

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

**JAWBONE INNOVATIONS, LLC'S
PATENT OWNER RESPONSE**

Case No. IPR2023-01291

TABLE OF CONTENTS

	<u>Page(s)</u>
I. INTRODUCTION	1
II. THE '543 PATENT	2
III. THE scope and content of the RELEVANT ALLEGED PRIOR ART	6
A. Andrea (U.S. Patent No. 5,732,143)	6
B. Hussain (A New Metric for Selecting Sub-band Processing in Adaptive Speech Enhancement Systems)	7
C. Holzrichter (U.S. Patent No. 6,006,175)	10
D. Stevens (A Miniature Accelerometer for Detecting Glottal Waveforms and Nasalization)	12
E. Alcivar (U.S. Patent No. 3,746,789)	12
IV. CLAIM CONSTRUCTION	13
A. “voice detection subsystem,” “voice activity signals,” “human voicing activity”	14
B. “generating a noise waveform estimate associated with noise of the acoustic signals”	16
V. LEVEL OF ORDINARY SKILL IN THE ART	23
VI. ARGUMENT	24
A. GROUND 1: Andrea, Hussain, and Holzrichter Do Not Render Obvious Claims 1-2 and 5-26	24
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]	24

i.	The Combination Does Not Render the Limitation Obvious Under the Proper Construction.....	25
ii.	The Combination Does Not Render the Limitation Obvious Under the Claim Language as Written.....	28
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].....	30
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,” as Required by Claims 1-2 and 5-26 [1I]	33
i.	Andrea Does Not Teach This Limitation.....	33
ii.	Holzrichter Does Not Teach or Render Obvious This Limitation.....	34
4.	Andrea, Hussain, and Holzrichter Do Not Disclose or Render Obvious Claim 26.....	38
B.	GROUND 2-3: These Grounds Fail Based on the Failure of Ground 1	38
VII.	CONCLUSION.....	38

TABLE OF AUTHORITIES

	Page(s)
Cases	
<i>DePuy Spine, Inc. v. Medtronic Sofamor Danek, Inc.</i> , 469 F.3d 1005 (Fed. Cir. 2006).....	14
<i>Nidec Motor Corp. v. Zhongshan Broad Ocean Motor Co.</i> , 868 F.3d 1013 (Fed. Cir. 2017).....	13
<i>Pause Tech., LLC v. TiVo, Inc.</i> , 419 F.3d 1326 (Fed. Cir. 2005).....	18, 29
<i>Phillips v. AWH Corp.</i> , 415 F.3d 1303 (Fed. Cir. 2005) (<i>en banc</i>).....	14
<i>Rovi Guides, Inc. v. Vidal</i> , 2020-1994, 2022 WL 1152311 (Fed. Cir. Apr. 19, 2022).....	18, 26
<i>Thorner v. Sony Comput. Ent. Am. LLC</i> , 669 F.3d 1362 (Fed. Cir. 2012).....	14
Statutes	
35 U.S.C. § 282(b).....	13
Other Authorities	
37 C.F.R. § 42.100(b).....	13

EXHIBIT LIST

Exhibit No.	Description of Document
2001	Deposition of David Anderson, taken June 14, 2024
2002	<i>unused</i>
2003	Declaration of Akbar M. Sayeed, Ph.D.
2004	Curriculum Vitae of Akbar M. Sayeed, Ph.D.
2005	Deposition of David V. Anderson, Ph.D., taken March 1, 2023, in IPR2022-01027

I. INTRODUCTION

The Board should issue a Final Written Decision finding all the Challenged Claims not unpatentable because Meta's obviousness combination, even taken as Meta describes it, fails to disclose or render obvious the key limitation "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."

Meta relies on the Andrea reference to supply this limitation, but Andrea does not generate any noise waveform estimate. Instead, Andrea simply subtracts the signal received at one microphone from another. The '543 Patent specifically distinguishes its invention from this sort of noise cancellation, and Meta does not even attempt to explain how a POSITA would overcome the differences between the prior art and the claimed invention.

Moreover, Meta selectively removes portions (in particular, the vocal fold EM sensor) of the Holzrichter reference for use as a voice activity detector (VAD). But Holzrichter explains that the vocal fold sensor, alone, cannot determine whether a user is speaking. Instead, Holzrichter requires the use of multiple sensors, including acoustic microphones, to properly detect speech.

Accordingly, for the reasons set forth below, and supported by the testimony of Dr. Akbar M. Sayeed (Ex. 2003) and the cross-examination testimony of Meta's

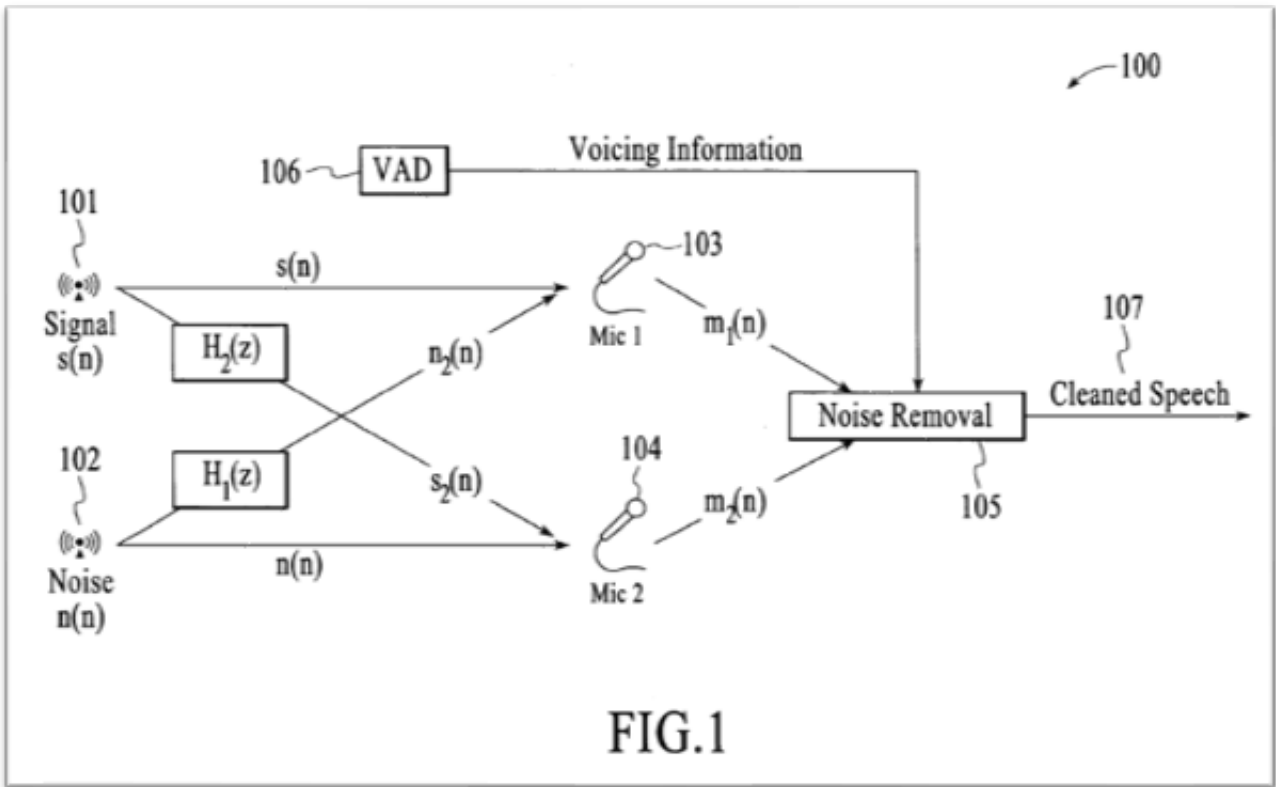
expert Dr. David Anderson (Exs. 2001, 2005¹), the Board should issue a Final Written Decision finding all the Challenged Claims to be not unpatentable.

II. THE '543 PATENT

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 (Ex. 1001). Embodiments of the system use two microphones and a voice activity detection (VAD) system, which includes both a VAD device and a VAD algorithm. *Id.*, 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. Ex. 2003, ¶ 28.

¹ Ex. 2005 is the deposition of Dr. Anderson taken in IPR2022-01027, on the same declaration as in this case. Ex. 2001, 13:7-15. Dr. Anderson confirmed that his deposition testimony in that case was accurate and that his opinions have not changed in the interim. *Id.*

The '543 Patent describes the embodiment of Figure 1:



'543 Patent, FIG. 1; Ex. 2003, ¶ 29.

In this embodiment, “[t]he 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.” ’543 Patent, 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. Ex. 2003, ¶ 30.

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

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

$$M_2(z) = N(z) + S(z)H_2(z)$$

'543 Patent, 14:52-60. Ex. 2003, ¶31.

The '543 Patent uses “[a] VAD signal 106, derived in some manner, . . . to control the method of noise removal.” '543 Patent, 14:51-52. 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. Ex. 2003, ¶ 32.

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:

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

’543 Patent , 16:2-3. Ex. 2003, ¶ 33.

The ’543 Patent teaches 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. 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.* Ex. 2003, ¶ 34.

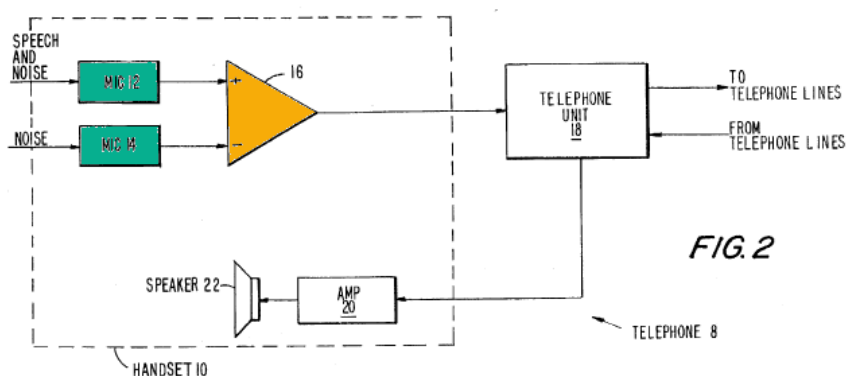
Using multiple subbands in processing the acoustic signals results in significant noise suppression. ’543 Patent , 16:9-11. The signals from MIC 1 and MIC 2 are filtered into multiple subbands and the resulting data from each subband is sent to its own adaptive filter. *Id.*, 16:31-36. The results from each subband are

added back together to form the final denoised signal. *Id.*, 16:38-40. Ex. 2003, ¶ 35.

III. THE SCOPE AND CONTENT OF THE RELEVANT ALLEGED PRIOR ART

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

U.S. Patent No. 5,732,143 (Ex. 1005) (“Andrea”) is directed to a primitive two-microphone handset or headset apparatus. Ex. 1005, FIG. 2. The first microphone receives speech and noise, and the second microphone receives only noise. *Id.*, 12:55-62. The speech-and-noise signal is denoised in the op-amp by subtracting the noise signal of the second microphone from the speech and noise signal of the first microphone. *Id.*, 12:63-13:1. This architecture is shown in Andrea FIG. 2, reproduced below, with microphones highlighted in green and op-amp in yellow. Ex. 2003, ¶¶ 58-62.

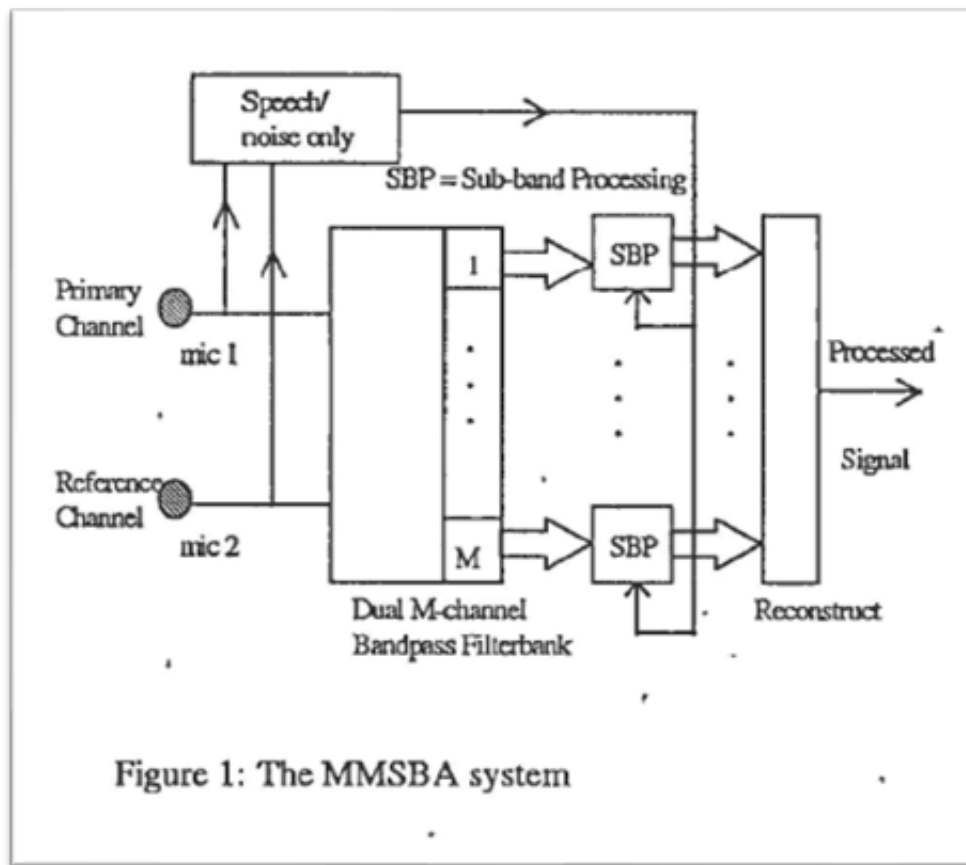


Ex. 1005, FIG. 2 (annotations added). While MIC14 is labeled “noise,” it will inevitably pick up speech as well as noise in operation. Ex 2005, 23:13-26:6. Thus, the output from MIC 14 will include some speech. Ex. 2003, ¶ 61.

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. Ex. 2003, ¶ 62.

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

The paper *A New Metric for Selecting Sub-band Processing in Adaptive Speech Enhancement Systems* (“Hussain”) (Ex. 1006) was purportedly published in 1997. Hussain is directed to “A multi-microphone adaptive speech enhancement system employing diverse sub-band processing.” Ex. 1006, 2611. Ex. 2003, ¶¶ 63-68.



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.” Ex. 1006, 2611 (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.* Ex. 2003, ¶ 64.

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.” Ex. 1006, 2612 (emphasis added). Hussain derives the following equation for the MSC:

$$\overline{\text{MSC}}(f_k) = \frac{1}{l} \sum_{i=1}^l \text{MSC}(f_k, i)$$

Ex. 1006, 2613. 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). Ex. 2003, ¶ 65.

Hussain uses the MSC to choose between using an “intermittent coherent noise canceller” and an “incoherent noise canceller.” Ex. 1006, 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.” *Id.*, 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.” *Id.*, 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.* Ex. 2003, ¶ 66.

Hussain’s system uses an array of “[t]wo or more relatively closely spaced microphones” in an adaptive noise cancellation scheme. Ex. 1006 at 2611. 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 2612. These adaptive filters are applied in each subband and the resulting signals are reconstructed to yield the denoised speech signal. *Id.*, Fig. 1. Ex. 2003, ¶ 67.

Hussain assumes that an effective voice activity detector (VAD) is available for use in his system. Ex. 1006 at 2612. Hussain does not disclose any specific voice activity detector. Ex. 2003, ¶ 68.

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

U.S. Patent No. 6,006,175 (“Holzrichter”) (Ex. 1007) 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.

Ex. 2003, ¶¶ 69-72.

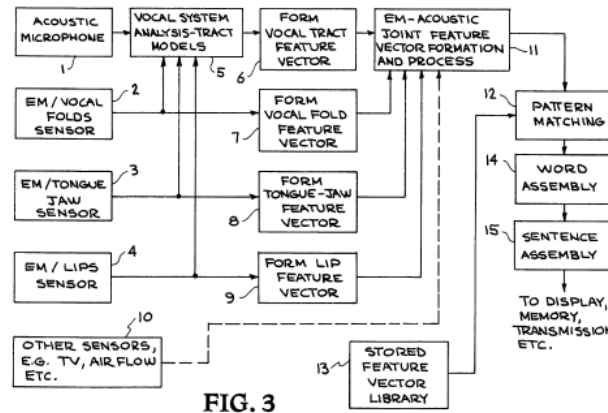


FIG. 3

Ex. 1007, FIG. 3.

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.* Ex. 2003, ¶ 70.

Holzrichter pattern matches the combined feature vector against a feature vector library to identify the sound represented by the feature vector. Ex. 1007, 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. Ex. 2003, ¶ 71.

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. Ex. 1007, 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. Ex. 2003, ¶ 72.

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

A Miniature Accelerometer for Detecting Glottal Waveforms and Nasalization ("Stevens") (Ex. 1008) 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. Ex. 1008 at 7. Ex. 2003, ¶ 73.

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

U.S. Patent No. 3,746,789 ("Alcivar") (Ex. 1009) 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. Ex. 2003, ¶ 74.

IV. CLAIM CONSTRUCTION

Claim construction is required regarding the terms “voice detection subsystem,” “voice activity signals,” “human voicing activity,” and “generating a noise waveform estimate associated with noise of the acoustic signals” because these constructions are dispositive of the issues presented in the Petition. *See Nidec Motor Corp. v. Zhongshan Broad Ocean Motor Co.*, 868 F.3d 1013, 1017 (Fed. Cir. 2017) (“Because we need only construe terms ‘that are in controversy, and only to the extent necessary to resolve the controversy,’ we need not construe [a particular claim limitation] where the construction is not ‘material to the . . . dispute.’” (citation omitted)).

The Board interprets the claims of the challenged patent:

[U]sing the same claim construction standard that would be used to construe the [claims] in a civil action under 35 U.S.C. 282(b), including construing the [claims] in accordance with the ordinary and customary meaning of such [claims] as understood by one of ordinary skill in the art and the prosecution history pertaining to the patent.

37 C.F.R. § 42.100(b) (2020). “In determining the meaning of [a] disputed claim limitation, we look principally to the intrinsic evidence of record, examining the claim language itself, the written description, and the prosecution history, if in

evidence.” *DePuy Spine, Inc. v. Medtronic Sofamor Danek, Inc.*, 469 F.3d 1005, 1014 (Fed. Cir. 2006). Claim terms are given their plain and ordinary meaning as would be understood by a person of ordinary skill in the art at the time of the invention and in the context of the entire patent disclosure. *Phillips v. AWH Corp.*, 415 F.3d 1303, 1313 (Fed. Cir. 2005) (*en banc*). “There are only two exceptions to this general rule: 1) when a patentee sets out a definition and acts as his own lexicographer, or 2) when the patentee disavows the full scope of a claim term either in the specification or during prosecution.” *Thorner v. Sony Comput. Ent. Am. LLC*, 669 F.3d 1362, 1365 (Fed. Cir. 2012).

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

Jawbone’s Construction	Petitioner’s Construction
“a subsystem that detects the presence of speech, including voiced and unvoiced speech” “signals that indicate speech activity, including voice and unvoiced speech activity” “human speech activity, including voiced speech and unvoiced speech”	No construction necessary.

The terms “voice detection subsystem,” “voice activity signals,” and “human voicing activity” should be construed to refer to subsystems, signals, and activity, respectively, that relate to both voiced and unvoiced speech. Voiced speech occurs

when the vocal cords are vibrating, and unvoiced speech occurs when the vocal cords are not vibrating. Ex. 2003, ¶¶ 37-38; *see also* Ex. 1007, 12:60-65, 32:44-51.

The intrinsic evidence compels a construction that is consistent with the specification's disclosure 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. Determining when there is noise only, *i.e.*, no speech, is required so that the transfer function that characterizes the noise signal can be correctly calculated. 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)." '543 Patent, 16:55-67 (emphasis added). Otherwise, if the voice detection system cannot detect certain types of speech, such as unvoiced speech, 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 all speech, including 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. Ex. 2003, ¶¶ 40-41.

The specification explains that when no distinction is made between voiced and unvoiced speech, then the terms “speech” and “voice” apply to both voiced and unvoiced speech. According to the specification, when the term “speech” or “voice” is used, “[u]nvoiced speech or voiced speech is distinguished where necessary.” ’543 Patent, 4:35-36. In discussing voice activity detection, the ’543 Patent does not distinguish between voiced and unvoiced speech. Ex. 2003, ¶ 40.

For these reasons, the Board should construe “voice detection subsystem” as a subsystem that detects the presence of speech, including voiced and unvoiced speech; “voice activity signals” as signals that indicate speech activity, including voice and unvoiced speech activity; and, “human voicing activity” as human speech activity, including voiced speech and unvoiced speech. Ex. 2003, ¶ 41.

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

Jawbone’s Construction	Petitioner’s Construction
“calculating a noise estimate based on the noise associated with all the acoustic signals”	No construction necessary.

The Board should construe “generating a noise waveform estimate associated with noise of the acoustic signals” as “calculating a noise estimate based on all the acoustic signals.” This limitation is recited in independent Claims 1 and 26. Ex. 2003, ¶¶ 42-58.

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

’543 Patent, Claim 1. Ex. 2003, ¶ 43.

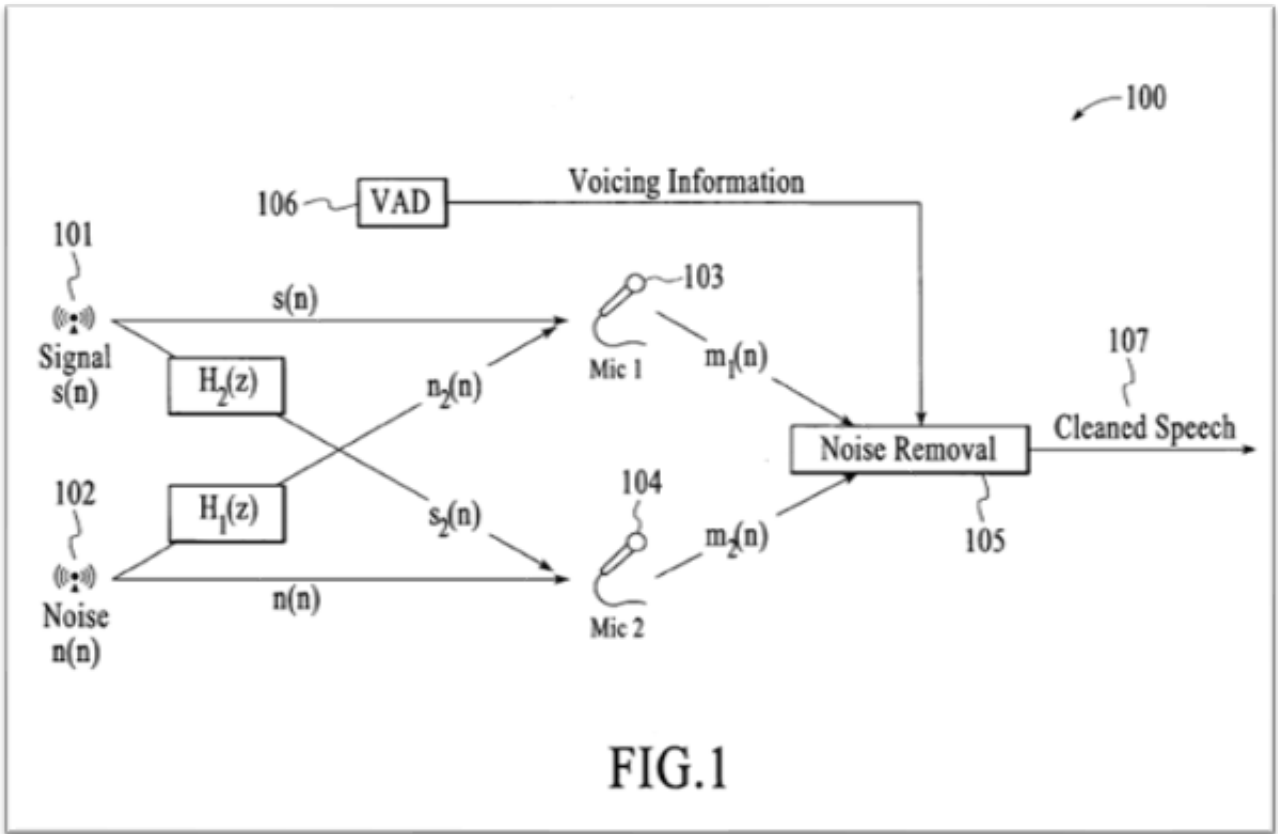
In other words, the noise waveform estimate is generated based on all the “acoustic signals,” while that generated waveform is subtracted from a single “acoustic signal.” Ex. 2003, ¶ 44.

Moreover, the surrounding claim language makes clear that the noise waveform estimate is not simply 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.” ’543 Patent, Claim 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. Ex. 2003, ¶ 45.

The specification similarly uses “acoustic signals” to include the signals coming *from* the microphones. *E.g.*, ’543 Patent, FIG. 1 (showing $m_1(n)$ and $m_2(n)$ output from the microphone); 14:52-54. The use of terms to denote an additional operation (generating) and a new element (the noise waveform estimate), demonstrates that the noise waveform estimate is both distinct from the acoustic signals and is the product of a subsequent operation. Ex. 2003, ¶ 46. A contrary decision would incorrectly render both “generating” and term meaningless. *Pause Tech., LLC v. TiVo, Inc.*, 419 F.3d 1326, 1334 (Fed. Cir. 2005) (“In construing claims, however, we must give each claim term the respect that it is due.”); *Rovi Guides, Inc. v. Vidal*, 2020-1994, 2022 WL 1152311, at *2 (Fed. Cir. Apr. 19, 2022) (“The available-to-a-user limitation requires the claimed system be configured to check whether an archived program is available to a *specific* user. A broader construction, which would only require the system be configured to check whether an archived program exists, would render the ‘to the user’ language meaningless.”).

FIG. 1 shows that the $m_1(n)$ and $m_2(n)$ acoustic signals emanate from the microphones:



As shown in FIG. 1, both Mic 1 and Mic 2 each receive both signal and noise.
'543 Patent, 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."). Ex. 2003, ¶¶ 47-48.

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

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

'543 Patent, 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. Ex. 2003, ¶ 49. In practical applications, $H_2(z)H_1(z)$ will be negligible. '543 Patent, 15:64-67. Thus, the signal is:

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

Id., 15:64-16:4. Ex. 2003, ¶ 50. $M_2(z)H_1(z)$ is thus the noise waveform estimate that is subtracted from the acoustic signal $M_1(z)$. *Id.*, ¶ 51.

$H_1(z)$ 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. *Id.*, ¶ 52.

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

Thus, the denoised signal can be calculated as:

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

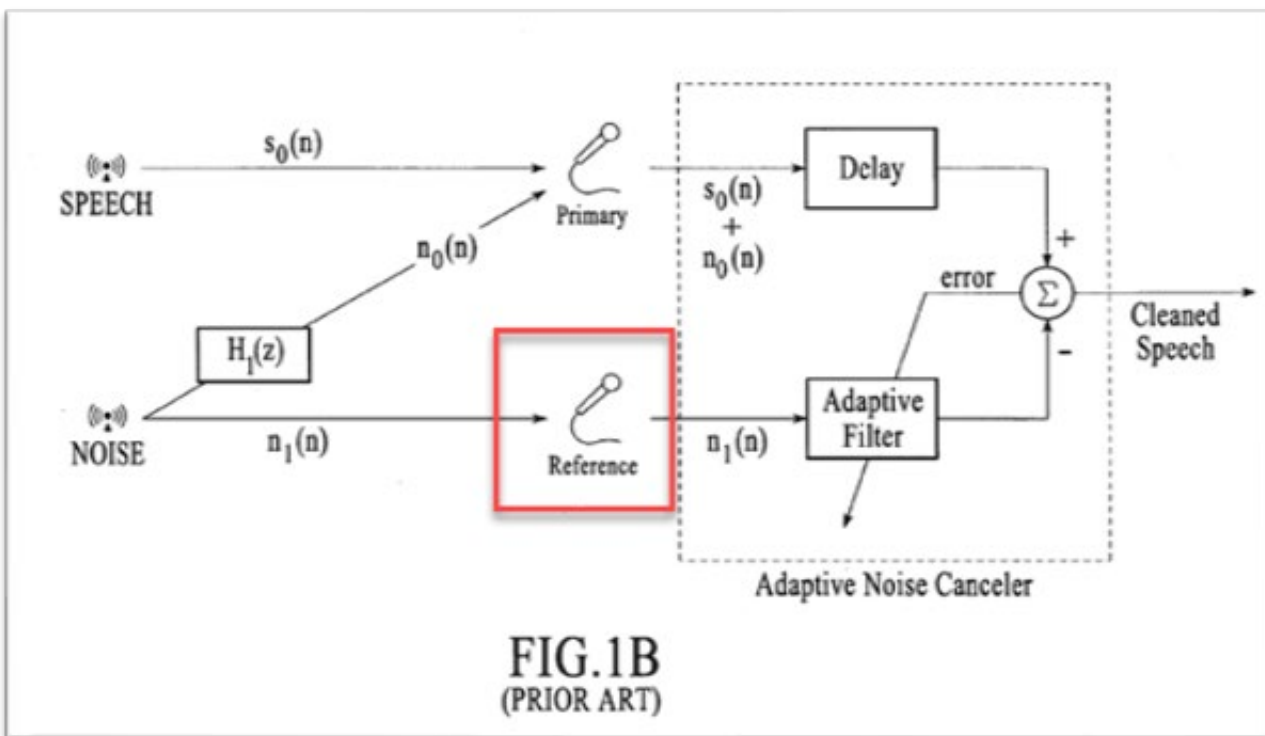
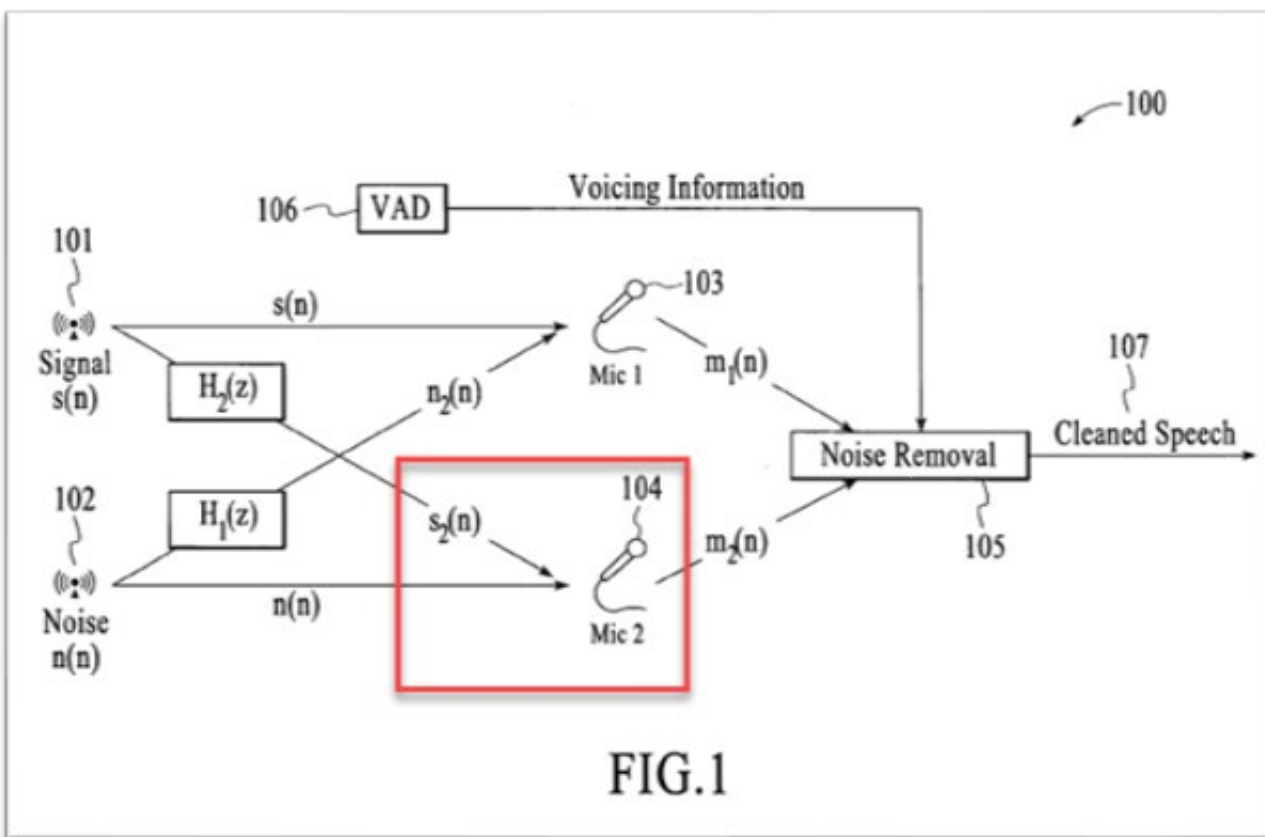
$$S(z) \approx M_1(z) - M_2(z) \frac{M_{1n}(z)}{M_{2n}(z)}$$

The denoised signal $S(z)$ is therefore calculated by subtracting the acoustic signal from MIC2 multiplied by $\frac{M_{1n}(z)}{M_{2n}(z)}$. The calculation of the noise estimate

waveform is thus based on the noise associated with all the acoustic signals. Ex. 2003, ¶¶ 53-54.

The specification further compels Jawbone’s construction because it 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. ’543 Patent, 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 of the invention over the prior art. *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.”). Allowing the microphones to be positioned closer to each other results in benefits. *Id.*, 17:28-31. (“This closer spacing simplifies the adaptive filter calculations and enables more compact microphone configurations/solutions.”). Ex. 2003, ¶ 55.

FIG. 1 compared with FIG. 1B demonstrates this key difference:



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. Ex. 2003, ¶¶ 56-57. Accordingly, the Board should construe “generating a noise waveform estimate associated with noise of the acoustic signals” as “calculating a noise estimate based on all the acoustic signals.”

V. LEVEL OF ORDINARY SKILL IN THE ART

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.” Ex. 2003, ¶¶ 25-26. Petitioner’s proposed level of skill is similar. Pet. at 4-5 (internal citations omitted).

VI. ARGUMENT

A. GROUND 1: Andrea, Hussain, and Holzrichter Do 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]

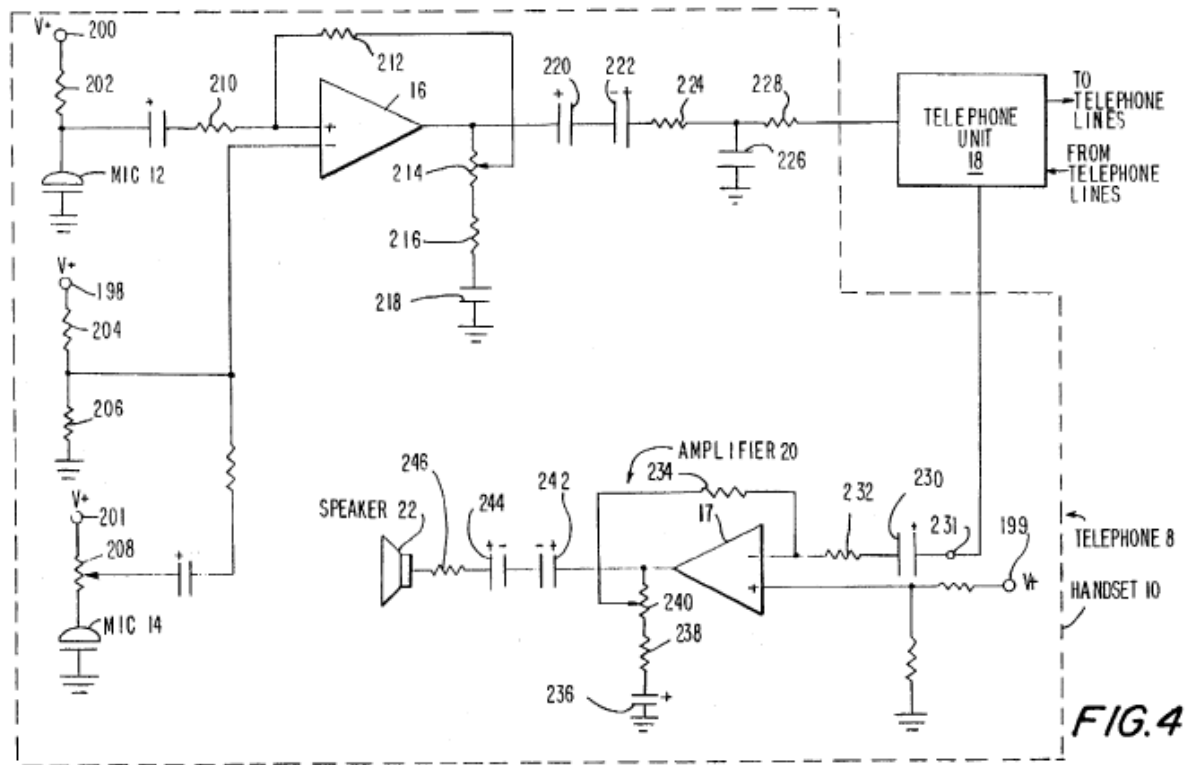
Petitioner argues that this limitation is disclosed or suggested by Andrea alone. Pet., 40-41. In particular, Petitioner states that "converting the background noise supplied to the second microphone into a corresponding electrical signal" is "generating a noise waveform estimate associated with noise 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.*

Petitioner is wrong. Under either the proper construction of the term, or simply the plain meaning as written, Andrea cannot teach or render obvious this limitation. In particular, Andrea does not "generate a noise waveform estimate" and the microphone signal it discloses is not associated with noise of the acoustic signals. Ex. 2003, ¶¶ 75-90.

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

As shown above, the Board should construe “generating a noise waveform estimate associated with noise of the acoustic signals” as “calculating a noise waveform estimate based on the noise associated with all the acoustic signals.” Ex. 2003, ¶ 76.

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. Ex. 2003, ¶ 77.



However, 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*”. '543 Patent, Claim 1. The only signal a microphone can provide (and a denoising subsystem can receive) is an electrical signal. Ex. 2003, ¶ 78.

The claims next require that the denoising method “generat[e]” a “noise waveform estimate.” The additional recitation of “generating” indicates that additional processing (a calculation) must be performed on the acoustic signal. *Rovi*

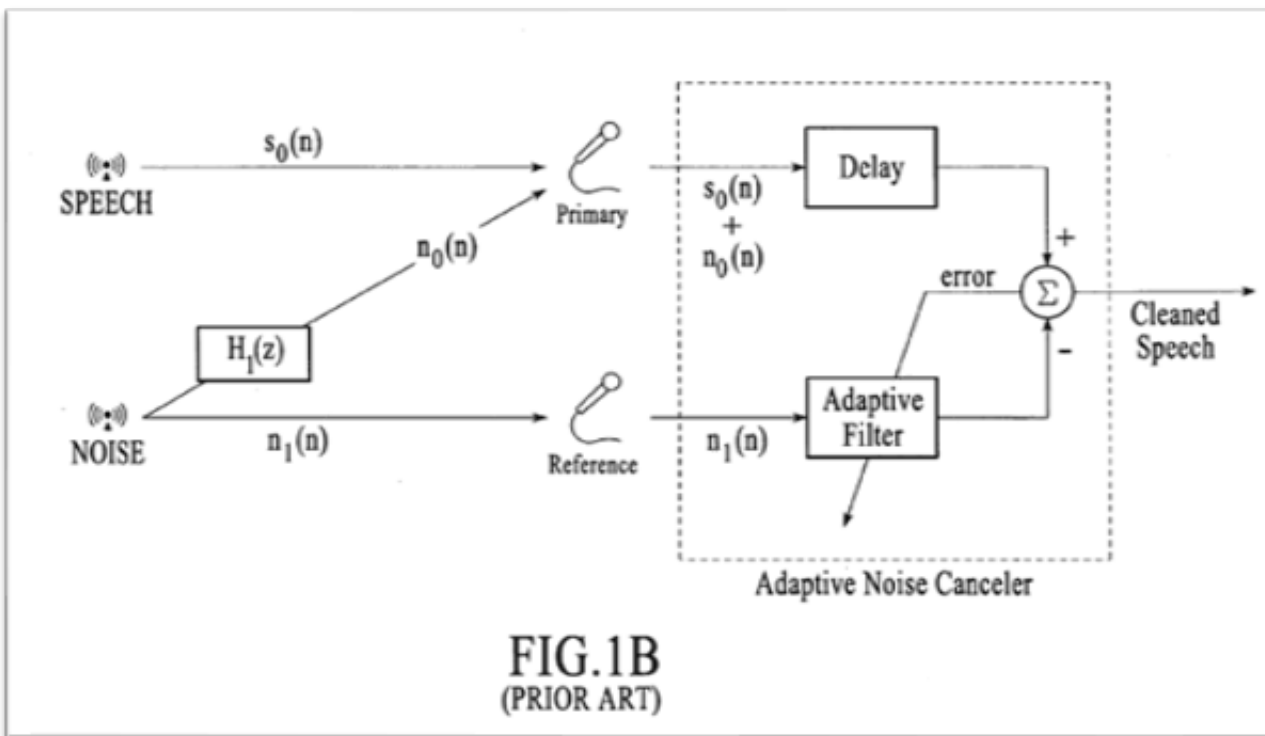
Guides, Inc., 2022 WL 1152311, at *2. Similarly, the use of a new term – “noise waveform estimate” – for the output of that processing indicates that the acoustic signal has been transformed. Ex. 2003, ¶ 79.

Andrea, however, simply passes an “acoustic signal”² to the op-amp. Ex. 2003, ¶ 80.

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 current-limiting resistor does not calculate a noise estimate waveform, but merely applies a gain to that same signal. Ex. 1005, 15:25-40. Ex. 2003, ¶ 81.

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. Ex. 2003, ¶ 82.

² “acoustic signal” refers to acoustic signal as used in the ’543 Patent’s claim language.



Accordingly, Meta’s combination does not disclose or render obvious this limitation under Jawbone’s proposed construction. Ex. 2003, ¶ 83.

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

Even if the Board were to decline to construe this term, Meta’s combination does not disclose it or render it obvious under the claim language as written. Ex. 2003, ¶¶ 84-90.

Meta’s combination does not perform any “generating” as required. Instead, as explained 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.” ’543 Patent, Claim 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. Ex. 2003, ¶ 85.

Moreover, the claims explicitly recite the noise waveform estimate to be 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. Ex. 2003, ¶ 86. Reading out claim terms in such a manner is improper. *Pause Tech., LLC*, 419 F.3d at 1334 (“we must give each claim term the respect that it is due”).

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, and is instead simply the sound picked up from one microphone. That signal includes both speech and noise. Ex. 2003, ¶ 87.

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. Thus the noise waveform estimate cannot simply be the signal from a microphone. Ex. 2003, ¶¶ 88-89.

Accordingly, even under the claim language as written, Petitioner's combination does not render this limitation obvious. Ex. 2003, ¶ 90.

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]

Meta relies on a combination of Andrea and Hussain to meet this limitation. However, 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. Ex. 2003, ¶ 91.

Andrea discloses an analog noise suppression apparatus, whereas Hussain discloses a more advanced digital noise suppression system. 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 2612. The resulting denoised

subband signals are reconstructed to yield the denoised speech signal. *Id.*, Fig. 1. Ex. 2003, ¶ 92.

A POSITA would not consider the noise-reduction processing of Andrea to be comparable or compatible 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. This combination was characterized by Petitioner’s expert, Dr. Anderson, 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. Ex. 2003, ¶ 93.

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 does not rely on Hussain for the subtraction limitation. Therefore, the Petition contemplates a combination of Andrea’s op-amp 16 for subtracting a noise signal from a speech-and-noise signal with Hussein’s digital subband adaptive filter solution. Ex. 2003, ¶ 94. On cross-examination, Dr. Anderson attempted to introduce a new argument that the Hussein portion of the combination could perform the subtraction. Ex. 2001,

31:3-32:22. However, this testimony conflicts with the Petition and Dr. Anderson's direct testimony, and the Board should not rely on it. Pet., 40-41; Ex. 1003, ¶¶ 132-133.

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 entail processing 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 2613. 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.

Ex. 2003, ¶ 95.

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. This limitation is therefore not rendered obvious by Meta's combination. Ex. 2003, ¶ 96.

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," as Required by Claims 1-2 and 5-26 [1I]

Petitioner's combination fails to disclose or render obvious this limitation because Andrea does not disclose or render obvious "using a sensor independent from the microphone array" and the combination does not disclose or render obvious "a voice detection subsystem." Ex. 2003, ¶¶ 97-106.

i. Andrea Does Not Teach This Limitation

Meta and its expert admit that Andrea fails to disclose a voice detection subsystem configured to receive voice activity signals "using a sensor independent from the microphone array." Pet. at 41; Ex. 2001, 22:7-24:3. 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. Accordingly, 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. Ex. 2003, ¶ 98.

ii. Holzrichter Does Not Teach or Render Obvious This Limitation

Meta further 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. *Id.* 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. Ex. 2003, ¶ 99.

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. Ex. 2003, ¶ 100.

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. Ex. 2003, ¶ 101.

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). 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. Ex. 2003, ¶ 102.

Contrary to Meta’s arguments, 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 2612, 2614, 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 2612. 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. Ex. 2003, ¶ 103.

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 2612. 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. Ex. 2003, ¶ 104.

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. Ex. 2003, ¶ 105.

Accordingly, Meta's combination does not disclose or render obvious this limitation.

4. Andrea, Hussain, and Holzrichter Do Not Disclose or Render Obvious Claim 26

Petitioner relies on the same arguments for limitations [26A]-[26I] as it does for Claim 1. For the same reasons as stated above, Petitioner's combination therefore does not render obvious Claim 26.

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

Petitioner relies on the combination of Andrea, Hussain, and Holzrichter with Stevens (Ground 2) or Alcivar (Ground 3) with respect to dependent Claims 3 and 4. Neither of the Grounds 2 and 3 combinations remedies the deficiencies of Ground 1. Accordingly, no claims are rendered obvious under Grounds 2 and 3.

VII. CONCLUSION

As discussed above, Petitioner's combinations fail to disclose key limitations of each claim of the '543 Patent, and therefore, do not render obvious any claims. The Board should, therefore, issue a Final Written Decision finding all Challenged Claims not unpatentable.

Respectfully submitted,

Dated: June 26, 2024

By: /Peter Lambrianakos /
Peter Lambrianakos (Reg. No. 58,279)
FABRICANT LLP
411 Theodore Fremd Avenue,
Suite 206 South
Rye, New York 10580
Tel. 212-257-5797
Fax. 212-257-5796
Email: plambrianakos@fabricantllp.com

CERTIFICATE OF WORD COUNT

The undersigned hereby certifies that the portions of the above-captioned JAWBONE INNOVATIONS, LLC'S RESPONSE specified in 37 C.F.R. § 42.24 has 6,836 words in compliance with the 14,000 word limit set forth in 37 C.F.R. § 42.24. This word count was prepared using Microsoft Word for Office 365.

June 26, 2024

By: /Peter Lambrianakos /
Peter Lambrianakos (Reg. No. 58,279)
Lead Counsel for Patent Owner
FABRICANT LLP
411 Theodore Fremd Avenue,
Suite 206 South
Rye, New York 10580
Tel. 212-257-5797
Fax. 212-257-5796

CERTIFICATE OF SERVICE

A copy of the foregoing Jawbone Innovations, LLC's Patent Owner's Response and Exhibits 2001 through 2005 have been served on Petitioner's counsel of record as follows:

Grace I. Wang
Email: PH-Meta-Jawbone-IPR@paulhastings.com
PAUL HASTINGS LLP
200 Park Avenue
New York, New York 10166

Lisa K. Nguyen
Eric E. Lancaster (*Pro Hac Vice* to be filed)
Email: PH-Meta-Jawbone-IPR@paulhastings.com
PAUL HASTINGS LLP
1117 South California Avenue
Palo Alto, California 94303

David M. Tennant
Alan M. Billharz
Email: PH-Meta-Jawbone-IPR@paulhastings.com
PAUL HASTINGS LLP
2050 M Street, N.W.
Washington, D.C. 20036

Attorneys for Meta Platforms, Inc.

June 26, 2024

By: /Peter Lambrianakos /
Peter Lambrianakos (Reg. No. 58,279)
Lead Counsel for Patent Owner
FABRICANT LLP
411 Theodore Fremd Avenue,
Suite 206 South
Rye, New York 10580
Tel. 212-257-5797
Fax. 212-257-5796