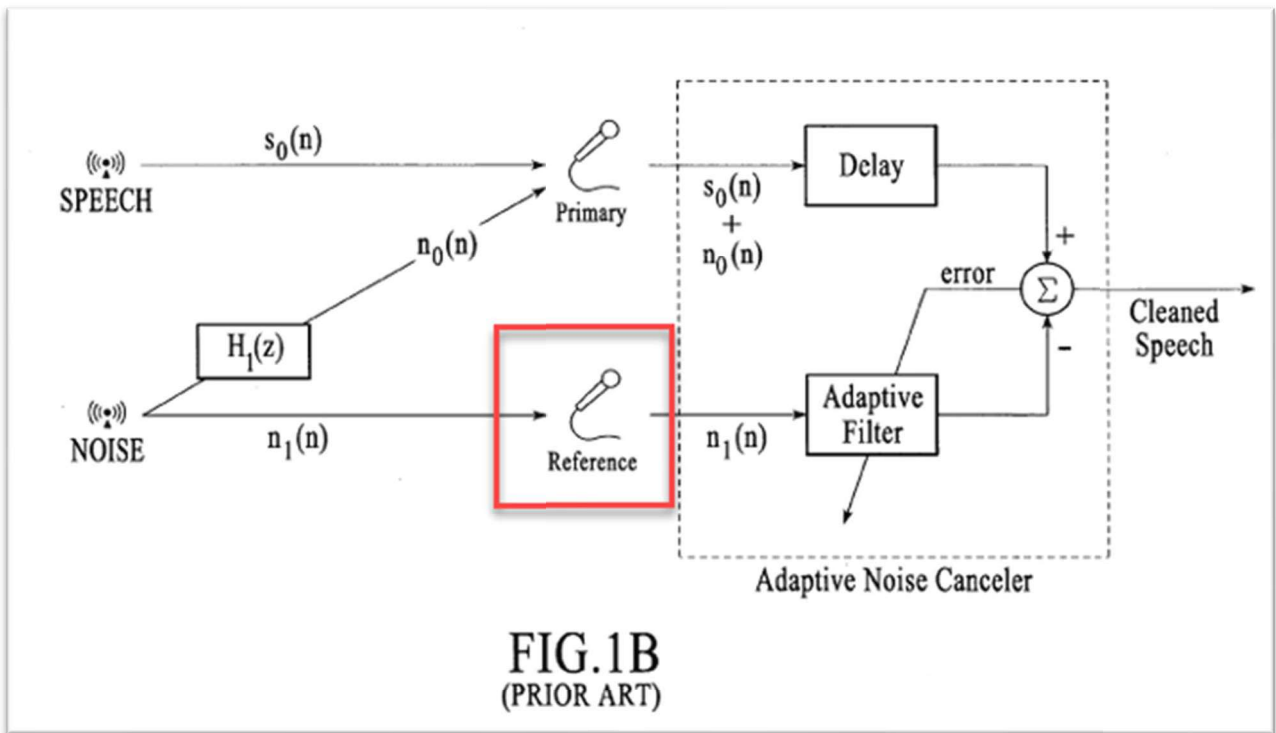
54. Therefore, the signal is calculated by subtracting the acoustic signal from MIC2 multiplied by the ratio of the noise-only component of the acoustic signals from MIC1 and MIC2—which are all the acoustic signals in the examples.

55. In contrast, the specification describes as “prior art” systems where the signal from a “reference” microphone is subtracted from a “primary” microphone, without calculating an estimate based on the noise of all microphones. Ex. 1001, FIG. 1B, 2:30-31, 16:44-54. Indeed, “supporting operation with the acoustic signal of interest in the reference microphone of the system” is called out as a specific differentiating feature. *Id.*, 16:44-54. *See also id.*, 6:58-63 (“Unlike classical adaptive noise cancellation (ANC), which specifies that there can be no speech

signal in the noise microphone, Pathfinder allows speech signal to be present in both microphones which means the microphones can be placed very close together. . . .”); 17:25-28 (“Allowing the acoustic signal to be received by the reference microphone means that the microphones can be much more closely positioned relative to each other (on the order of a centimeter) than in classical ANC configurations.”). “This closer spacing simplifies the adaptive filter calculations and enables more compact microphone configurations/solutions.” *Id.*, 17:28-31.

56. Comparing FIG. 1 with FIG. 1B shows this key difference:



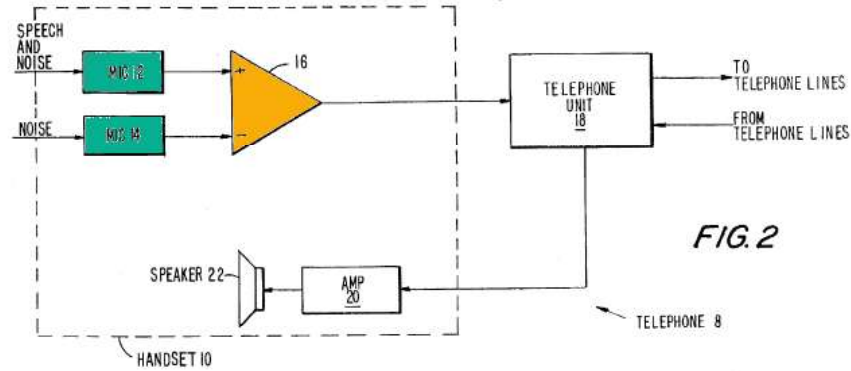


57. A POSITA would accordingly recognize that the denoising method must include subtracting a noise waveform estimate, where that noise waveform estimate is calculated based on the noise associated with all the acoustic signals.

VIII. THE ALLEGED PRIOR ART

A. Andrea (U.S. Patent No. 5,732,143)

58. I have reviewed U.S. Patent No. 5,732,143 (“Andrea”) (Ex. 1005). Andrea was filed on June 7, 1995, and issued on March 24, 1998. Andrea is directed to a primitive, analog two-microphone handset or headset apparatus. Ex. 1005, FIG. 2.



59. Andrea's FIG. 2, shown above, shows two microphones (green) and an op-amp (yellow) used to subtract the noise signal of the second microphone from the speech and noise signal of the first microphone.

60. The first microphone, MIC 12, receives speech and background noise, and the second microphone, MIC 14, receives background noise. *Id.*, 12:55-62. The speech-and-noise signal is denoised in the op-amp 16 by subtracting the noise signal of the second microphone from the speech and noise signal of the first microphone. *Id.*, 12:62-13:1.

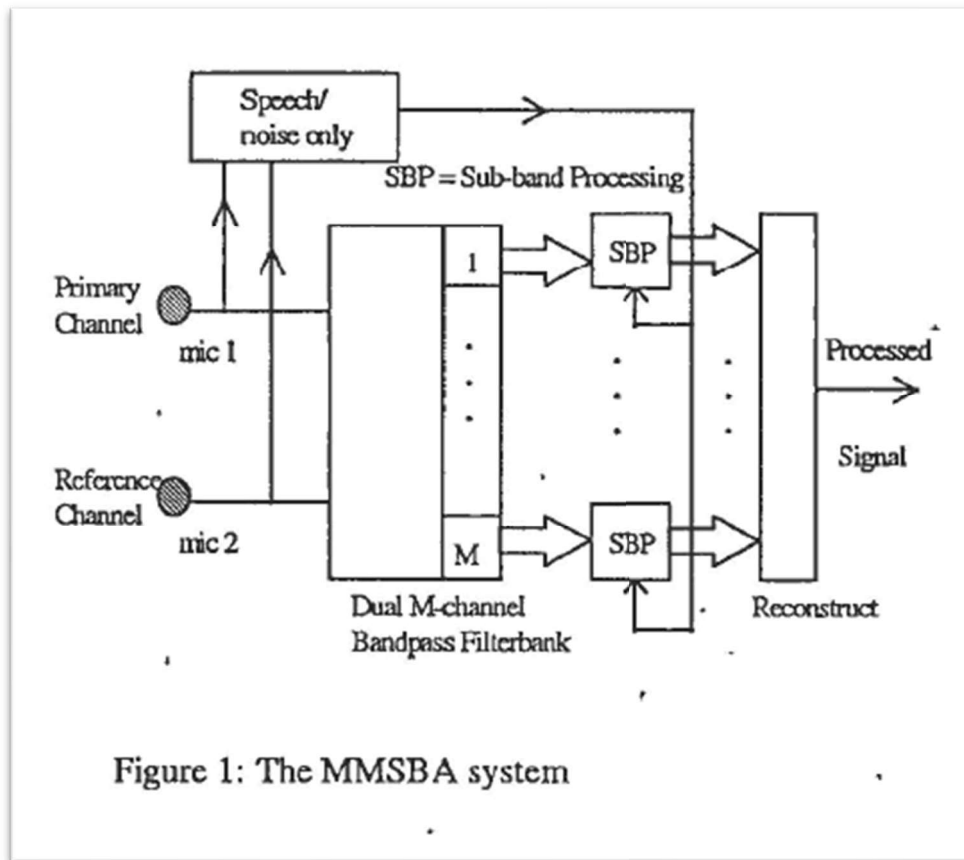
61. I agree with Dr. Anderson that the second microphone (Mic 14 in FIG. 2 above) will invariably pick up both the background noise, as well as some speech. Ex 2005, 23:13-26:6. Thus, the output of the second microphone will not be noise only.

62. Prior to the subtraction of the signal from the second microphone from the signal in the first microphone to create the denoised signal, Andrea matches the

level of the output signal of the second microphone to the level of the output signal of the first microphone. Ex. 1005, 15:29-34. This is done to reduce the amplitude of the difference between the output level of the microphone to improve noise suppression. *Id.*, 15:34-38.

B. Hussain (*A New Metric for Selecting Sub-band Processing in Adaptive Speech Enhancement Systems*)

63. I have reviewed the paper *A New Metric for Selecting Sub-band Processing in Adaptive Speech Enhancement Systems* (“Hussain”) (Ex. 1006). Hussain was purportedly published in 1997. Hussain is directed to “A multi-microphone adaptive speech enhancement system employing diverse sub-band processing.” Ex. 1006 at 10.



64. Hussain discloses that it has developed a “new robust metric . . . which is capable of real-time implementation, in order to *automatically select* the best form of processing within each sub-band.” *Id.* (emphasis added). “In order to realize this, a robust practical metric has been developed, which is capable of realtime implementation, based on the inter-channel Magnitude Squared Coherence (MSC) relationship *in order to automatically select the best SBP [Sub-Band Processing] option.*” *Id.* (emphasis added). Based on the MSC, Hussain automatically chooses one of three options for subband processing: “(i) no processing; (ii) intermittent coherent noise canceller; and (iii) continuous incoherent noise canceller.” *Id.*

65. Hussain “propose[s] the use of a modified MSC as a part of a system for selecting the best SBP option in a [multi-microphone sub-band adaptive] speech enhancement system.” Hussain, 2612 (emphasis added). Hussain derives the following equations for the MSC:

$$\overline{\text{MSC}}(f_k) = \frac{1}{l} \sum_{i=1}^l \text{MSC}(f_k, i)$$
$$\text{MSC}(f_k, l) = \frac{|S'_{x_1 x_2}(f_k, l)|^2}{S_{x_1 x_1}(f_k, l) S_{x_2 x_2}(f_k, l)}$$

Hussain, 2612-13. According to Hussain:

The above adaptively averaged MSC criterion can thus be used as an effective means for determining the level of correlation between the disturbing noises at various frequencies bands (by averaging the above MSC over each respective sub-band), during the noise alone period in intermittent speech. *The subsequent form of processing in each respective frequency band can therefore be selected on the basis of the inter-channel correlation.*

Id. (emphasis added).

66. Hussain uses the MSC to choose between using an “intermittent coherent noise canceller” and an “incoherent noise canceller.” Hussain, 2612-13. According to Hussain, “[o]n initial trials, a threshold value of 0.6 for the adaptive

MSC has been found to be suitable for distinguishing between highly correlated and weakly correlated sub-band noise signals.” Hussain, 2613. Hussain teaches that when “the noise power is significant and the noise between two channels is significantly correlated in a sub-band, then perform adaptive intermittent noise cancellation.” Hussain, 2612. But, “[i]f the noise power is significant but not highly correlated between the two channels in a sub-band, then the incoherent noise cancellation approach . . . may be applied during the noisy speech period.” *Id.*

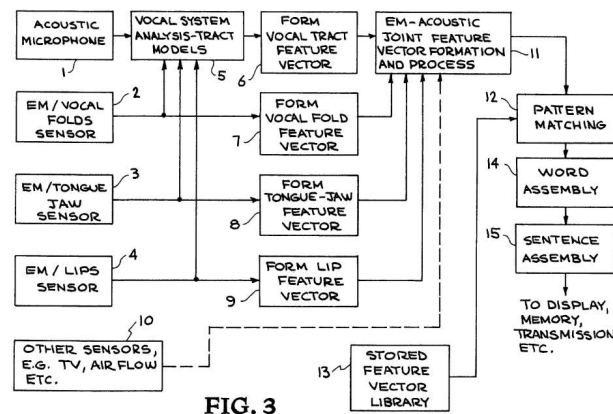
67. Hussain’s system uses an array of “[t]wo or more relatively closely spaced microphones” in an adaptive noise cancellation scheme. Ex. 1006 at 10. Hussain uses a filter bank to divide the incoming signals into subbands. *Id.* During the noise-alone period, *i.e.*, when there is no speech, Hussain determines adaptive filters which model the transfer functions between the microphones when there is only noise. *Id.* at 11. These adaptive filters are applied in each subband and the resulting signals are reconstructed to yield the denoised speech signal. *Id.*, FIG. 1.

68. Hussain assumes that an effective voice activity detector (VAD) is available for use in his system. *Id.* at 11. Hussain does not disclose any specific voice activity detector.

C. Holzrichter (U.S. Patent No. 6,006,175)

69. I have reviewed U.S. Patent No. 6,006,175 (“Holzrichter”) (Ex. 1007). Holzrichter was filed on February 6, 1996, and issued on December 21, 1999.

Holzrichter “relates generally to speech recognition and more particularly to the use of nonacoustic information in combination with acoustic information for speech recognition.” Ex. 1007, 1:12-15.



70. Holzrichter discloses using EM sensors for measuring the positions and movements of certain organ positions, including the vocal folds, velum, tongue, and lips. *Id.*, 16:42-49. Feature vectors are generated which describe the signals generated by the EM sensors which are then combined with feature vectors describing acoustic information “to produce a more complex, more information-rich feature vector describing the speech unit time frame.” *Id.*

71. Holzrichter’s pattern matches the combined feature vector against a feature vector library to identify the sound represented by the feature vector. *Id.*, 17:32-33. Words and sentences are assembled from the collection of identified sounds to yield the results of Holzrichter’s voice recognition system. *Id.*, 17:34-36.

72. Holzrichter teaches that the algorithm for determining when speech begins, *i.e.*, the onset of speech, requires the use of an EM sensor and an acoustic sensor. *Id.*, 31:33-32:54. “The algorithm detects the onset of speech by using both an acoustic and one or more EM sensor modules at the same time.” *Id.*, 31:51-53. Although the EM sensors are effective for detecting voiced speech, the acoustic sensor is used to detect unvoiced speech when no vocal fold motion is detected. *Id.*, 32:7-10, FIG 19, element 79.

D. Stevens (*A Miniature Accelerometer for Detecting Glottal Waveforms and Nasalization*)

73. I have reviewed *A Miniature Accelerometer for Detecting Glottal Waveforms and Nasalization* (“Stevens”) (Ex. 1008). Stevens was purportedly published in 1975. Stevens is directed to a neck-attached accelerometer used in a computer-based system of speech-training aids for deaf children and a system for the training of pronunciation of a second language. *Id.* at 7.

E. Alcivar (U.S. Patent No. 3,746,789)

74. I have reviewed U.S. Patent No. 3,746,789 (“Alcivar”) (Ex. 1009). Alcivar was filed on October 20, 1971, and issued on July 17, 1973. Alcivar is directed to a tissue-conduction microphone in an underwater communications system that acts as a switch to enable a transmitter in a communications system when the user is speaking. Ex. 1009, Abstract, 1: 9-13.

IX. PETITIONER’S COMBINATIONS DO NOT RENDER ANY CLAIM OBVIOUS

A. GROUND 1: The Combination of Andrea, Hussain, and Holzrichter Does Not Render Obvious Claims 1-2 and 5-26

1. Meta’s Combination Does Not Render Obvious “wherein the denoising method includes generating a noise waveform estimate associated with noise of the acoustic signals,” [1H]

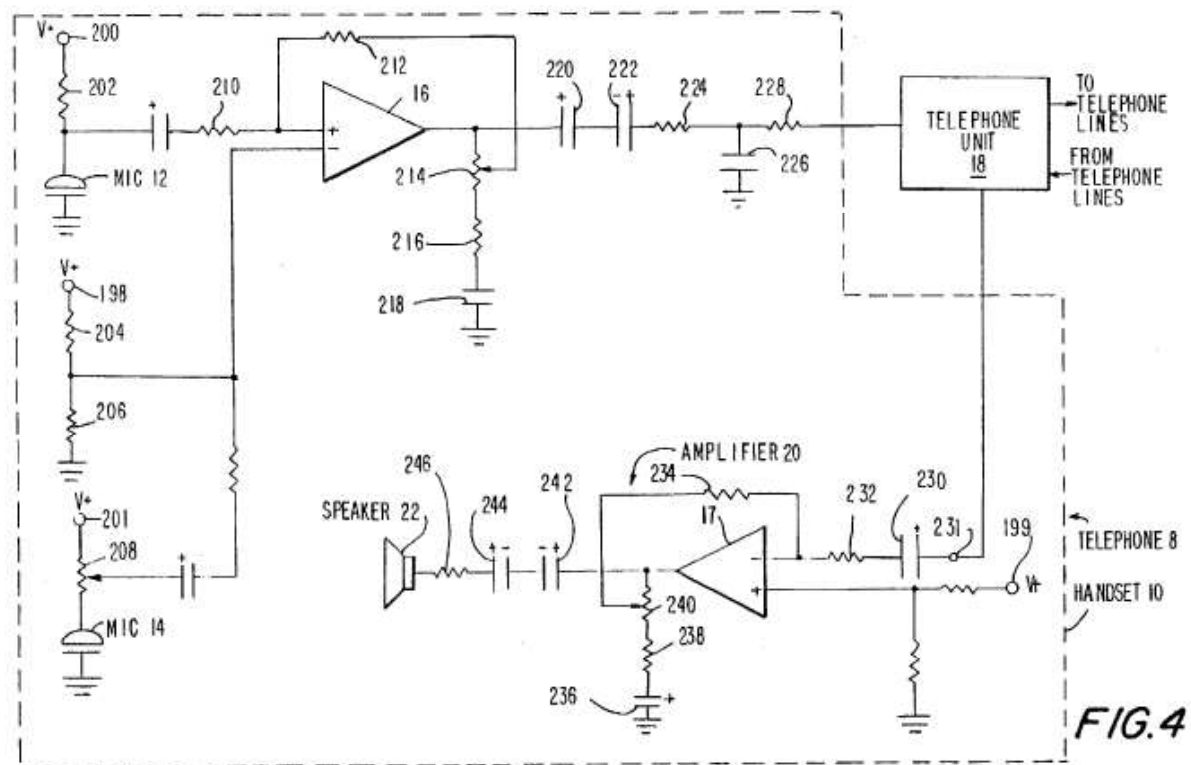
75. Petitioner argues that this limitation is disclosed or suggested by Andrea alone. Pet., 40-41. In particular, Petitioner states that “converting the background noise supplied to the second microphone into a corresponding electrical signal” is “generating a noise waveform estimate associated with noise of the acoustic signals.” Petitioner further alleges “A POSITA would have understood the “corresponding electrical signal” and/or the level-matched signal from second microphone 14 is a noise waveform estimate of the background noise.” *Id.* I disagree.

i. The Combination Does Not Render the Limitation Obvious Under the Proper Construction

76. As I explain above, “generating a noise waveform estimate associated with noise of the acoustic signals” should be construed as “calculating a noise waveform estimate based on the noise associated with all the acoustic signals.”

77. Petitioner’s combination cannot meet this limitation as properly construed. Andrea simply takes the output of a microphone (Mic 14 below) and

connects it to the input of an op-amp, which subtracts that signal from that received from the other microphone (Mic 12). Ex. 1005, 12:59-62, 14:23-29. That signal may be passed through a current-limiting resistor 208 which matches the level of the second microphone signal to that of the first microphone signal. *Id.*, 15:25-40. As Petitioner admits, the only alleged generating is when the microphone converts the received audio into an electrical signal. Pet., 40.



78. It is clear, however, that the claims of the '543 Patent require more than just converting a sound to an electrical signal. First, the claims uniformly use “acoustic signal” to denote the signal coming from the microphone array. For example, Claim 1 reads “the microphone array *providing acoustic signals* of an

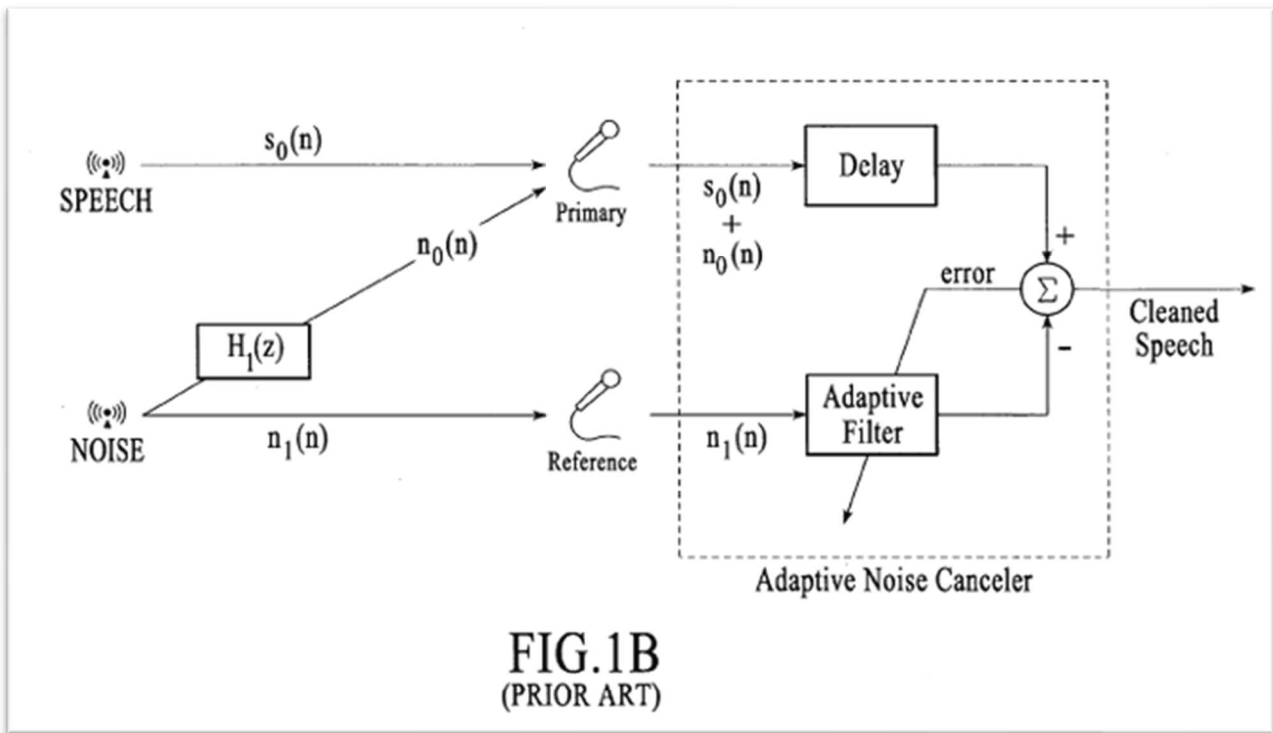
environment *to components of the denoising subsystem*". Ex. 1001, cl. 1. The only signal a microphone can provide (and a denoising subsystem can receive) is an electrical signal.

79. But the claims require *generating a noise waveform estimate*. The use of an active operation -- "generating" -- and the use of a different term for the output -- "noise waveform estimate" -- makes clear that the noise waveform estimate is not simply the acoustic signal from one microphone.

80. Andrea, both in its own disclosure and as used in Petitioner's combination, simply passes an acoustic signal (in the language of the '543 Patent) to the op-amp.

81. Petitioner's combination thus does not calculate a noise waveform estimate, as required, but instead simply passes a microphone signal to a subtractor. Similarly, the resistor does not calculate a noise estimate waveform, but merely applies a gain to that same signal. Ex. 1005, 15:25-40.

82. Further, the Andrea microphone signal is not based on the noise associated with all the acoustic signals. There is no connection between the Mic 12 and Mic 14 inputs, and thus Andrea's alleged noise waveform estimate is not based on the noise received at Mic 12. Instead, Andrea discloses much the same system as the FIG. 1B prior art system discussed in the '543 Patent.



83. Accordingly, in my opinion, Meta’s combination does not disclose or render obvious this limitation.

ii. The Combination Does Not Render the Limitation Obvious Under the Claim Language

84. Even if the Board were to decline to construe the “generating. . .” term, Meta’s combination does not disclose it or render it obvious.

85. Meta’s combination does not perform any “generating” as required. Instead, as I explain above, it simply receives a sound at a microphone and passes it (as an electrical signal) to an op-amp. But the claims use “acoustic signals” to include the electrical signals emanating *from* the microphone. For example, Claim

1 states that “the microphone array providing acoustic signals of an environment to components of the denoising subsystem.” Ex. 1001, cl. 1. Claim 1 further states that “the denoising method includes . . . subtracting the noise waveform estimate from the acoustic signal.” *Id.* The specification similarly uses “acoustic signals” to include the signals coming *from* the microphone. *E.g.*, FIG. 1 (showing $m_1(n)$ and $m_2(n)$ output from the microphone); 14:52-54.

86. Moreover, the claims explicitly recite that the noise waveform estimate is the product of “generating.” If, as Meta argues, the signal output from a microphone could be the noise waveform estimate, it would render that part of the limitation meaningless.

87. Similarly, a microphone signal is not a “noise waveform estimate.” The term itself requires that it be an estimation of the noise. The microphone signal is not the product of any estimation. Instead, it is simply the sound picked up from one microphone, and includes both speech and noise.

88. Finally, the Andrea microphone signal is not “associated with noise of the acoustic signals.” It is associated with, at most, the noise of one acoustic signal.

89. A POSITA would thus understand that the noise waveform estimate cannot simply be the signal from a microphone.

90. Accordingly, under the claim language of the limitation as written, Petitioner’s combination does not render this limitation obvious.

2. Meta’s Combination Does Not Disclose or Render Obvious “subtracting the noise waveform estimate from the acoustic signal when the acoustic signal includes speech and noise,” [1H]

91. In my opinion, a POSITA would not have been motivated to combine Andrea and Hussain in the manner proposed by Meta, and a POSITA would not have reasonably expected success in making the combination because the systems in Andrea and Hussain are incompatible with one another.

92. Andrea discloses an analog noise suppression apparatus, whereas Hussain discloses a more advanced digital noise suppression system. I understand that Dr. Anderson agrees with this assessment. Ex. 2005, 59:14-19. Andrea denoises an analog speech-and-noise signal by subtracting the noise signal of the second microphone from the speech-and-noise signal of the first microphone using an op-amp 16. Ex. 1005, 12:63-13:1. Hussain converts the input analog signal to a digital signal and applies digital signal processing methods to these signals using processor functionality. Hussain uses this digital processing to filter a speech-and-noise signal into subbands and further calculates and applies adaptive filters to remove the noise in each subband. Ex. 1006 at 11. The resulting denoised subband signals are reconstructed to yield the denoised speech signal. *Id.*, FIG. 1.

93. A POSITA would recognize that the noise-reduction processing of Andrea is completely different from and incompatible with that disclosed in Hussain.

The Petition argues that a POSITA “would have been motivated to implement Hussain’s sub-band processing and VAD in Andrea’s noise cancellation apparatus to improve noise cancellation across frequencies and improve performance in reverberant environments.” Pet. at 16. Dr. Anderson characterized it as a “drop-in” replacement where Andrea’s handset or headset configuration, microphones, and microphone pre-amps are “drop[ped] . . . directly into Hussain as the microphone array.” Ex. 2005, 41:14-24.

94. Dr. Anderson’s description of the combination is different from the argument made in the Petition. The Petition states that Andrea’s op-amp 16 meets the limitation of “subtracting the noise waveform estimate from the acoustic signal when the acoustic signal includes speech and noise.” Pet. at 40-41. The Petition and Dr. Anderson rely solely on Andrea for the subtraction limitation and do not reference Hussein at all. Pet., 40-41; Ex. 1003, ¶¶ 132-33. Therefore, Petitioner’s combination uses Andrea’s op-amp 16 for subtracting a noise signal from a speech-and-noise signal with Hussein’s digital subband adaptive filter solution.

95. A POSITA would not have been motivated to combine Andrea and Hussain, as described in the Petition, and would not have reasonably expected to succeed in making this combination because Andrea’s subtraction operation, generally, and its op-amp, specifically, would have no role in Hussain’s complete digital denoising solution. Hussain’s noise reduction solution does not process

analog signals beyond the conventional preprocessing needed before performing an analog to digital conversion that is required in all digital signal processing systems. Furthermore, Hussain does not perform an analog subtraction operation between signals before his system would convert to a digital signal. In fact, Hussain's simulation does not use any analog processing and is performed only in the digital processing domain as simulations from two independent white noise sources corrupting a speech signal sampled at 10 KHz. Ex. 1006 at 12. The Petition does not contend that including Andrea's teaching of subtracting signals would improve the denoising performed by Hussain's system. A POSITA would understand that no such role is technologically plausible because Andrea's low-cost analog subtraction operation and Hussain's digital-processing-based adaptive filtering are technological alternatives to one another and not combinable in a coherent solution.

96. Accordingly, a POSITA would not be motivated to combine Andrea and Hussain as Meta proposes, and the combination would not have been reasonably likely to be successful. In my opinion, this limitation is not rendered obvious by Meta's combination.

3. Meta’s Combination Does Not Disclose or Render Obvious “wherein the voice detection subsystem is configured to receive the voice activity signals using a sensor independent from the microphone array and to output the control signals generated from the voice activity signals to the denoising system” [1I]

97. Meta’s combination does not disclose or render obvious “wherein the voice detection subsystem is configured to receive the voice activity signals using a sensor independent from the microphone array and to output the control signals generated from the voice activity signals to the denoising system.”

i. Andrea Does Not Teach This Limitation

98. Meta does not argue that Andrea discloses a voice detection subsystem configured to receive voice activity signals “using a sensor independent from the microphone array.” Pet. at 41. Meta states that Andrea’s VOX circuit 1950 is a voice detection system. Pet. at 13. The VOX circuit 1950 is disclosed as “connected to the microphone output of the microphone amplifier 1940.” Ex. 1005, 34:14-16. The VOX circuit uses this input to determine whether “the signal at the output of the microphone is usable audio.” *Id.*, 34:22-24. For this reason, it is my opinion that Andrea’s VOX circuit does not receive voice activity signals using a sensor independent from Andrea’s microphone array because it uses a microphone output as its input.

**ii. Holzrichter Does Not Teach or Render
Obvious This Limitation**

99. Meta also argues that Holzrichter renders obvious a voice detection subsystem using a sensor independent from the microphone array to receive the voice activity signals and output the control signals. Pet. at 41. Meta argues that the claimed voice detection subsystem includes the EM sensor module 23 and the functional block that forms the vocal fold feature vector from the output of the EM sensor 23. Pet. at 42-43 (citing Ex. 1007, FIG. 3, elements 2 and 7); Ex. 2005, 85:25-86:6. The EM sensor detects motion in the vocal folds and glottis. Ex. 1007, 17:62-63. Meta argues that the EM sensor 23 is independent from the microphone array because Holzrichter discloses a microphone 24 which obtains simultaneous acoustic information but which is shown separately in FIG. 4 of Holzrichter. Pet. at 42.

100. In my opinion, Holzrichter does not disclose or render obvious a voice detection system. The voice detection system identified by the Petition—elements 2 and 7 of FIG. 3—is used to identify only the *voiced* portions of speech. Pet. at 45 (“Holzrichter’s control signals . . . provide voice activity information on whether *voiced* speech is occurring. . . .”) (emphasis added); Ex. 1003, ¶¶ 130, 140. The claimed voice detection system, properly construed, must detect speech which includes both voiced and unvoiced speech.

101. Holzrichter teaches that the algorithms for detecting speech, the “onset of speech time algorithm” and the “end-of-speech algorithm,” use both an acoustic detector, *i.e.*, a microphone, and one or more EM sensor modules at the same time. Ex. 1007, 31:33-33:18. The algorithm in FIG. 19 uses the EM sensor to identify voiced speech and the acoustic detector to identify unvoiced speech because the EM sensor does not pick up unvoiced speech such as “s” in “sam or in “sixteen.” *Id.*, 31:51-60.

102. Holzrichter states:

If acoustics and vocal fold motion are present, then a voiced speech unit has occurred and the unvoiced timer, T , is set to zero in box 78, and processing on frame t_i continues. If satisfactory acoustics, *but no vocal fold motion occurs* in box 72, the sound in frame t_i is labeled as nonvoiced speech in box 79 and processing continues.

Ex. 1007, 32:4-10 (emphasis added). Holzrichter discloses that vocal fold motion detected by an EM sensor can detect voiced speech but not unvoiced speech. “The issue for this algorithm is that some words begin with unvoiced sounds such as ‘s’ in ‘sam’ or in ‘sixteen’; thus the algorithm must be able to back up *to catch sounds that may be missed by the vocal fold motion onset detection by the EM sensor.*” *Id.*, 31:53-57 (emphasis added). In my opinion, a POSITA would, therefore, recognize that the Petition’s proposed voice detection system would not meet the requirements of the claims because it does not detect unvoiced speech.

103. In my opinion, a POSITA would not have “found it obvious to use Holzrichter’s control signals in Hussain’s sub-band processing noise-cancellation system to provide voice activity information on whether voiced speech is occurring to control processing based on whether there is a noise-only or speech-plus-noise period occurring.” Pet. at 45. Hussain’s system includes a “speech/noise only” block whose output identifies two time periods: a “speech plus noise period” and a “noise alone period.” Ex. 1006 at 11, 13, FIG. 1. The output from this block is used by Hussain to create the adaptive filters used to denoise the noisy speech signal. Ex. 1006 at 11. The control signals from Petitioner’s “voice detection system” based solely on an EM sensor which detects only voiced speech would not provide an accurate “noise only” signal to Hussain because periods of unvoiced speech would be undetected as speech and mischaracterized as noise. For this reason, Holzrichter pairs the EM sensor with an acoustic sensor to pick up unvoiced speech to identify the entirety of a speech signal and compensate for the failings of the EM sensor.

104. A POSITA would not be motivated to combine Holzrichter’s control signals from the EM sensor with Hussain because it would reduce the effectiveness of Hussain’s denoising system. Hussain assumes “that an effective voice activity detector (VAD) is available” because the efficacy of the adaptive filters which model the transfer function between the microphones depends on the determination of the adaptive filters occurring during the noise only period. Ex. 1006 at 11. If the VAD

used by Hussain is inaccurate and speech is occurring when Holzrichter's control signals indicate a noise only period, the adaptive filters will not properly model the transfer function and, therefore, the noise suppression of the Hussain system would suffer.

105. A POSITA would not have reasonably expected success in combining the control signals from Holzrichter's EM sensor with the Hussain system for the same reasons that would have dissuaded a POSITA from making the combination in the first instance: the resulting combination would not be a successful denoising system. A POSITA would not have regarded a VAD which senses only voiced speech as an accurate VAD and would not have expected the resulting combination to yield a successfully accurate denoising system. A POSITA reviewing the Holzrichter reference would recognize that to detect unvoiced speech, the output of the EM sensor was paired with the output of an acoustic sensor, but Meta's combination explicitly excludes the acoustic sensor and its output from its identification of the claimed voice detection system.

106. Accordingly, I disagree that Meta's combination discloses or renders obvious this limitation.

B. GROUND 2-3: These Grounds Fail Based on the Failure of Ground 1

107. Petitioner relies on the combination of Andrea, Hussain, and Holzrichter with Stevens (Ground 2) or Alcivar (Ground 3). Neither Grounds 2 nor 3 remedies the deficiencies of Ground 1. Accordingly, Grounds 2 and 3 are deficient for the same reasons discussed with respect to Ground 1.

X. CONCLUSION

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

I hereby 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 this 26th day of June, 2024

Respectfully submitted,

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

Akbar M Sayeed
