

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

---

**DECLARATION OF AKBAR M. SAYEED, PH.D.**

Case No. IPR2023-01352

---

**TABLE OF CONTENTS**

|                                                                         | <b><u>Page(s)</u></b> |
|-------------------------------------------------------------------------|-----------------------|
| I. INTRODUCTION .....                                                   | 1                     |
| A. Background and Qualifications .....                                  | 1                     |
| II. COMPENSATION .....                                                  | 6                     |
| III. MATERIALS CONSIDERED .....                                         | 6                     |
| IV. Legal Principles .....                                              | 7                     |
| A. Level of Skill in the Art.....                                       | 8                     |
| V. THE CLAIMED INVENTION OF THE '058 PATENT .....                       | 9                     |
| VI. THE ALLEGED PRIOR ART .....                                         | 10                    |
| A. Meisel (U.S. Patent No. 4,718,096) .....                             | 10                    |
| B. Kaneda (U.S. Patent No. 5,208,864).....                              | 14                    |
| C. Freidin (U.S. Patent No. 6,148,313).....                             | 14                    |
| D. Lyons .....                                                          | 14                    |
| E. Sakoe (U.S. Patent No. 4,239,936) .....                              | 14                    |
| VII. CLAIM CONSTRUCTION .....                                           | 15                    |
| A. “voicing sensor” .....                                               | 15                    |
| B. “physiological information” .....                                    | 19                    |
| C. “cross-correlation data” .....                                       | 21                    |
| VIII. PETITIONER’S COMBINATIONS DO NOT RENDER ANY<br>CLAIM OBVIOUS..... | 24                    |

|    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                |    |
|----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| A. | GROUND 1: Meisel in View of Kaneda and Freidin Does Not Render Obvious Claims 1-2 and 4-5 .....                                                                                                                                                                                                                                                                                                                                                                | 24 |
| 1. | Petitioner’s Combination 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.....                                                                                                                                              | 24 |
| 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 ..... | 31 |
| 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 orrelation threshold,” as Required by Claims 1 and 2.....                                                                                                                                        | 37 |
| 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 .....                                                                                                                                                                                                                      | 39 |
| 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 Recited in Claim 5.....                                                                                                                                                      | 42 |

|     |                                                                                                 |    |
|-----|-------------------------------------------------------------------------------------------------|----|
| B.  | GROUND 2: Meisel, in View of Kaneda, Freidin, and Lyons<br>Does Not Render Obvious Claim 3..... | 43 |
| C.  | GROUND 3: Meisel, in View of Kaneda, Freidin, and Sakoe<br>Does Not Render Obvious Claim 4..... | 43 |
| IX. | Conclusion .....                                                                                | 43 |

I, Akbar M. Sayeed, declare as follows:

1. I have been asked by counsel for Patent Owner Jawbone Innovations, LLC (“Jawbone” or “Patent Owner”) to review U.S. Patent No. 7,246,058 (the “’058 Patent”) entitled DETECTING VOICED AND UNVOICED SPEECH USING BOTH ACOUSTIC AND NONACOUSTIC SENSORS, and to provide my technical review, analysis, insights, and opinions regarding the ’058 Patent in view of the prior art cited by Petitioner Meta Platforms, Inc. (“Meta” or “Petitioner”). I submit this Declaration in support of Patent Owner’s Response in this IPR proceeding. I have personal knowledge of the matters stated herein and would be competent to testify to them if required.

2. I have been retained on behalf of Jawbone for the above-captioned *inter partes* review proceeding. I understand that the ’058 Patent is currently assigned to Jawbone.

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

## **I. INTRODUCTION**

### **A. Background and Qualifications**

4. I have reviewed and am familiar with the specification of the ’058 Patent. I understand the ’058 Patent has been provided as Exhibit 1001. I will cite to the specification using the following format: ’058 Patent as col.:line.

5. My CV is being submitted simultaneously herewith. (Ex. 2005)

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

## **II. COMPENSATION**

16. My compensation for time worked on this proceeding is not dependent on any issues related to the '058 Patent, the outcome of this proceeding, or the substance of my opinions. My compensation for time worked on this proceeding is at my customary rate of \$200 per hour. I have no financial interest in, or affiliation with, the Patent Owner or any of the real parties in interest.

## **III. MATERIALS CONSIDERED**

17. In providing my technical review, analysis, insights, and opinions, I have considered the '058 Patent and its prosecution history.

18. I have also considered the Petition filed by the Petitioner in this proceeding and the relevant exhibits relied on by Petitioner, including the expert declaration submitted by David V. Anderson.

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

#### **IV. LEGAL PRINCIPLES**

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

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

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

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

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

**A. Level of Skill in the Art**

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

26. In my opinion, a person of ordinary skill in the art would have a bachelor's degree in computer engineering, computer science, electrical engineering, mechanical engineering, or a similar field, and approximately 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.

## V. THE CLAIMED INVENTION OF THE '058 PATENT

27. 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.” '058 Patent, 1:34-37.

28. In contrast to previous methods which rely on acoustic data, the '058 Patent utilizes a voicing sensor that is unaffected by environmental noise to receive physiological information. *Id.*, 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.

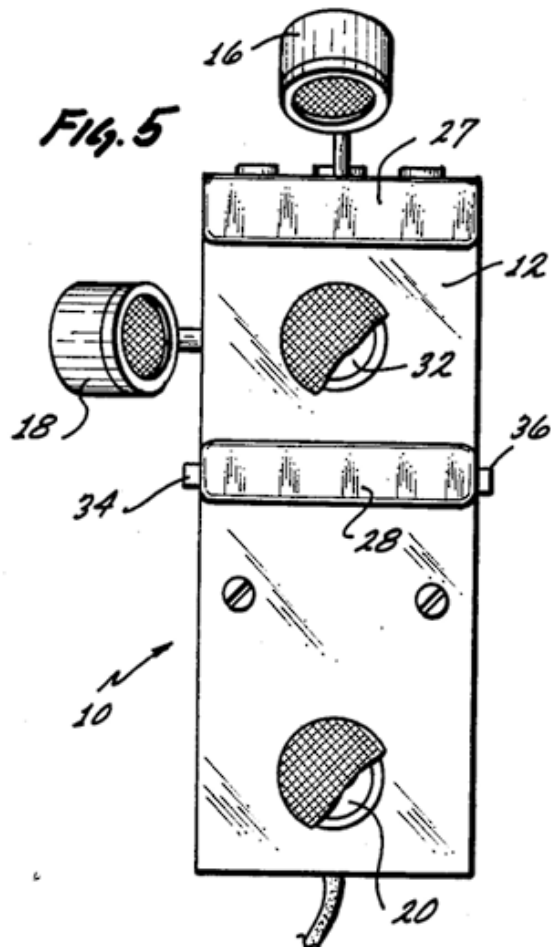
29. In particular, the '058 Patent receives acoustic signals at two microphones. *Id.*, 3:29-38. The system also receives physiological information from the voicing sensor. *Id.*, 5:26-31, 5:46-65. By correlating the acoustic signals with the physiological information, the system can determine when a 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.*

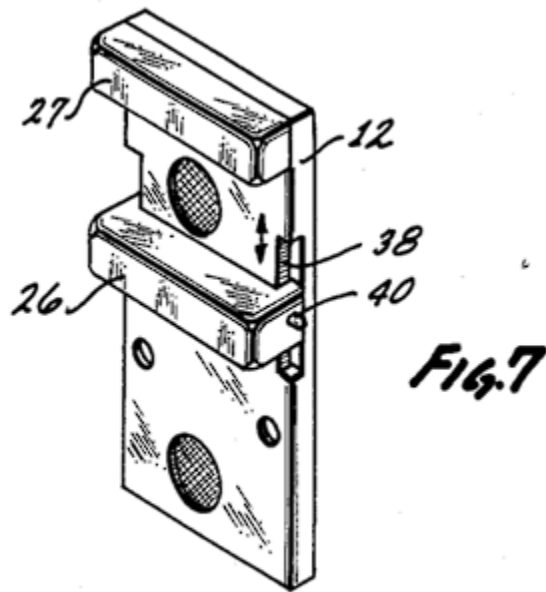
## **VI. THE ALLEGED PRIOR ART**

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

30. I have reviewed U.S. Patent No. 4,718,096 (Ex. 1005, “Meisel”). Meisel is directed to a speech recognition system. Meisel, 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). Meisel, 5:19-36.

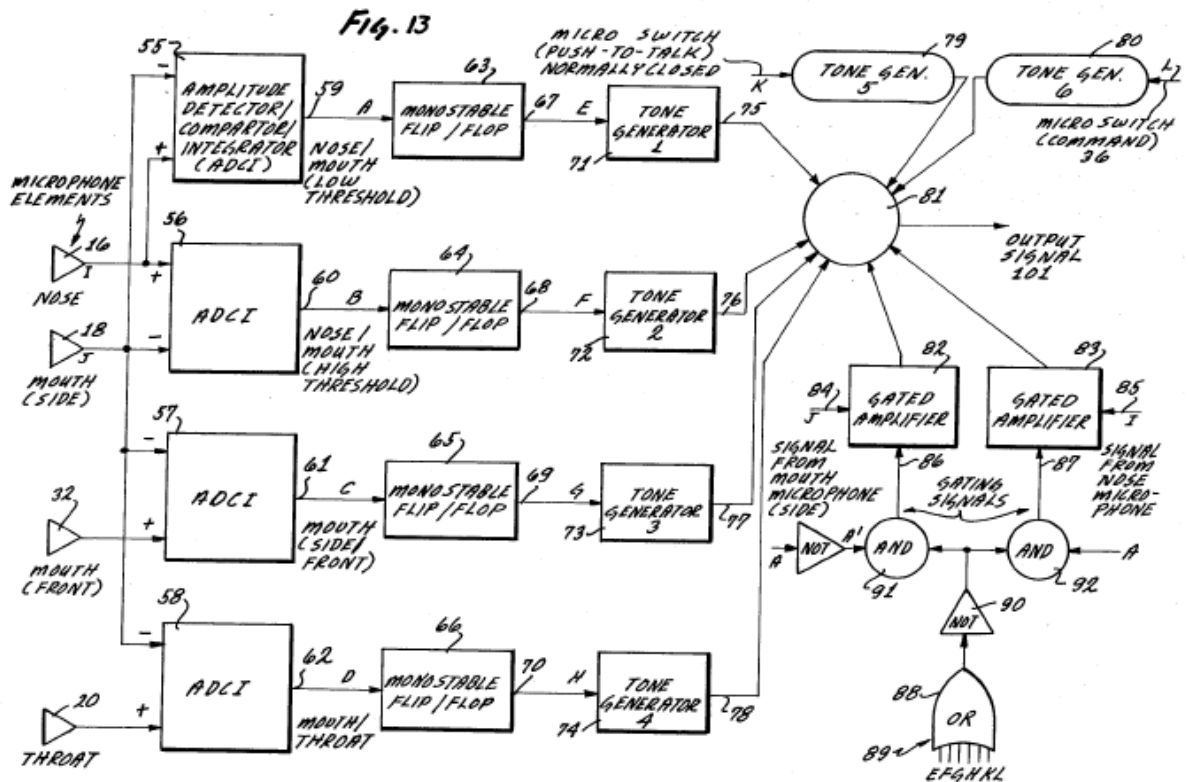


Meisel, Fig. 5. The microphones are enclosed in a housing (12) which includes spacers (26, 27) that prevent the microphones from touching the user's skin. *Id.*, 5:37-46. The spacers rest, for example, on a user's upper lip and chin. *Id.*



Meisel, Fig. 7.

31. Meisel seeks to detect and classify sounds as speech. Meisel, 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.



32. For example, 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 side 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 Figure 13 will be logically true. *Id.*, 9:60-10:11.

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

33. I have reviewed U.S. Patent No. 5,208,864 (Ex. 1006, “Kaneda”). Kaneda is directed to a method of detecting an acoustic signal. Kaneda, 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.

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

34. I have reviewed U.S. Patent No. 6,148,313 (Ex. 1007, “Freidin”). Freidin is directed to a “programmable digital correlator device, and associated correlation method, with a very efficient structure.” Freidin, 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.

**D. Lyons**

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

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

36. I have reviewed U.S. Patent No. 4,239,936 (Ex. 1008, “Sakoe”). Sakoe is directed to a speech recognition system for use in noisy environments. Ex. 1008 at 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.*

## VII. CLAIM CONSTRUCTION

### A. “voicing sensor”

37. A POSITA, reading the '058 Patent in combination with its prosecution history, would understand the term “voicing sensor” to mean “a sensor that is able to detect human tissue motion associated with speech and is unaffected by environmental noise.”

38. This construction comes directly from the specification, which explains that “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.” '058 Patent at 3:61-63.

39. A POSITA would understand that the voicing sensor must be unaffected by environmental noise. Every embodiment disclosed in the '058 Patent, including those using GEMS sensors, ultrasound devices, acoustic throat microphones, and airflow detectors, utilizes a non-acoustic sensor, which would be unaffected by environmental acoustic noise. *See* '058 Patent, Claim 5. For example, some of the '058 Patent embodiments utilize a General Electromagnetic Movement

Sensor (GEMS). *E.g., id.* at 3:57-63. “The GEMS is a radio frequency device (2.4 GHz) that allows the detection of moving human tissue dielectric interfaces.” *Id.* at 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.* at 4:1-5; *see also id.* at 3:64-4:13.

40. The ’058 Patent also explicitly defines the voicing sensor as a non-acoustic sensor. *Id.* at 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.* at 2:32-34 (discussing “voicing activity detectors or non-acoustic sensors”); 3:1-28 (discussing use of “non-acoustic voicing sensor”); 3:50-4:13; 5:63-65.

41. Moreover, being unaffected by environmental noise is key to the invention and required for good performance. *E.g., id.* at 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)). 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.

42. I understand that Claim 5 of the '058 Patent recites “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.” This recitation does not change my opinion. A POSITA would understand that a throat microphone is (and must be) unaffected by environmental noise, by touching the skin of a user to pick up vibration.

43. A throat microphone, also known as a laryngophone, is a well-known term of art for a type of skin surface microphone, originally developed for military aircraft pilots operating in extremely high noise environments, in which a vibration receiving diagram is in contact with a user's throat over the larynx, from which it receives vibrations directly. This contrasts with a conventional acoustic microphone which receives acoustic waves through air molecules. For example, definitions of “throat microphone” and “laryngophone” in the Merriam Webster Dictionary captures attached hereto as Exs. 2007, 2008 would be consistent with a POSITA's understanding of the ordinary meaning of a throat microphone as recited in the '058 Patent. As a further example, the attached Speech Intelligibility in Noise Using Throat and Acoustic Microphones (Ex. 2009), a 2004 publication by the U.S. Army Aeromedical Research Laboratory, 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."***

Ex. 2009 at 8.

44. A POSITA would further not understand the '058 Patent's reference to a throat microphone as "acoustic" 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. 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.

45. 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). *See* 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.*

46. Accordingly, for the reasons above, in my opinion, a POSITA would understand "voicing sensor" to mean "a sensor that is able to detect human tissue motion associated with speech and is unaffected by environmental noise."

**B. "physiological information"**

47. A POSITA, reading the '058 Patent in combination with its prosecution history, would understand the term "physiological information" to mean "human tissue motion information."

48. This construction 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). A POSITA would recognize that the human tissue motion described as being sensed by a voicing sensor throughout the specification is the claimed physiological information.

49. A POSITA would further understand that this construction is supported by the claim language itself. 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; *see also* Claim 2. A POSITA would read this to mean that “physiological information” is distinct from “acoustic signals.” A POSITA reading the claims in light of the

specification would thus understand physiological information to refer to human tissue motion, which the '058 Patent teaches is received independent of acoustic signals in the same manner. '058 Patent, 3:61-63, 3:64-4:13, 5:11-14, 5:26-27.

50. Accordingly, for the reasons above, in my opinion, a POSITA would understand “physiological information” to mean “human tissue motion information.”

**C. “cross-correlation data”**

51. A POSITA, reading the '058 Patent in combination with its prosecution history, would understand the term “cross-correlation data” to mean “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.” In this way the cross correlation function can measure the similarities of the two sequences if the sequences are time aligned or offset in time.

52. This construction is based on the plain meaning of “cross-correlation,” as well as the '058 Patent’s definition of “cross-correlation” as the XCORR function. *Id.* at 5:1-3.

53. Cross-correlation is defined as:

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

Ex. 2004 (citing Ex. 2005). The  $\tau$  refers to a time-lag and controls the relative time delay between samples. This definition is how a POSITA would have understood cross-correlation in 2001. Dr. Anderson uses a similar formula:

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

Ex. 1003, ¶ 60 (quoting Ex. 1016 at 20). These formulas track my construction above. The cross-correlation function  $R$  is the summation over a time window of the multiplication of samples from  $x$  at time  $n$  by samples from  $y$  at time  $n - \tau$ . The  $\tau$  value accounts for the shifting in time of my construction.

54. 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.

55. Petitioner’s exhibits use the same definition. 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

multiplies sample  $g$  at indices  $i$  and  $j$  by sample  $t$  at indices  $(i-m)$  and  $(j-n)$ .. Ex. 1023 at 279.

56. 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.”).

57. I note that Petitioner utilizes these definitions. Pet. at 34-35.

58. A POSITA would know the definition of cross-correlation as discussed above. Moreover, the definition in the '058 Patent of cross-correlation as 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.

59. Indeed, the '058 Patent explains that its use of cross-correlation (XCORR) requires the ability to shift the samples in time. '058 Patent at 5:11-14 (“The XCORR is only calculated to 15 delays, which corresponds to just under 2 milliseconds at 8000 Hz.”). 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. *Id.* at 5:31-39 (“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).”).

60. Accordingly, in my opinion, a POSITA would understand “cross-correlation data” to mean “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.”

## **VIII. PETITIONER’S COMBINATIONS DO NOT RENDER ANY CLAIM OBVIOUS**

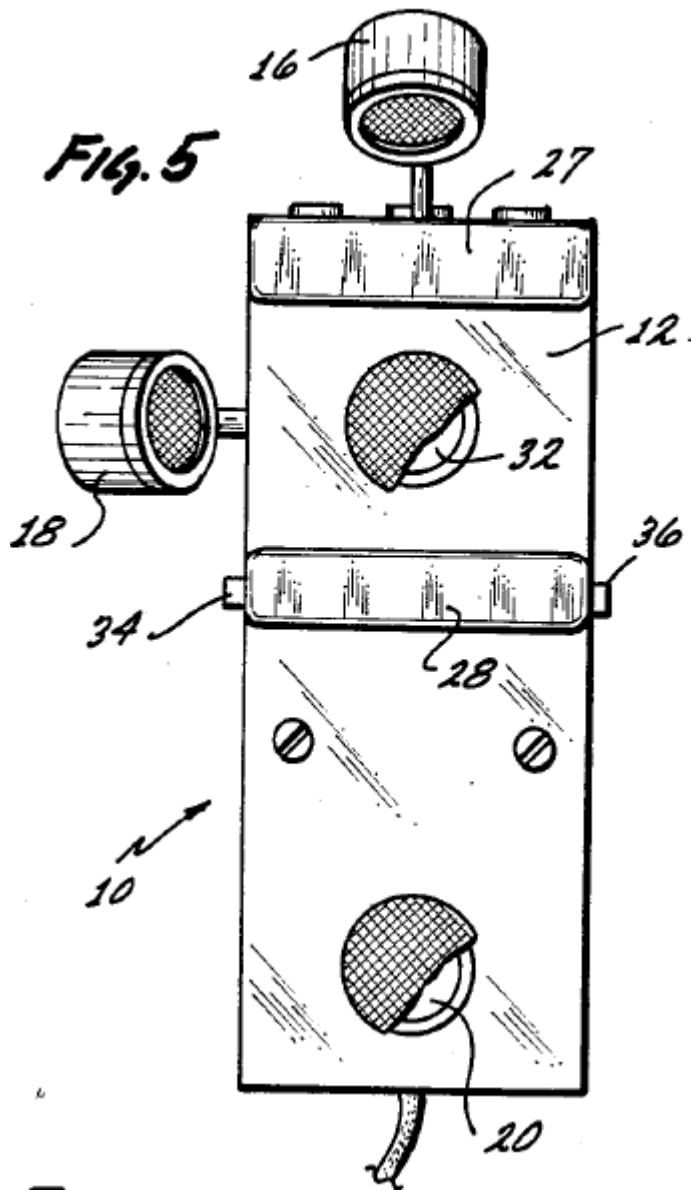
### **A. GROUND 1: Meisel in View of Kaneda and Freidin Does Not Render Obvious Claims 1-2 and 4-5**

#### **1. Petitioner’s Combination 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**

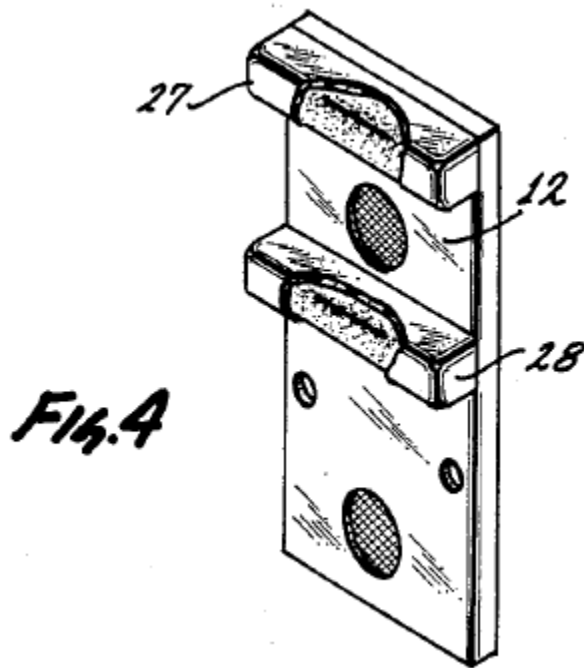
61. As I explain above, “voicing sensor” should be construed as “a sensor that is able to detect human tissue motion associated with speech and is unaffected by environmental noise” and “physiological information” should be construed as “human tissue motion information.”

62. Petitioner relies on Meisel’s throat microphone as the voicing sensor. Pet. at 21. In my opinion, Meisel’s “throat microphone” does not disclose or render obvious this limitation under the correct construction of “voicing sensor” or “physiological information” that

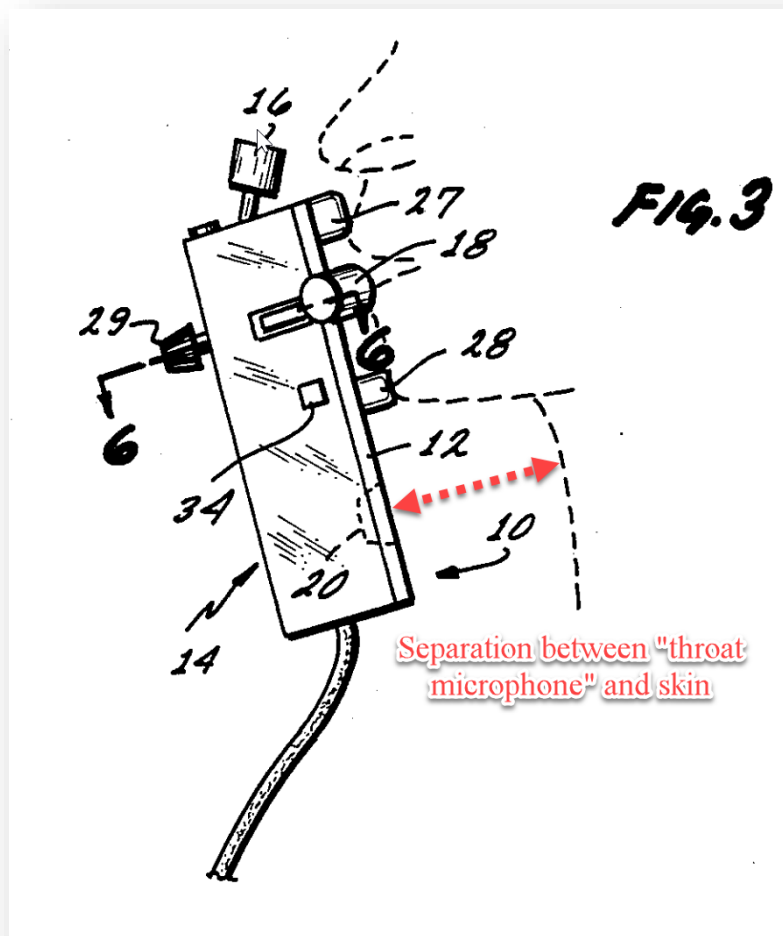
63. As an initial matter, a POSITA would not understand microphone that Meisel describes as a “throat microphone” to be a throat microphone as the term is used in the art, or in the ’058 Patent. Meisel’s “throat microphone” is simply an acoustic microphone pointed at a user’s throat that receives acoustic waves generated by “sounds emanating from the throat of a speaker.” and simply Meisel uses the term “throat microphone to distinguish between that lower microphone, and it’s upper “nose microphone” and “mouth microphone.” Meisel at 5:33-35; *id.* at 5:47-50 (“The distance between spacers 27 and 28 is individual to the speaker to assure that the microphones 16, 18, and 20 will be disposed respectively adjacent to the nose, mouth, and throat of the speaker.”) Meisel’s throat microphone is illustrated in Figure 5 as identifier 20:



64. The throat microphone is also shown (but not numbered) in Figures 4 and 7.



65. The spacers 27 and 28 shown in the Figures will rest against the user's upper lip and chin, respectively. Meisel at 5:37-46. There will, therefore, be a substantial separation between the user's skin and the throat microphone, as illustrated in Figure 3 (excerpted below). 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. Meisel, 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.



Meisel, Fig. 3 (annotations added).

66. A POSITA would understand that Meisel's throat microphone does not receive physiological information in the form of "human tissue motion," as required by both constructions. 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 information on motion of human tissue. See Meisel, 8:55-59 ("Similar proof can be provided of the importance of recording

the *sounds* emanating from the throat. . .”); *id.* at 8:37. Indeed, Meisel describes that “it is a further aspect of this invention to allow close talking microphones to be used,” and a POSITA would understand microphone 20 to be such a close-talking microphone (not a throat microphone). A POSITA would understand that the conventional close-talking microphones disclosed in Meisel detect acoustic sound waves through air, not human tissue vibration, especially when they are not even in contact with human tissue as disclosed in Meisel. 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. 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.

67. 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,” I disagree. As discussed above in Section VII.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.

68. A POSITA would understand that Meisel’s throat microphone is affected by environmental noise. 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. 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 ones 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. The throat microphone therefore is not “unaffected by environmental noise.”

69. I have reviewed Petitioner’s remaining arguments as to this limitation, and they do not remedy these deficiencies.

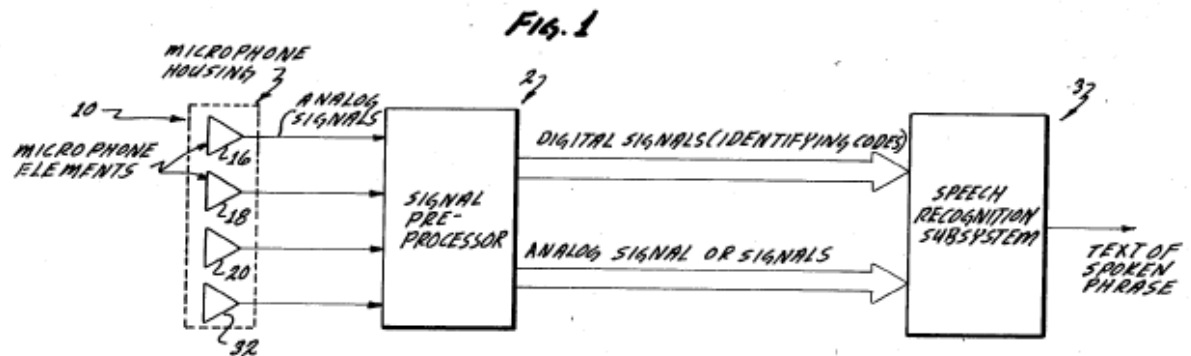
70. Therefore, in my opinion, Petitioner’s combination does not render this element of Claims 1 and 2 obvious. Accordingly, in my opinion, Claims 1 and 2 are not obvious under Ground 1.

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**

71. Petitioner's combination does not render this limitation obvious because: i) Meisel's imbalances are not cross-correlation data; and ii) a POSITA would not combine Freidin with Meisel.

**i. Meisel's "imbalances" Are Not "cross correlation data"**

72. Meisel explains that "The 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." Meisel at 4:35-39. Meisel's preprocessor passes those imbalances as Boolean (that is, true or false) values to the speech recognition system. *Id.* at 4:39-42.



73. As Meisel further explains, the preprocessor generates the imbalance data through the use of an ADCI (amplitude detector/comparator/integrator). *Id.* at 8:60-9:2. The ADCI performs a simple compare operation between two inputs by inverting the second input and subtracting it from the first. *Id.* at 9:3-59.

74. As I explain 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.”

75. Meisel’s imbalance data is not generated by multiplying any samples. Thus it is not cross correlation data and does not render this limitation obvious.

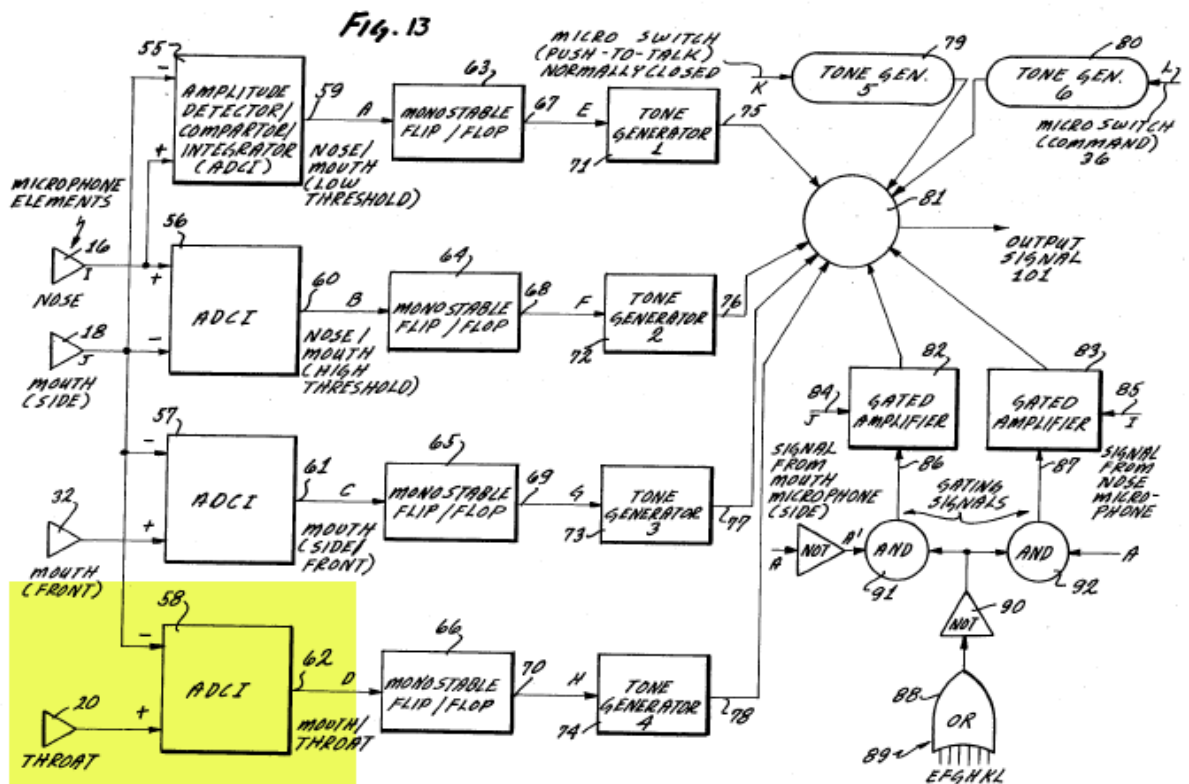
76. I further disagree with Dr. Anderson and Petitioner when they state that the imbalances are cross correlation data, even if the term is given its plain meaning. As I explain above, the plain meaning of cross correlation requires multiplication, not merely a comparison or difference operation. At most, the imbalances are a *correlation* but not a *cross correlation* as required.

**ii. A POSITA Would Not Modify Meisel to  
Include Freidin’s Cross-correlation**

77. Petitioner and Dr. Anderson assert that a POSITA would have modified Meisel to include Freidin’s cross-correlation. I disagree.

78. Petitioner and Dr. Anderson assert that 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.

79. I disagree. The purpose of the throat microphone in Meisel is to differentiate voiced sounds from unvoiced sounds. Meisel at 2:31-48. 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. Meisel at 10:9-11.



Meisel, Fig. 13 (highlighting added).

80. 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. 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.

81. 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. Even if this is true—and they cite no evidence that it is—they do not explain how such correlation would “improve the accuracy of voiced and unvoiced speech identification already performed by Meisel.” *Id.* 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.

82. 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,’ for example, ‘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.*

83. I disagree that the statement above shows any motivation. At most, Freidin’s statement demonstrates one use of correlators; it does not suggest any use for correlators in a Meisel system. 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. Detecting similarities goes against the actual inventive concept in Meisel.

84. 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* § IX.A.4.e) to a correlation threshold to identify the speech type as voiced speech in Meisel.” Pet. at 40; Ex. 1003, ¶ 72. I disagree.

85. As I explain above, Meisel neither uses nor suggests a use for correlation values. Meisel is intrinsically directed to examining differences between microphone signals, and a correlation value measures the level of similarity. In

particular, no portion of Meisel uses, or could use data suggesting that the throat microphone and mouth microphone both have high amplitudes. Indeed, neither Petitioner nor Dr. Anderson identify a case disclosed in Meisel where accuracy would be lost because both signals have high amplitudes and “the difference (calculated by Meisel’s ADCI) may be small.” Pet. at 40; Ex.1003, ¶ 73. Instead, they simply rely on hindsight to assert that such correlation would “provide[] additional accuracy” without explaining why or how. In my opinion, for the reasons discussed above, a POSITA would not recognize any increase in accuracy or “objectiv[ity]” from inserting cross correlation into Meisel.

86. In my opinion, therefore, a POSITA 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.

87. Further, Petitioner and Dr. Anderson do not explain how a POSITA would integrate Freidin’s correlation with Meisel’s system. As I explain above, Meisel does not use correspondences or similarities between signals for any purpose, and only utilizes differences or imbalances. A POSITA would therefore need to completely redesign Meisel’s system to use correspondences. Petitioner and Dr. Anderson do not explain how it would be obvious to do so, and in my opinion, it would not be obvious.

88. Accordingly, in my opinion, Petitioner has not shown that this limitation is rendered obvious.

**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**

89. As I discuss 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.

90. In my opinion, 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 looks for a *lack of correlation* between the throat microphone and the mouth microphone.

91. Meisel explains that "Stop 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." Meisel, 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.”).

92. 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. Meisel at 2:39-45.

93. Thus, in my opinion, Meisel will not indicate voiced speech when the “correlation data exceeds a correlation threshold.”

94. A POSITA would understand that, according to Meisel’s disclosure, a greater difference, and thus a lower correlation, between the throat and mouth microphone is indicative of voiced speech. Meisel therefore does not disclose this limitation.

95. While Petitioner seeks to use Freidin with Meisel, for the reasons discussed above, a POSITA would not combine Freidin with Meisel.

96. Accordingly, in my opinion, Petitioner’s combination does not render this limitation obvious.

**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**

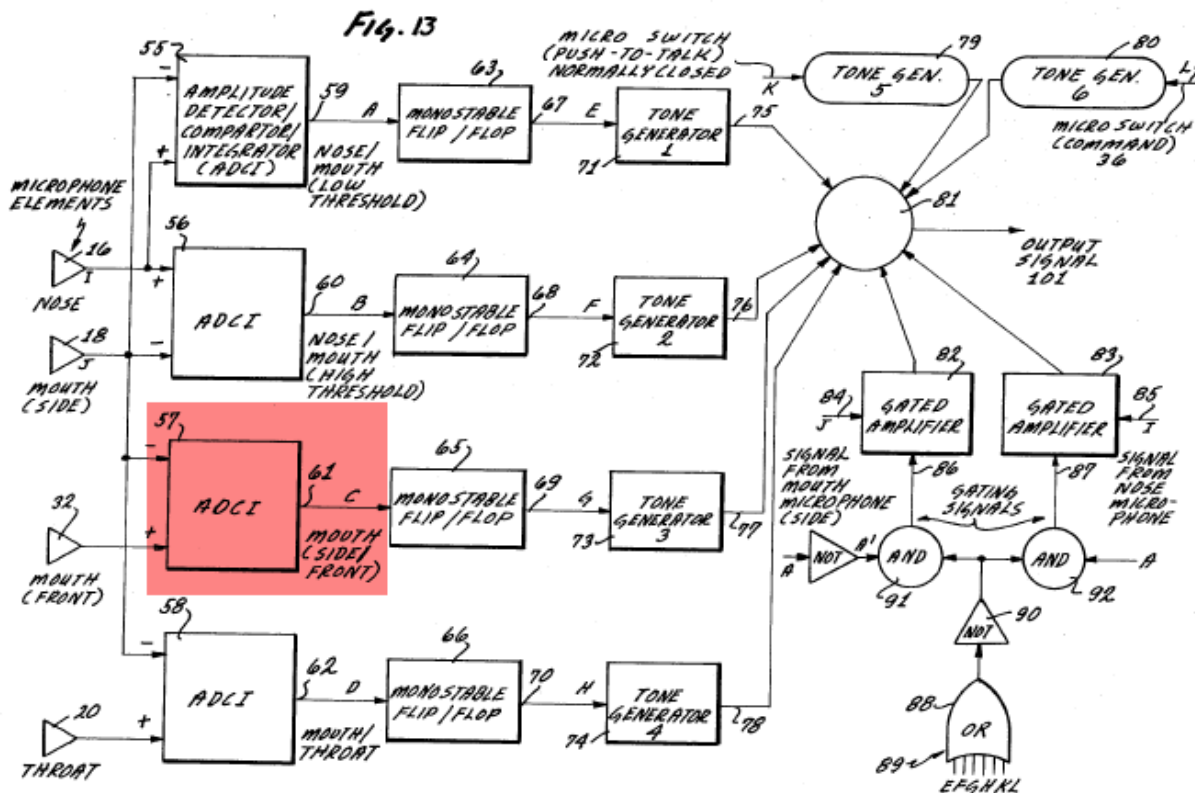
97. In my opinion, Meisel generally assumes that received signals will not include noise because of Meisel's use of “close-talking microphones.” Meisel, 2:54-60 (“Close-talking microphones have the secondary advantage that they respond very little to background noise, such as may occur in an office environment.”). Meisel also assumes that the placement of the microphones will prevent noise. *Id.*, 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.”).

98. Meisel cites to Sakoe, which “uses two microphones.” Meisel at 1:46-47. In Sakoe's system, “The second microphone is placed so as to avoid the speech signal and to detect background noise with a view toward minimizing its effect on recognition accuracy.” *Id.* at 1:47-50. Petitioner cites to this discussion as evidence that Meisel requires separate noise cancellation. Pet. at 56; Ex. 1003, ¶ 89.

99. I disagree. Meisel cites Sakoe in the context of discussing the shortcomings of prior art systems. In particular, Meisel faults Sakoe for using one of its two microphones for non-speech purposes. Meisel at 1:46-50. One of skill in the art would not recognize this discussion as suggesting the addition of noise cancellation.

100. According to Petitioner, “When 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.

101. I disagree. Meisel makes no such identification. Meisel compares the two mouth microphones in ADCI 57, which generates a Boolean output 61. Meisel, 9:60-10:11, Fig. 13.



Meisel, Fig. 1 (annotated).

102. Instead, Meisel only uses that difference to determine if a plosive or fricative is being pronounced. Meisel, 10:6-9 (“The output 61 will become logically true when a plosive or fricative is being pronounced.”). The determination of whether a sound is a plosive or fricative is not the same as identifying a signal as noise. Indeed, Meisel explains that some sounds are not plosives or fricatives, such as nasal sounds. Meisel, 10:4-9. These sounds are not noise but would arbitrarily be classified as such under Petitioner’s interpretation, in an apparent hindsight-based attempt to reach the claim language.

103. 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.*

104. 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. Meisel already compares the signals of the two mouth microphones and does not consider those signals noise. Instead, if the mouth microphones’ amplitudes are similar, Meisel interprets that sound as not fricative or plosive. Meisel, 10:7-9. 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. Meisel, 10:4-7.

105. 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. 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. Thus, Meisel's output would misclassify critical pieces of data as noise, frustrating its speech classification function.

106. Accordingly, in my opinion, Petitioner's combination does not render this limitation 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 Recited in Claim 5**

107. As I explain 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. 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.

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

108. Ground 2 is directed to dependent claims. Petitioner's Ground 2 arguments do not address the deficiencies within its Ground 1 arguments discussed above. Pet. at 68-73. Accordingly, in my opinion, Claim 3 is not obvious under Ground 2.

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

109. Ground 3 is directed to dependent claims. Petitioner's Ground 3 arguments do not address the deficiencies within its Ground 1 arguments discussed above. Pet at 73-75. Accordingly, Petitioner has not shown that Claim 4 is obvious unpatentable under Ground 3.

**IX. CONCLUSION**

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

I hereby declare that all statements made herein of my own knowledge are

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

Executed this 26th day of June, 2024.

Respectfully submitted,

A handwritten signature in black ink, appearing to read "Akbar M. Sayeed", with a horizontal line drawn underneath the signature.

Akbar M. Sayeed