

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. 7,246,058
Filing Date: May 30, 2002
Issue Date: July 17, 2007

Inventor: Gregory C. Burnett
Title: DETECTING VOICED AND UNVOICED SPEECH USING
BOTH ACOUSTIC AND NONACOUSTIC SENSORS

PATENT OWNER'S RESPONSE

Case No. IPR2023-01352

TABLE OF CONTENTS

	<u>Page(s)</u>
I. INTRODUCTION	1
II. THE '058 PATENT	2
III. THE ALLEGED PRIOR ART.....	3
A. Meisel (U.S. Patent No. 4,718,096)	3
B. Kaneda (U.S. Patent No. 5,208,864).....	7
C. Freidin (U.S. Patent No. 6,148,313).....	7
D. Lyons.....	7
E. Sakoe (U.S. Patent No. 4,239,936).....	7
IV. LEVEL OF ORDINARY SKILL IN THE ART.....	8
V. CLAIM CONSTRUCTION	8
A. “voicing sensor”	9
B. “physiological information”	13
C. “cross-correlation data”	15
VI. ARGUMENT.....	19
A. GROUND 1: Meisel in View of Kaneda and Freidin Does Not Render Obvious Claims 1-2 and 4-5.....	19
1. Meisel Does Not Disclose “at least one voicing sensor that receives physiological information associated with human voicing activity,” as Required by Claim 1 and “receiving physiological information associated with human voicing activity at a voicing sensor,” as Required by Claim 2.....	19

2.	Petitioner’s Combination Does Not Disclose or Render Obvious “at least one processor. . .wherein the at least one processor generates cross correlation data between the physiological information and an acoustic signal received at one of the two microphones” and “generating cross correlation data between the physiological information and an acoustic signal received at one of the two receivers,” as Required by Claims 1 and 2, Respectively.....	26
3.	Petitioner’s Combination Does Not Disclose or Render Obvious “identif[ies/ying] information of the acoustic signals as voiced speech when the cross correlation data corresponding to a portion of the acoustic signal received at the one receiver exceeds a correlation threshold,” as Required by Claims 1 and 2	32
4.	Petitioner’s Combination Does Not Disclose or Render Obvious “[identifies/identifying] information of the acoustic signals as noise when the difference parameters are less than the gain threshold,” as Required by Claims 1 and 2	34
5.	Petitioner’s Combination Does Not Disclose or Render Obvious “wherein the voicing sensor includes at least one detector selected from a group including radio frequency devices, electroglottographs, ultrasound devices, acoustic throat microphones, and airflow detectors,” as Required by Claim 5.....	37
B.	GROUND 2: Meisel, in View of Kaneda, Freidin, and Lyons Does Not Render Obvious Claim 3.....	38
C.	GROUND 3: Meisel, in View of Kaneda, Freidin, and Sakoe Does Not Render Obvious Claim 4.....	38
VII.	CONCLUSION.....	38

TABLE OF AUTHORITIES

	Page(s)
Cases	
<i>Kyocera Senco Indus. Tools Inc. v. Int’l Trade Comm’n</i> , 22 F.4th 1369 (Fed. Cir. 2022).....	10
<i>Phillips v. AWH Corp.</i> , 415 F.3d 1303 (Fed. Cir. 2005) (<i>en banc</i>).....	10
<i>In re Suitco Surface, Inc.</i> , 603 F.3d 1255 (Fed. Cir. 2010).....	9
<i>Thorner v. Sony Comput. Ent. Am. LLC</i> , 669 F.3d 1362 (Fed. Cir. 2012).....	16, 19

EXHIBIT LIST

Exhibit No.	Description of Document
2001	Transcript of the Deposition of David S. Anderson, Ph.D., taken February 3, 2023 in IPR2022-00623
2002	Wave Walker DSP, Cross Correlation Explained With Real Signals, Dec. 1, 2021, available at https://www.wavewalkerdsp.com/2021/12/01/cross-correlation-explained-with-real-signals/
2003	Gardner, William A., Statistical Spectral Analysis: A Non-Probabilistic Theory, 1988
2004	Declaration of Akbar M. Sayeed, dated June 26, 2024
2005	Curriculum Vitae of Akbar M. Sayeed
2006	Transcript of the Deposition of David S. Anderson, Ph.D., taken June 11, 2024
2007	Merriam Webster Dictionary Definition of “Throat Microphone”
2008	Merriam Webster Dictionary Definition of “Laryngophone”
2009	Acker-Mills et al., Speech Intelligibility in Noise Using Throat and Acoustic Microphones, U.S. Army Aeromedical Research Laboratory, April 2004

I. INTRODUCTION

Petitioner Meta Platforms, Inc. (“Meta”) fails to show that any Challenged Claim of U.S. Patent No. 7,246,058 (Ex. 1001, the “’058 Patent”) would have been obvious because its combination fails to disclose a “voicing sensor” that “receives physiological information associated with human voicing activity” under the proper construction of those terms.

Petitioner also fails to show that the recited cross-correlation data would have been obvious. Instead of analyzing cross-correlation according to how a person of skill in the art (“POSITA”) would understand the term, Petitioner simply asserts that any comparison meets the cross-correlation limitations. But cross-correlation is a term of art with a specific meaning and requires at least multiplication and time shifting. Moreover, Meisel’s invention, which is directed toward examining *differences* between microphone signals, is intrinsically incompatible with cross-correlation, which examines similarities.

Accordingly, as explained in detail below and supported by the Declaration of Dr. Akbar M. Sayeed (Ex. 2004, “Sayeed”)¹, the Board should issue a Final Written Decision finding all Challenged Claims not unpatentable.

¹ Dr. Sayeed’s CV is provided as Exhibit 2005.

II. THE '058 PATENT

The '058 Patent describes systems and methods for detecting voiced and unvoiced speech in acoustic signals using physiological information. '058 Patent, Abstract. Previous methods “for classifying voiced and unvoiced speech have relied mainly on the acoustic content of microphone data, which is plagued by problems with noise and the corresponding uncertainties in signal content.” *Id.*, 1:34-37. Ex. 2004, ¶ 27.

In contrast to previous methods which rely on acoustic data, the '058 Patent utilizes a voicing sensor which receives non-acoustic physiological information. '058 Patent, 3:10-15 (discussing “non-acoustic sensor to enable detection of voiced speech in any environment”). While many different sensors may be usable as the voicing sensor, the invention requires a “sensor that is able to detect human tissue motion associated with speech and is unaffected by environmental acoustic noise.” *Id.*, 3:61-63. Ex. 2004, ¶ 28.

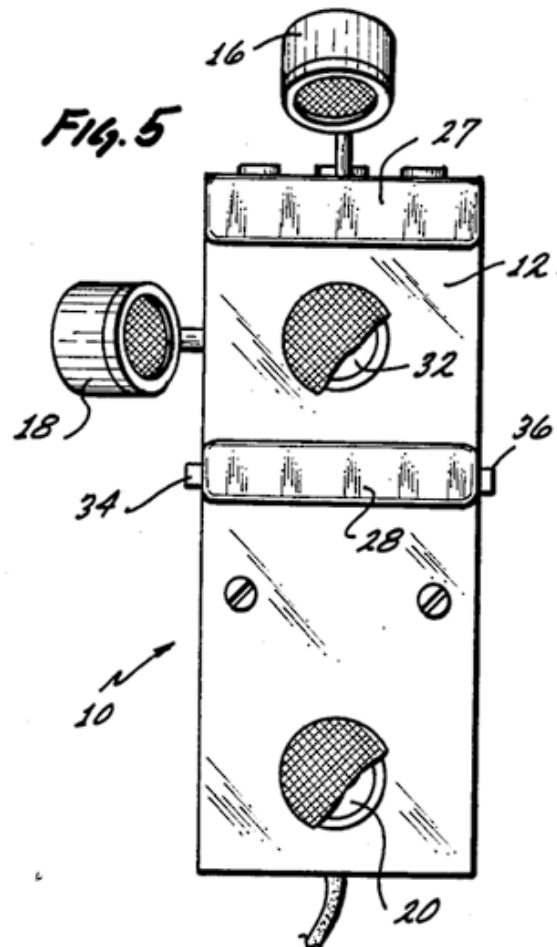
In particular, the '058 Patent receives acoustic signals at two microphones. '058 Patent, 3:29-38. The system also receives physiological information from the voicing sensor. *Id.*, 5:26-31, 5:46-65. By correlating the acoustic signals with the physiological information, the system can determine when a signal contains voiced speech. *Id.* The system also determines whether there is unvoiced speech by generating difference parameters between the acoustic signals received at the two

microphones. *Id.*, 7:52-8:48. When those parameters exceed a gain threshold, the signal is identified as unvoiced speech. *Id.* And when the parameters are beneath the threshold, the signal is identified as noise. *Id.* Ex. 2004, ¶ 29.

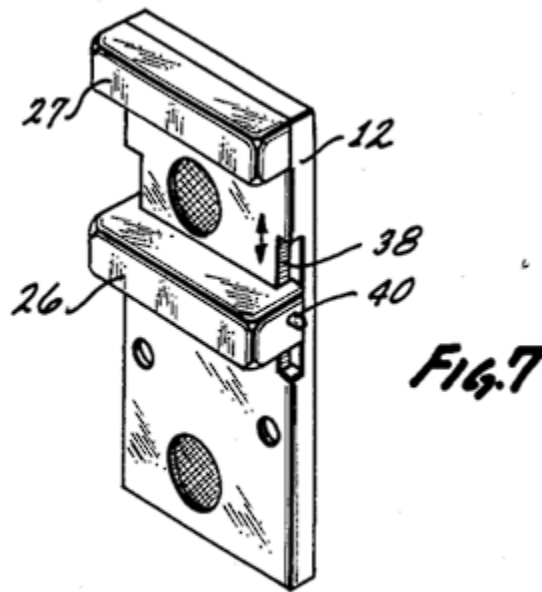
III. THE ALLEGED PRIOR ART

A. Meisel (U.S. Patent No. 4,718,096)

U.S. Patent No. 4,718,096 (Ex. 1005, “Meisel”) is directed to a speech recognition system. Ex. 1005, Abstract. In one embodiment, Meisel makes use of a system of four acoustic microphones: a nose microphone (16), throat microphone (20), mouth microphone (32), and side mouth microphone (18). *Id.*, 5:19-36. Ex. 2004, ¶ 30.

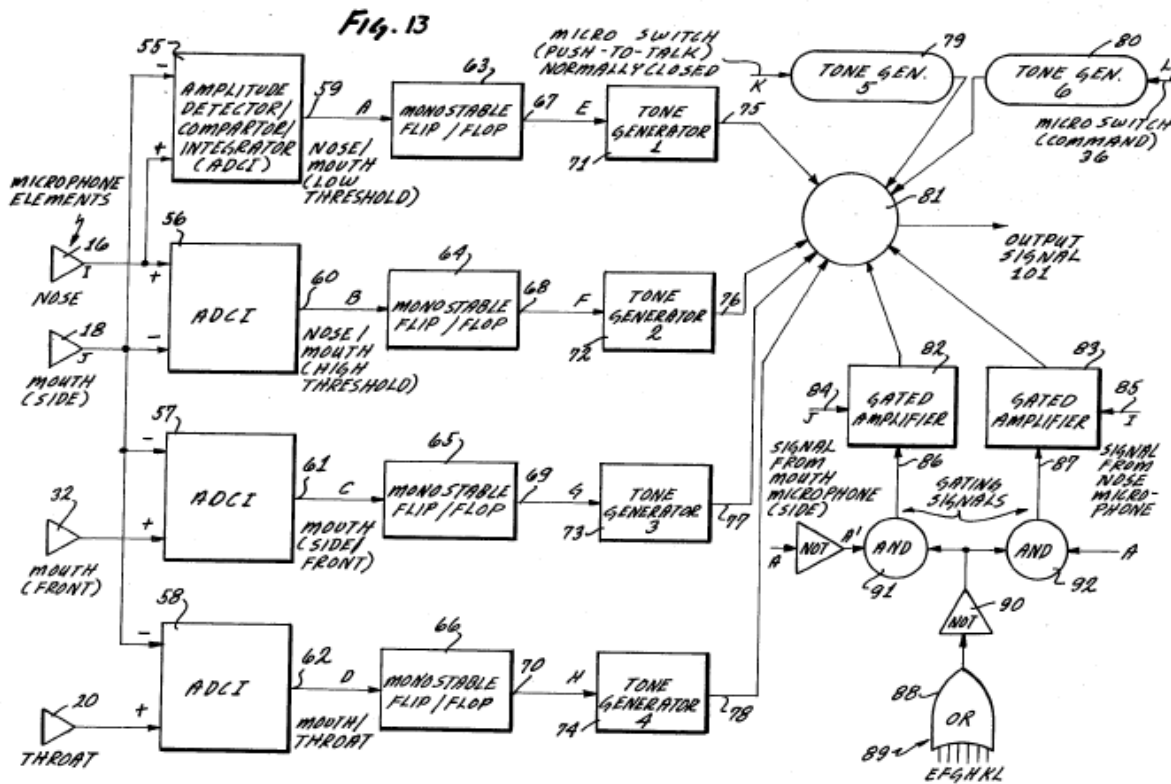


Ex. 1005, FIG. 5. The microphones are enclosed in a housing (12) which includes spacers (26, 27) that prevent the microphones from touching the user's skin. *Id.*, 5:37-46. The spacers rest, for example, on a user's upper lip and chin. *Id.* Ex. 2004, ¶ 30.



Ex. 1005, FIG. 7.

Meisel seeks to detect and classify sounds as speech. *Id.*, Abstract. To do so, Meisel compares the amplitudes of signals from its microphones to classify sounds as, for example, voiced or unvoiced stop consonants. Meisel uses the microphones' signals as inputs to ADCIs, which perform the amplitude comparison. Ex. 2004, ¶ 31.



Ex. 1005, FIG. 13.

Meisel uses its throat microphone to detect when a user pronounces certain consonant sounds. *Id.*, 2:31-53. In particular, when a user pronounces stop consonants, the vocal tract is closed briefly. *Id.*, 2:39-41. When the vocal tract is closed, “little sound comes from the mouth.” *Id.* However, the throat microphone can detect acoustic sounds emanating from the throat. By comparing the mouth microphone and throat microphone signals, the system can detect voiced stop consonants when the mouth signal is low, but the throat signal is high. *Id.*, 2:45-48. In such cases, signal 62 in FIG. 13 will be logically true. *Id.*, 9:60-10:11. Sayeed, ¶ 32.

B. Kaneda (U.S. Patent No. 5,208,864)

U.S. Patent No. 5,208,864 (Ex. 1006, “Kaneda”) is directed to a method of detecting an acoustic signal. Ex. 1006, 1:6-9. Kaneda uses “two sound receiving units for generating signals having different S/N ratios” that “are located at a single position.” *Id.*, 5:16-24. “[A] speech period is detected by using a power difference between the two output signals.” *Id.*, 5:22-24. Ex. 2004, ¶33.

C. Freidin (U.S. Patent No. 6,148,313)

U.S. Patent No. 6,148,313 (Ex. 1007, “Freidin”) is directed to a “programmable digital correlator device, and associated correlation method, with a very efficient structure.” Ex. 1007, Abstract. Freidin explains that “Correlators have been widely used in signal processing to measure the degree of correspondence between a data sequence and a predetermined pattern or reference sequence.” *Id.*, 1:11-20. Ex. 2004 Sayeed, ¶ 34.

D. Lyons

Understanding Digital Processing by Lyons (Ex. 1012, “Lyons”) is a textbook purportedly published in 1997.

E. Sakoe (U.S. Patent No. 4,239,936)

U.S. Patent No. 4,239,936 (Ex. 1008, “Sakoe”) is directed to a speech recognition system for use in noisy environments. Ex. 1008, Abstract. Sakoe makes use of two microphones. *Id.*, FIG. 1A, 1:59-2:39. One microphone is used to pick

up speech, another is used to receive ambient noise. *Id.* Sakoe includes a recognition system to recognize speech. *Id.* Sakoe also includes a rejection unit that rejects the recognition results when excessive noise is detected. *Id.* Ex. 2004, ¶ 36.

IV. LEVEL OF ORDINARY SKILL IN THE ART

The parties agree on the level of skill in the art—“a minimum of a bachelor’s degree in computer engineering, computer science, electrical engineering, mechanical engineering, or a similar field, and approximately three years of industry or academic experience in a field related to acoustics, speech recognition, speech detection, or signal processing. . . . Work experience can substitute for formal education and additional formal education can substitute for work experience.” Petition, Paper 1 (“Pet.”) at 7; Ex. 2004, ¶¶ 25-26.

V. CLAIM CONSTRUCTION

Petitioner asserts that no terms require construction. Pet. at 7. However, the Petition uses: i) “voicing sensor” and contrary to its definition within the ’058 Patent’s specification; and ii) “physiological information” and “cross-correlation data” contrary to their plain and ordinary meaning. The Board should therefore construe those terms as described below.

A. “voicing sensor”

The term “voicing sensor” appears in Clams 1, 2, and 5 of the ’058 Patent. The Board should construe it as “a sensor that is able to detect human tissue motion associated with speech and is unaffected by environmental noise.” Ex. 2004, ¶ 37.

Jawbone’s construction comes directly from the specification, which states that “[a]lternative embodiments, however, may use any sensor that is able to detect human tissue motion associated with speech and is unaffected by environmental acoustic noise.” ’058 Patent, 3:61-63. Ex. 2004, ¶ 38. Indeed, every embodiment in the patent, “including those using GEMS sensors, ultrasound devices, acoustic throat microphones, and airflow detectors, utilizes a non-acoustic sensor, which is unaffected by environmental noise. *Id.*, ¶ 39. The preferred embodiment, for example, uses the General Electromagnetic Movement Sensor (GEMS). *E.g.*, ’058 Patent, 3:57-63. “The GEMS is a radio frequency device (2.4 GHz) that allows the detection of moving human tissue dielectric interfaces.” *Id.*, 3:64-66. As the ’058 Patent explains, the GEMS “sends out weak electromagnetic waves (less than 1 milliwatt) that reflect off of whatever is around the sensor. The reflected waves are mixed with the original transmitted waves and the results analyzed for any change in position of the targets.” *Id.*, 4:1-5; *see also id.*, 3:64-4:13. Ex. 2004, ¶ 39. A POSITA would thus understand that the voicing sensor must be unaffected by environmental noise and the Board should construe the claim accordingly. *In re*

Suitco Surface, Inc., 603 F.3d 1255, 1260 (Fed. Cir. 2010) (“[C]laims should always be read in light of the specification and teachings in the underlying patent.”).

The '058 Patent also explicitly defines the voicing sensor as a non-acoustic sensor, which would necessarily be unaffected by environmental noise. '058 Patent, 4:55-57 (“the non-acoustic sensor (or *hereafter just the sensor*)” (emphasis added)). It also refers to the voicing sensor as a non-acoustic sensor. *E.g., id.*, 2:32-34 (discussing “voicing activity detectors or non-acoustic sensors”); 3:1-28 (discussing use of “non-acoustic voicing sensor”); *id.* at 3:50-4:13; 5:63-65; Ex. 2004, ¶ 40. The patent’s definition is controlling. *Kyocera Senco Indus. Tools Inc. v. Int’l Trade Comm’n*, 22 F.4th 1369, 1378-79 (Fed. Cir. 2022) (“Because the patentee clearly defined [claim term at issue] in the written description, that definition controls.”); *Phillips v. AWH Corp.*, 415 F.3d 1303, 1316 (Fed. Cir. 2005) (*en banc*) (holding that “the inventor’s lexicography governs”).

The '058 Patent refers to the sensor as unaffected by environmental noise for good reason. Indeed, using a non-acoustic sensor that is unaffected by environmental noise is “required to ensure good performance.” *E.g., '058 Patent*, 4:44-46 (noting that “external noise is not a problem” when using a non-acoustic sensor); 4:55-57 (“the non-acoustic sensor (or hereafter just the sensor) will be *required to ensure good performance.*” (emphasis added)). Ex. 2004, ¶ 41. A POSITA would understand this to be a requirement of all voicing sensors as claimed

in the '058 Patent, particularly given the distinction drawn between voicing sensors (which receive physiological information) and microphones (which receive acoustic signals) in the claims. *Id.*

Claim 5 of the '058 Patent recites that “wherein the voicing sensor includes at least one detector selected from a group including radio frequency devices, electroglottographs, ultrasound devices, acoustic throat microphones, and airflow detectors.” However, the listing of “acoustic throat microphones” does not remove the requirement of being unaffected by environmental noise. *Id.*, ¶ 42. A POSITA would understand that a throat microphone is (and must be) unaffected by environmental noise, by touching the skin of a user to sense vibration. *Id.*

A throat microphone, also known as a laryngophone, is a well-known term of art for a type of skin surface microphone, with which a POSITA would be familiar. *Id.*, ¶ 43. Throat microphones were originally developed for military aircraft pilots operating in extremely high noise environments, and operate using a vibration receiving diagram in contact with a user’s throat over the larynx, from which the microphone receives vibrations directly. *Id.* This contrasts with a conventional acoustic microphone which receives acoustic waves through air molecules. *Id.* Dr. Sayeed confirms that the Merriam Webster definitions attached hereto as Exs. 2007 and 2008 would be consistent with a POSITA’s understanding of the ordinary meaning of a throat microphone, as it is recited in the '058 Patent. *Id.* Dr. Sayeed

further confirms that the Speech Intelligibility in Noise Using Throat and Acoustic Microphones (Ex. 2009) 2004 publication by the U.S. Army Aeromedical Research Laboratory, further discusses throat microphones consistent with a POSITA's understanding, noting that "[t]hroat microphones were developed in response to [noise issues] during the later years of WWII," and that "since acoustic microphones convert sound energy (spoken words) into electrical energy to be transmitted into the communication system, they must necessarily also transmit any ambient noise around the talker's mouth. Removing this ambient noise from the communication signal should improve the signal-to-noise ratio for the listener. *Throat microphones are designed to pick up (transduce) the vibrations of the vocal apparatus at the throat instead of the vibrations of air molecules at the mouth.*" *Id.* (citing Ex. 2009 at 8).

Moreover, a POSITA would not understand the '058 Patent's recitation of a throat microphone as "acoustic" in dependent Claim 5 to suggest it must be susceptible to environmental acoustic noise, as opposed to claiming a throat microphone, conventionally used to detect acoustic waves through skin, which is implemented to detect human tissue motion within the scope of the claims. Ex. 2004, ¶ 44. A POSITA would recognize that the primary purpose of a throat microphone is to avoid environmental acoustic noise. A POSITA with the benefit of the '058 Patent specification would further recognize that human tissue motion

corresponding with voicing activity, distinct from acoustic signals, could be sensed by the diaphragm of an acoustic throat microphone in the same manner as with a GEMS sensor.

A POSITA's understanding that an "acoustic throat microphone" is "a sensor that is able to detect human tissue motion associated with speech and is unaffected by environmental noise" would be reinforced by the claim language itself, which draws a distinction between "acoustic signals" received by microphones and "physiological information" received by voicing sensors (such as acoustic throat microphones). *Id.*, ¶ 45 (citing Claims 1 and 2). The '058 Patent plainly recites a voicing sensor as something other than a conventional microphone, and a POSITA would understand that is because it must detect human tissue motion associated with speech, as opposed to acoustic signals which would be affected by environmental noise. *Id.*

The Board should therefore construe "voicing sensor" to mean "a sensor that is able to detect human tissue motion associated with speech and is unaffected by environmental noise." *Id.*, ¶ 46.

B. "physiological information"

The term "physiological information" appears in Claims 1 and 2 of the '058 Patent. The Board should construe it as "information on human tissue motion." *Id.*, ¶ 47.

Jawbone’s construction similarly comes directly from the specification, which explains that the type of “physiological information” which voicing sensors detect is “human tissue motion.” ’058 Patent, 3:61-63 (“Alternative embodiments, however, may use any sensor that is able to detect human tissue motion associated with speech and is unaffected by environmental acoustic noise.”); *id.*, Abstract (“embodiments of the system include non-acoustic sensors that receive physiological information to aid in identifying voiced speech”); *id.*, 3:64-4:13 (“The GEMS is a radio frequency device (2.4GHz) that allows the detection of moving human tissue. . . . Anything that *moves* near the sensor will cause a change in phase of the reflected wave that will be amplified and displayed. . . . A similar sensor is described in [] “The physiological basis of glottal electromagnetic micropower sensors (GEMS) and their use in defining an excitation function for the human vocal tract”. . .); *see also id.* at 5:11-14, 5:26-27). Ex. 2004, ¶ 48.

The claim language itself further supports Jawbone’s construction by drawing a distinction between “physiological information” and “acoustic signals.” For example, Claim 1 recites “at least two **microphones** that receive the **acoustic signals**; at least one voicing sensor that receives physiological information associated with human voicing activity . . . [wherein the at least one processor] generates cross correlation data *between the physiological information and an acoustic signal.*” ’058 Patent, Claim 1 (emphasis added); *see also* Claim 2. A

POSITA would understand that the patentee plainly used the term “physiological information” distinctly from “acoustic signals” to mean different things. Ex. 2004, ¶ 49. A POSITA reading the claims in light of the specification would understand physiological information to refer to human tissue motion, is received independent of acoustic signals in the same manner. *Id.* (citing ’058 Patent, 3:61-63, 3:64-4:13, 5:11-14, 5:26-27).

In addition to being supported by the claims and specification, Jawbone’s construction is necessary because Petitioner incorrectly reads “physiological information” to include acoustic signals received in a conventional microphone. This reading would erase the distinction between “physiological information” and “acoustic signals” claimed by the patentee, and Jawbone’s construction is necessary to preserve that distinction in the face of Meta’s reading of the claims.

The Board should therefore construe “physiological information” to mean “information on human tissue motion.” Ex. 2004, ¶ 50.

C. “cross-correlation data”

The ’058 Patent uses the plain and ordinary meaning of “cross-correlation data”: “data generated by multiplying and adding the result of a windowed set of samples at one time from one source by a windowed set of samples from another source where the two sets of samples may be shifted in time relative to each other.” As Petitioner and Dr. Anderson acknowledge, “cross-correlation” is well-known in

the art. Pet. at 38-39; Ex. 1003, ¶¶ 59-60. The Board should afford the term its customary meaning. *Thorner v. Sony Comput. Ent. Am. LLC*, 669 F.3d 1362, 1365 (Fed. Cir. 2012) (“The words of a claim are generally given their ordinary and customary meaning as understood by a person of ordinary skill in the art when read in the context of the specification and prosecution history.”)

A POSITA would understand that cross-correlation function is given by:

$$R_{xy}[\tau] = \sum_n x[n]y^*[n - \tau].$$

Ex. 2004, ¶ 53 (citing Ex. 2004, 2005). Dr. Anderson uses essentially the same formula in his Declaration:

$$r_{x,y}(\ell) = \sum_{n=-\infty}^{\infty} x(n)y(n - \ell)$$

Ex. 1003, ¶ 60 (quoting Ex. 1016 at 20).

The formula tracks Patent Owner’s construction. The cross-correlation R is obtained by summing multiples of samples from x at time n with samples from y at time n – τ. *Id.* The “τ” controls the time shifting in Patent Owner’s construction. *Id.*

Petitioner provided multiple exhibits, all of which use variations of the above definition. *Id.*, ¶¶ 55-56. For example, Ex. 1016 provides precisely the same formula. Ex. 1016 at 20. Ex. 1023 shows cross-correlation for a two-dimensional

image, but uses the same basic concept to multiply sample g at indices i and j by sample t at indices $(i-m)$ and $(j-n)$. Ex. 1023 at 279. Ex. 1011 similarly describes cross-correlation. Ex. 1011 at 152 (“Next, the indicated samples from the received signal fall into the correlation machine, and are multiplied by the corresponding points in the target signal.”). Ex. 1009, “Signal Processing Toolbox for use with MATLAB” similarly defines cross-correlation in the context of the XCORR function:

The true cross-correlation sequence is

$$\gamma_{xy}(m) = E\{x_n y_{n+m}^*\}$$

where x_n and y_n are stationary random processes, $-\infty < n < \infty$, and $E\{\}$ is the expected value operator. `xcorr` must estimate the sequence because, in practice, only a finite segment of the infinite-length random process is available.

See also Ex. 1016 at 14. Ex. 2004, ¶¶ 55-56.

A POSITA, reading the '058 Patent, would know the above definitions of cross-correlation. *Id.*, ¶ 58. The '058 Patent's use of cross-correlation and reference to XCORR would further inform a POSITA that “cross-correlation” was being used in its typical meaning as being generated by multiplying two samples together, with those samples being at the same or a different time and adding up this series of multiplications over a finite time window. *Id.*

The '058 Patent explains that its use of cross-correlation (XCORR) requires the ability to shift the samples in time. '058 Patent, 5:11-14 (“The XCORR is only

calculated to 15 delays, which corresponds to just under 2 milliseconds at 8000 Hz.”); Ex. 2004, ¶ 59. The ’058 Patent explains that this is because the voicing sensor data arrives earlier than the acoustic signals, which are bound by the speed of sound. ’058 Patent, 5:31-38 (“The acoustic data is expected to lag behind the sensor data by about 0.1 to 0.8 milliseconds (or about 1-7 samples) as a result of the delay time due to the relatively slower speed of sound (around 330 m/s).”); Ex. 2004, ¶ 59.

When deposed regarding statements in an identical declaration he submitted in IPR2022-00623, Dr. Anderson further confirmed that a POSITA would understand that “XCORR”, which the ’058 Patent identifies as cross correlation, involves “multiplying a sample from one set of samples [] by a sample from another set of samples”. Ex. 2001, Dr. Anderson Dep., Feb. 3, 2023, 28:18-32:24. Dr. Anderson confirmed that he has not formed any new opinions since that time and that the statements he made in that deposition were true to the best of his knowledge. Ex. 2006, Dr. Anderson Dep., June 11, 2024, 9:22-10:3. Dr. Anderson further confirmed his understanding that “XCORR” is “used in some programming languages such as MATLAB.” Dr. Anderson also stated that “cross-correlation was well understood in the art,” while himself representing cross-correlation with a formula that involves multiplying and adding the result of a windowed set of samples at one time from one source by a windowed set of samples from another source in

the same manner as Jawbone’s proposal. *Compare Id.*, 13:7-14:11 with Ex. 1003, ¶ 60 (quoting Ex. 1016 at 20).

Accordingly, all of the intrinsic and extrinsic evidence confirms that “cross-correlation data” means “data generated by multiplying and adding the result of a windowed set of samples at one time from one source by a windowed set of samples from another source where the two sets of samples may be shifted in time relative to each other.” The Board should therefore construe the term as Patent Owner requests. *Thorner*, 669 F.3d at 1365.

VI. ARGUMENT

A. **GROUND 1: Meisel in View of Kaneda and Freidin Does Not Render Obvious Claims 1-2 and 4-5**

1. **Meisel Does Not Disclose “at least one voicing sensor that receives physiological information associated with human voicing activity,” as Required by Claim 1 and “receiving physiological information associated with human voicing activity at a voicing sensor,” as Required by Claim 2²**

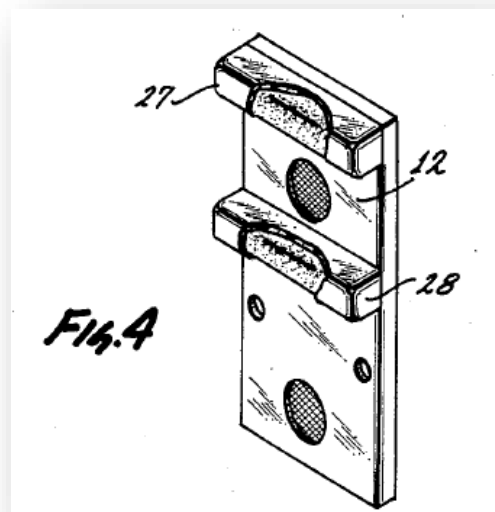
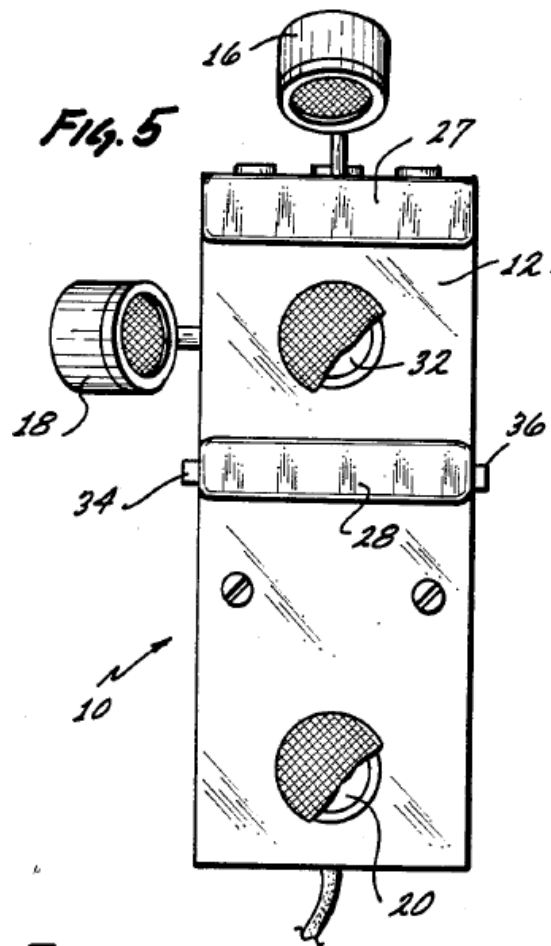
Meisel does not disclose or render obvious “at least one voicing sensor that receives physiological information associated with human voicing activity,” as required by Claim 1, nor does it disclose “receiving physiological information

² Petitioner relies on its Claim 1 arguments for all limitations of Claim 2. Pet. at 65-66. Thus, both independent claims are addressed together herein.

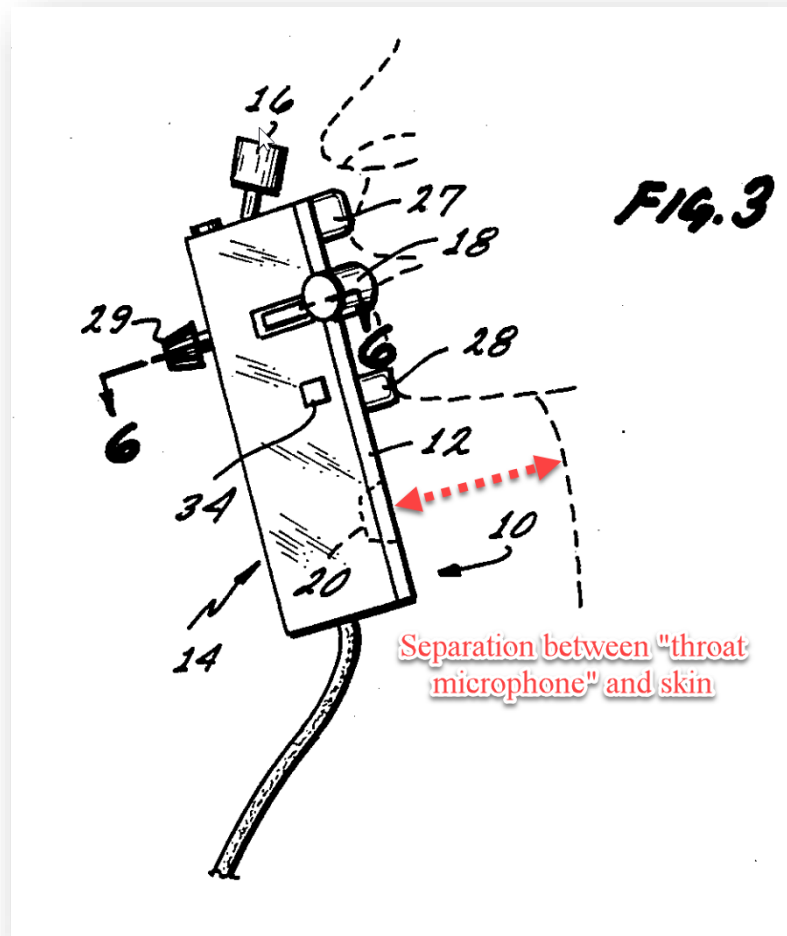
associated with human voicing activity at a voicing sensor,” as required by Claim 2. Petitioner argues that Meisel’s throat microphone discloses a voicing sensor. Pet. at 21. Petitioner does not raise any obviousness arguments with respect to this limitation. *See id.*

As discussed above, “voicing sensor” should be construed as “a sensor that is able to detect human tissue motion associated with speech and is unaffected by environmental noise,” and “physiological information” should separately be construed as “human tissue motion information.” Meisel’s throat microphone does not meet either construction because it is merely a conventional microphone pointed at a throat – it detects acoustic signals through air, not “human tissue motion,” and it is affected by environmental noise. Ex. 2004, ¶¶ 61-63. Even in the absence of a construction, Meisel does not disclose receiving “physiological information” under any fair reading of the term’s ordinary meaning.

Meisel’s throat microphone is illustrated in FIGS. 4 and 5 of the ’058 Patent, reproduced below.



Ex. 1005, FIGS. 4, 5. Meisel's "throat microphone" is identifier 20 in Fig. 5 and shown, but not labeled in Fig. 4. Ex. 2004, ¶¶ 64-65. The spacers 27 and 28 rest against the user's upper lip and chin, respectively. Ex. 1005 at 5:37-46. Ex. 2004, ¶ 65; Ex. 2008, 40:11-41:18. There will, therefore, be a substantial separation between the user's skin and the throat microphone, as illustrated in FIG. 3 (excerpted below). Ex. 2004, ¶ 65 (citing Ex. 1005, FIG. 3). This is consistent with Meisel's description of its throat microphone as "adjacent" to the "throat of the speaker" rather than in contact with the speaker's skin. Ex. 1005, 5:32-36. Meisel does not disclose any embodiments where the throat microphone is on (or even in close proximity to) the user's skin. Ex. 2004, ¶ 65.



Ex. 1005 FIG. 3, annotated by Dr. Sayeed.

Meisel's throat microphone does not receive physiological information in the form of "human tissue motion," as required by both constructions, and a POSITA would recognize as much. Ex. 2004, ¶ 66. Meisel's throat microphone is described as a conventional microphone pointed at a speaker's throat as depicted above, and is described as receiving *sounds* (*i.e.*, acoustic signals), not motion of human tissue. See Ex. 1005, 8:55-59 ("Similar proof can be provided of the importance of

recording the *sounds* emanating from the throat . . .”) (emphasis added). Indeed, Meisel describes that “it is a further aspect of this invention to allow close talking microphones to be used,” and even Dr. Anderson conceded that this “provide[s] guidance as to what sort of microphones would be used in Meisel.” Ex. 2001, 47:8-24. A POSITA would understand that the conventional close-talking microphones disclosed in Meisel detect acoustic sound waves through air, not human tissue movement, especially when they are not even in contact with human tissue as disclosed in Meisel. Ex. 2004, ¶ 66. Even without an explicit construction of “physiological information,” a POSITA would understand that the acoustic signals received by Meisel’s throat microphone are not physiological information. *Id.* For example, a POSITA would recognize that the claims themselves distinguish between “acoustic signals” of the sort received by Meisel’s “throat microphone,” and “physiological information” which requires information about to a user’s physiology. *Id.*

To the extent that Petitioner argues that Meisel’s “throat microphone” is a type of “voicing sensor,” as claimed in the ’058 Patent based on Claim 5’s recitation of “wherein the voicing sensor includes at least one detector selected from a group including . . . acoustic throat microphones,” a POSITA would not recognize it as a throat microphone within the ordinary meaning of the term (as recited in the ’058 Patent). *Id.*, ¶ 67. As discussed above in Section V.A, a POSITA would recognize

that the '058 Patent recites an acoustic throat microphone consistent with the well-understood ordinary use of that term in the art. This contrasts with the usage of the term “throat microphone” in Meisel, which simply refers to an acoustic microphone pointed at a throat. *Id.*

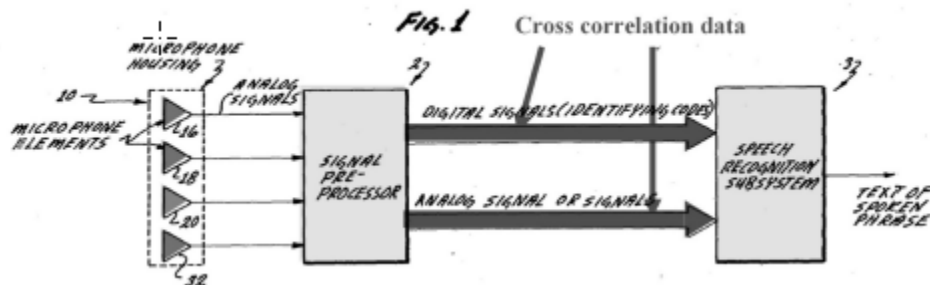
Moreover, a POSITA would understand that the “throat microphone” described in Meisel, is affected by environmental noise. *Id.*, ¶ 68. The Figures above show that the portion of housing 10 containing the throat microphone are open on the sides and below, which would allow for substantial environmental noise to reach the throat microphone. *Id.* Meisel further contemplates Substantial space between Microphone 20 and the user’s throat, as shown in Fig. 3. *Id.* A POSITA would understand that a close talking microphone, such as the one used in Meisel, would be affected by environmental noise, particularly given that it is open to the air. Given the spacing between the throat microphone and the user’s skin, and the openings to both sides and below the microphone, a POSITA would understand that the throat microphone receives acoustic signals generated by environmental noise sources. *Id.* The throat microphone therefore is not “unaffected by environmental noise.”

Claims 1 and 2 are therefore not rendered obvious by Petitioner’s combination.

2. Petitioner’s Combination Does Not Disclose or Render Obvious “at least one processor. . . wherein the at least one processor generates cross correlation data between the physiological information and an acoustic signal received at one of the two microphones” and “generating cross correlation data between the physiological information and an acoustic signal received at one of the two receivers,” as Required by Claims 1 and 2, Respectively

Petitioner’s combination does not disclose or render obvious these limitations because Meisel does not generate cross-correlation data and a POSITA would not combine Freidin with Meisel. Ex. 2004, ¶ 71.

Petitioner argues that Meisel’s statement that “[t]he pre-processor 2 processes the signals from the microphones 16, 18, 20 and 32 to determine at each instant imbalances in the relative amplitude of each microphone,” discloses the necessary cross correlation data. Pet. at 25 (citing Meisel, 4:35-42, FIG. 1).



Pet. at 25.

Meisel explains that “[t]he pre-processor 2 processes the signals from the microphones 16, 18, 20 and 32 to determine at each instant imbalances in the relative amplitude of each microphone.” Ex. 1005, 4:35-39. Ex. 2004, ¶ 72. Meisel’s

preprocessor passes those imbalances as Boolean (that is, true or false) values to the speech recognition system. Ex. 1005, 4:39-42. Ex. 2004, ¶ 72. As Meisel further explains, the preprocessor generates the imbalance data through the use of an ADCI (amplitude detector/comparator/integrator). Ex. 1005 at 8:60-9:2; Ex. 2004, ¶ 73. The ADCI performs a simple compare operation between two inputs by inverting the second input and subtracting it from the first. Ex. 1005, 9:3-59; Ex. 2004, ¶ 73.

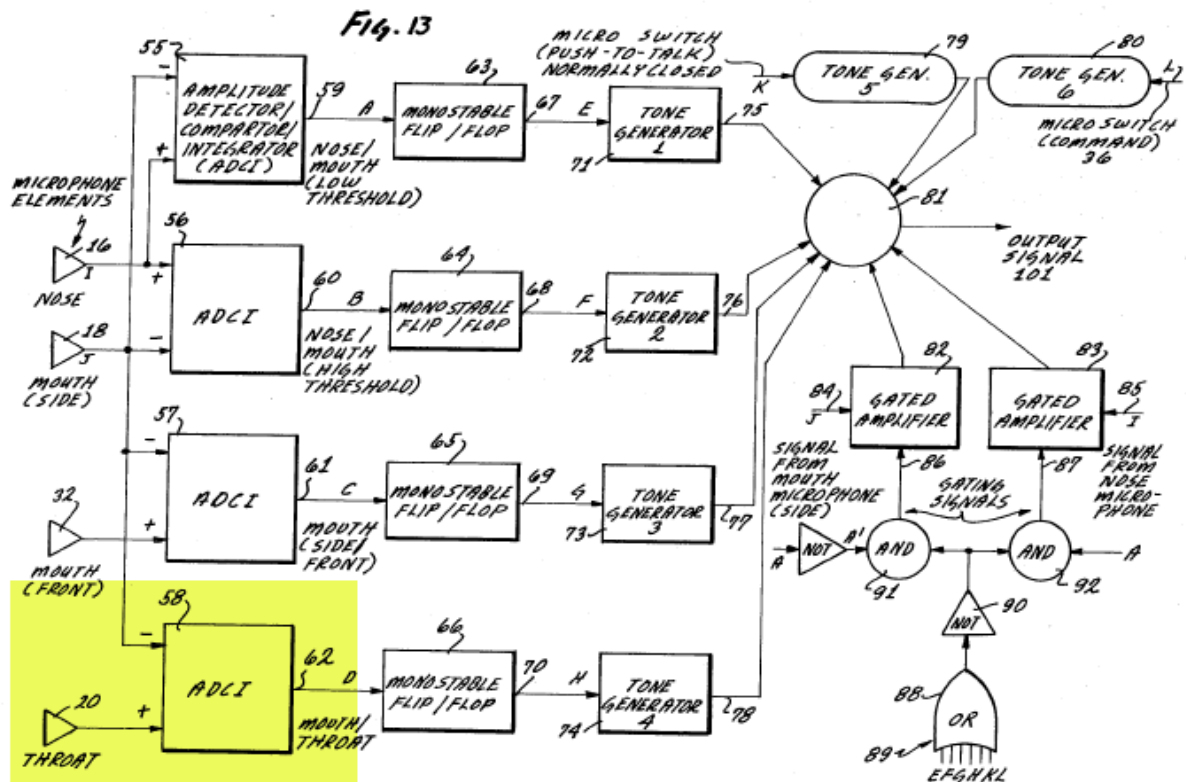
As discussed above, the proper construction of “cross-correlation data” is “data generated by multiplying and adding the result of a windowed set of samples at one time from one source by a windowed set of samples from another source where the two sets of samples may be shifted in time relative to each other.” Meisel’s data does not involve multiplying and adding the results of a windowed set of samples in any way. *Id.*, ¶¶ 72-74.

Even under the plain meaning of cross-correlation data, Meisel’s imbalances cannot satisfy the limitation. As explained above, the plain meaning of cross-correlation requires multiplication and time shifting. Meisel provides neither. Meisel’s imbalances and resulting signals are, at most, a comparison, not a correlation. *Id.*, ¶¶ 75-76.

Petitioner next argues that it would be obvious to modify Meisel to include Freidin’s cross-correlation. Pet. at 30-36. According to Petitioner, a POSITA would have been motivated to make the combination “identify voiced speech types because

such voiced speech types correlate when the signals at the throat microphone and signal microphone (*e.g.*, mouth microphone) are high and the differences of the ADCI calculation may therefore be small.” Pet. at 32; Ex. 1003, ¶ 62.

Petitioner is wrong. The purpose of the throat microphone in Meisel is to differentiate voiced sounds from unvoiced sounds. Ex. 1005, 2:31-48; Ex. 2004, ¶¶ 77-79. Meisel outputs a 1 from the ADCI 58 when a sound is voiced, in which case the throat microphone amplitude is greater than the side mouth microphone amplitude. Ex. 1005, 10:9-11; Ex. 2004, ¶ 79.



Ex. 1005, FIG. 13 (highlighting added).

A POSITA would not see a need to modify Meisel to add a cross-correlation step because Meisel already includes all the necessary functionality to determine the classification of a sound. Ex. 2004, ¶ 80. In particular, Meisel does not state or identify a need to determine a period where both a mouth microphone and throat microphone have high amplitudes. *Id.*

Neither Dr. Anderson nor Petitioner identify such a need. Instead, they simply state that voiced speech types correlate when signals at the throat and a signal microphone are high. Pet. at 32; Ex. 1003, ¶ 62. Petitioner and Dr. Anderson cite no evidence that such a relationship exists, and they do not explain how such correlation would “improve the accuracy of voiced and unvoiced speech identification already performed by Meisel.” *Id.*; Ex. 2004, ¶ 81. In particular, Petitioner and Dr. Anderson do not identify any way that a POSITA would perceive Meisel’s speech detection as lacking accuracy or even being susceptible to an improvement in accuracy. *Id.*

Finally, Dr. Anderson and Petitioner state that a “POSITA would also have been motivated to use Freidin’s cross-correlation to identify voiced speech in Meisel because Freidin discloses that ‘[c]orrelators have been widely used in signal processing to measure the degree of correspondence between a data sequence and a predetermined pattern or reference sequence,’ including ‘in a number of applications including speech recognition.’” Pet. at 33; Ex. 1003, ¶ 63 (quoting Ex. 1007, 1:7-

26). According to Dr. Anderson and Petitioner, “A POSITA would have recognized that Meisel’s signal processing techniques would be most easily implemented using digital processors, and Freidin provides the guidance needed to find signal correlations.” *Id.*

The statements above fail to demonstrate any motivation. Ex. 2004, ¶¶ 82-83. At most, Freidin’s statement demonstrates one use of correlators; it does not suggest any use for correlators in a Meisel system. *Id.* Moreover, neither Dr. Anderson nor Petitioner suggest any need or use in Meisel for measuring the degree of correspondence between any two of Meisel’s microphones. Indeed, Meisel’s system is predicated on looking for *differences* (*i.e.*, imbalances) between microphones, not correspondence or similarities. *Id.* Detecting similarities goes against the actual inventive concept in Meisel, which is to classify speech by looking at imbalances between microphone signals. *Id.*

With respect to limitation [1e], Petitioner and Dr. Anderson further argue that a “POSITA would have been motivated to combine Meisel’s speech processing system with Freidin’s correlator to compare the correlation value (the cross correlation data, *supra* § V.A.4.e) to a correlation threshold to identify the speech type as voiced speech in Meisel.” Pet. at 40; Ex. 1003, ¶ 72.

However, Meisel neither uses nor suggests a use for correlation values. Ex. 2004, ¶¶ 84-85. Meisel is intrinsically directed to examining differences between

microphone signals, and a correlation value measures the level of similarity over a time sequence as it is shifted in time. *Id.* In particular, no portion of Meisel uses, or could use, data suggesting that the throat microphone and mouth microphone both have high amplitudes. *Id.*, ¶¶ 84-85. Indeed, neither Petitioner nor Dr. Anderson identify a case where data would be lost because both signals have high amplitudes and “the difference (calculated by Meisel’s ADCI) may be small.” Pet. at 41; Ex.1003, ¶ 73. Instead, they simply assert that such correlation would “provide[] additional accuracy” without explaining why or how. However, for the reasons discussed above, a POSITA would not recognize any increase in accuracy or “objectiv[ity]” from inserting cross correlation into Meisel. Ex. 2004, ¶¶ 84-85.

A POSITA therefore would not be motivated, except by a hindsight-based reason to match the ’058 Patent’s claims, to modify Meisel to include Freidin’s correlation. *Id.*, ¶ 86.

Further, Petitioner and Dr. Anderson do not explain how a POSITA would integrate Freidin’s correlation with Meisel’s system. Meisel does not use correspondences or similarities between signals for any purpose, and only utilizes differences or imbalances. *Id.*, ¶ 87. A POSITA would, therefore, need to completely redesign Meisel’s system to use correspondences. *Id.* It would not be obvious to do so or to incorporate Freidin. *Id.*

3. Petitioner’s Combination Does Not Disclose or Render Obvious “identif[ies/ying] information of the acoustic signals as voiced speech when the cross correlation data corresponding to a portion of the acoustic signal received at the one receiver exceeds a correlation threshold,” as Required by Claims 1 and 2

As explained above, Petitioner’s combination does not disclose or render obvious “cross correlation data” as claimed. However, even if the combination did render the required “cross correlation data” obvious, it still does not render this limitation obvious because Meisel does not disclose this limitation alone, and a POSITA would not combine Meisel and Freidin. Ex. 2004, ¶ 89.

Meisel does not disclose this limitation alone because it does not identify information of the acoustic signals as voiced speech when the digital code corresponding to the pattern of imbalances exceeds a correlation threshold. Instead, Meisel makes a comparison between the throat microphone and the mouth microphone without checking to see if anything exceeds a threshold. *Id.*, ¶ 90.

Instead, Meisel explains that “[s]top consonants (such as b, d, g, p, t, and k) are pronounced by closing the vocal tract briefly. While the vocal tract is closed, little sound comes from the mouth.” Ex. 1005, 2:39-41. Thus, the mouth microphone will show a low amplitude. An unvoiced consonant, such as p, t, or k, will similarly have a low amplitude from the throat microphone. *Id.*, 2:42-43 (“For unvoiced stops (p, t, and k), the vocal cords do not vibrate and little sound comes

from the throat as well.”). However, the vocal cords do vibrate for a voiced stop consonant. *Id.*, 2:44-45 (“For voiced stops (b, d, and g), however, the vocal cords vibrate during the closure.”). Accordingly, for a voiced consonant, the throat microphone will have a relatively high amplitude, while the mouth microphone has a lower amplitude because the vocal tract is closed. *Id.*, 2:39-45. Ex. 2004, ¶¶ 91-92.

Accordingly, Meisel will not indicate voiced speech when the “correlation data exceeds a correlation threshold.” *Id.*, ¶ 93. Indeed, a POSITA would understand that, according to Meisel’s disclosure, a greater difference, and thus a lower comparison, between the throat and mouth microphone is indicative of voiced speech. *Id.*, ¶ 94. Meisel therefore does not disclose this correlation limitation. *Id.*

While Petitioner seeks to use Freidin with Meisel to overcome this deficiency, for the reasons discussed above, a POSITA would not combine Freidin with Meisel. *Id.*, ¶ 95.

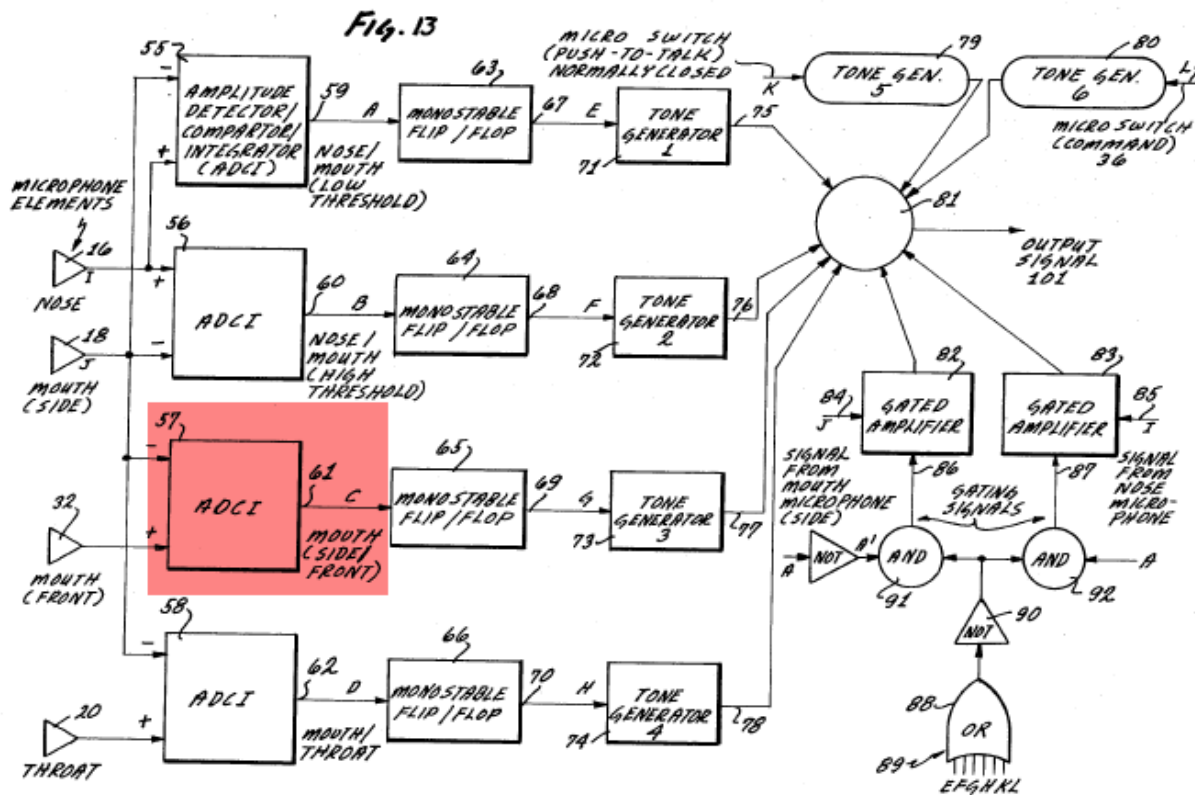
Accordingly, Petitioner’s combination fails to disclose or render obvious these limitations.

4. Petitioner’s Combination Does Not Disclose or Render Obvious “[identifies/identifying] information of the acoustic signals as noise when the difference parameters are less than the gain threshold,” as Required by Claims 1 and 2

Meisel does not discuss detecting noise. *See generally* Ex. 1005. Indeed, Meisel generally assumes that received signals will not include noise because of Meisel’s use of “close-talking microphones.” *Id.*, 2:54-60 (“Close-talking microphones have the secondary advantage that they respond very little to background noise, such as may occur in an office environment.”); Ex. 2004, ¶ 97; Ex. 2001, 48:5-9. Meisel also assumes that the placement of the microphones will prevent noise. Ex. 1005, 6:8-11 (“The microphone 18 detects speech sounds from the mouth; its placement is such that it is not overly affected by breath sounds.”); Ex. 2004, ¶ 97.

While Meisel cites to Sakoe, this citation does not suggest that Meisel needs a separate noise detection process. *Id.*, ¶ 98; *see* Pet. at 56, Ex. 1003, ¶ 89. Meisel cites Sakoe in the context of discussing the shortcomings of prior art systems. Ex. 2004, ¶ 98. In particular, Meisel faults Sakoe for using one of its two microphones for non-speech purposes. Ex. 1005, 1:46-50; Ex. 2004, ¶ 99. One of skill in the art would not recognize this discussion as suggesting the addition of noise cancellation; *Id.* However, to the extent that it does, it further supports that Meisel’s throat microphone is not a voicing sensor as claimed. *Id.*

Petitioner further states, “[w]hen the difference between the first mouth microphone input and the second mouth microphone input (A-B) is below the threshold T1, then the information of the acoustic signals from the side and front mouth microphones [] is identified as noise.” Pet. at 57. This is incorrect. Meisel makes no such identification. Meisel compares the two mouth microphones in ADCI 57, which generate a Boolean output 61. Ex. 1005, 9:60-10:11, FIG. 13. Ex. 2004, ¶¶ 100-101.



Ex. 1005, FIG. 1 (annotated).

Instead, Meisel only uses that Boolean variable and difference to determine if a plosive or fricative is being pronounced. *Id.*, 10:6-9 (“The output 61 will become logically true when a plosive or fricative is being pronounced.”); Ex. 2004, ¶ 102. The determination of whether a sound is a plosive or fricative is not the same as identifying a signal as noise. *Id.* Indeed, Meisel explains that some sounds are not plosives or fricatives, such as nasal sounds. Ex. 1005, 10:4-9; Ex. 2004, ¶ 102. These sounds are not noise but would be classified as such under Petitioner’s interpretation. *Id.*

Petitioner seeks to add Kaneda’s noise/speech detection to fill this hole in Meisel. Petitioner’s proposed system would calculate the difference between the power output of two microphones (the example given is the two mouth microphones) and, when that difference is greater than a threshold, speech is detected. Pet. at 57. Thus, under Petitioner’s interpretation, noise would be detected when the difference in power output of two microphones’ signals is below some threshold. *See id.*

However, even if the Kaneda process were added in parallel to the Meisel system, it would result in signals that Meisel considers speech to be classified as noise. Ex. 2004, ¶¶ 103-104. Meisel already compares the signals of the two mouth microphones and does not consider those signals noise. *Id.* Instead, if the mouth microphones’ amplitudes are similar, Meisel interprets that sound as not fricative or plosive. Ex. 1005, 10:7-9; Ex. 2004, ¶¶ 103-104. Similarly, if the Petition is

interpreted such that the inputs to the Kaneda system are the mouth microphone and nose microphone, Meisel uses those comparisons to detect partial closure of the mouth or nasal sounds. Ex. 1005, 10:4-7; Ex. 2004, ¶¶ 103-104.

Thus, under Petitioner’s proposed system, if there are similar amplitude signals from two microphones, that portion of the entire input would be classified as noise. Pet. at 55-64; Ex. 2004, ¶ 105. This would occur even if the Kaneda portion of the system were operating in parallel with, rather than in place of, Meisel’s recognition system. Ex. 2004, ¶ 105. Thus, Meisel’s output would lose critical pieces of data, frustrating its speech classification function. *Id.*

Accordingly, Petitioner has failed to show that this limitation is disclosed or rendered obvious.

5. Petitioner’s Combination Does Not Disclose or Render Obvious “wherein the voicing sensor includes at least one detector selected from a group including radio frequency devices, electroglottographs, ultrasound devices, acoustic throat microphones, and airflow detectors,” as Required by Claim 5

As explained above, a POSITA would not understand Meisel’s “throat microphone” to be a throat microphone as the term is understood in the art, and as it is recited in the ’058 Patent, Claim 5. *See supra* Section VI.A.1. Moreover, the “acoustic throat microphone” of Claim 5 must still meet the limitation that it is “unaffected by environmental noise.” Meisel’s throat microphone is affected by

environmental noise, and thus Petitioner's combination does not render Claim 5 obvious. Ex. 2004, ¶ 107.

B. GROUND 2: Meisel, in View of Kaneda, Freidin, and Lyons Does Not Render Obvious Claim 3

The Petition does not address the above deficiencies in its Ground 2 arguments, instead relying on its Ground 1 arguments as to those limitations. Pet. at 68-73. Accordingly, Petitioner has not shown obviousness under Ground 2. Ex. 2004, ¶ 108.

C. GROUND 3: Meisel, in View of Kaneda, Freidin, and Sakoe Does Not Render Obvious Claim 4

The Petition does not address the above deficiencies in its Ground 3 arguments, instead relying on its Ground 1 arguments as to those limitations. Pet. at 73-75. Accordingly, Petitioner has not shown obviousness under Ground 3. Ex. 2004, ¶ 109.

VII. CONCLUSION

As discussed above, Petitioner's combination fails to disclose or render obvious numerous limitations of the '058 Patent, and therefore, does not render obvious any claims. The Board should, therefore, issue a Final Written Decision holding all the Challenged Claims not unpatentable.

Respectfully submitted,

Dated: June 26, 2024

By: /Peter Lambrianakos /
Peter Lambrianakos (Reg. No. 58,279)

IPR2023-01352
PATENT NO. 7,246,058

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
Email: plambrianakos@fabricantllp.com

CERTIFICATE OF WORD COUNT

The undersigned hereby certifies that the portions of the above-captioned PATENT OWNER'S RESPONSE specified in 37 C.F.R. § 42.24 has 6,897 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.

Respectfully submitted,

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
Email: plambrianakos@fabricantllp.com

CERTIFICATE OF SERVICE

A copy of the foregoing Patent Owner's Response and Exhibits 2002 through 2009 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
Telephone: 212-257-5797
Fax: 212-257-5796