

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

META PLATFORMS, INC.,

Petitioner,

v.

JAWBONE INNOVATIONS, LLC,

Patent Owner.

Case No. IPR2023-01352

U.S. Patent No. 7,246,058

**DECLARATION OF DR. DAVID V. ANDERSON IN SUPPORT OF
PETITIONER'S REPLY**

TABLE OF CONTENTS

I.	INTRODUCTION	3
II.	QUALIFICATIONS	3
III.	ADDITIONAL MATERIALS CONSIDERED	4
IV.	LEGAL STANDARDS	5
V.	CLAIM CONSTRUCTION	6
	A. Voicing Sensor	6
	B. Physiological Information	10
	C. Cross Correlation Data	13
VI.	GROUND 1	16
	A. Meisel Discloses a “Voicing Sensor”	16
	B. Meisel Discloses a “Physiological Information”	18
	C. Meisel in View of Freidin Discloses the Cross Correlation Data Recited in Claims 1 and 2	19
	D. Meisel in View of Kaneda Discloses Identifying Information of Acoustic Signals as Noise When Difference Parameters are Less Than the Gain Threshold	22
VII.	FOUNDATIONS 2-3	25
VIII.	CONCLUSION	26

I. INTRODUCTION

1. I have been retained by Meta Platforms, Inc. (“Meta”) as an independent expert consultant in this proceeding.

2. I am over 21 years of age and, if I am called upon to do so, I would be competent to testify as to the matters set forth herein.

3. My compensation is in no way contingent on the nature of my findings, the presentation of my findings in testimony, or the outcome of any proceeding.

4. I previously provided a declaration in this proceeding as Ex. 1003. My CV was previously submitted as Ex. 1004.

5. I have been asked to provide this supplemental declaration in response to the issues raised in Dr. Akbar M. Sayeed’s declaration (Ex. 2004).

6. Throughout this declaration, I refer to specific pages, figures, and/or line numbers of various exhibits. These citations are illustrative and are not intended to suggest that they are the only support for the propositions for which they are cited.

II. QUALIFICATIONS

7. My qualifications have been provided in my prior declaration. Ex. 1003, ¶¶ 8-11.

III. ADDITIONAL MATERIALS CONSIDERED

8. In forming my opinions, I have reviewed my prior declaration, the following additional documents, and any other document cited in this declaration or my previous declaration:

Exhibit	Description
Ex. 2002	WaveWalker 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/
Ex. 2003	Gardner, William A., Statistical Spectral Analysis: A Non-Probabilistic Theory, 1988
Ex. 2004	Declaration of Akbar M. Sayeed
Ex. 1038	U.S. Provisional Application No. 60/368,209 (“the ’209 provisional”)
Ex. 1039	Deposition Transcript of Akbar M. Sayeed, Ph.D.
Ex. 1040	Gregory C. Burnett, <i>The Physiological Basis of Glottal Electromagnetic Micropower Sensors (GEMS) and Their Use in Defining an Excitation Functions for the Human Vocal Tract</i> , Thesis, University of California at Davis (1999)
Ex. 1041	U.S. Patent No. 9,196,261 to Burnett et al.

9. I have also relied on my education, experience, research, training, and knowledge in the relevant art, and my understanding of any applicable legal principles described in this declaration.

10. All of the opinions contained in this declaration are based on the documents I reviewed and my knowledge and professional judgment. My opinions have also been guided by my understanding of how a person of ordinary skill in the art

would have understood the claims of the '058 patent at the time of the alleged invention. For purposes of this declaration, I have been asked to assume that the date of the alleged invention is the earliest claimed priority date: May 30, 2001.

11. I reserve the right to supplement and amend any of my opinions in this declaration based on documents, testimony, and other information that becomes available to me after the date of this declaration.

IV. LEGAL STANDARDS

12. My understanding of the relevant legal standards is set forth in my previous declaration. Ex. 1003 at 6-10.

13. I have been informed and understand that patent claims, and their terms and phrases, are interpreted by a court according to their ordinary and customary meaning to a person of ordinary skill in the art as of the earliest filing date of claimed invention. I understand that interpreting a term or phrase contained in a patent claim involves reading the language of the claim and considering other materials relevant to the claim, including the patent specification and the prosecution history, which I have been informed constitutes the “intrinsic record” of the patent. I understand that evidence extrinsic to the patent and its prosecution history may also be consulted, but that such materials should only be considered so long as they are not inconsistent with the claim language, the patent specification and the prosecution history.

V. CLAIM CONSTRUCTION

A. Voicing Sensor

14. I have considered Dr. Sayeed's opinions regarding Jawbone's proposed construction of "voicing sensor" as "a sensor that is able to detect human tissue motion associated with speech and is unaffected by environmental noise." Ex. 2004, ¶ 37. I disagree with his opinions and Jawbone's proposed construction.

15. Dr. Sayeed states that "[e]very embodiment disclosed in the '058 utilizes a non-acoustic sensor, which would be unaffected by environmental acoustic noise," because "[f]or example, some of the '058 Patent embodiments utilize a General Electromagnetic Movement Sensor (GEMS)." Ex. 2004, ¶ 39. As an initial matter, I disagree that the '058 patent only discloses a non-acoustic sensor because it discloses a PSAD (Pathfinder Speech Activity Detection) system that uses entirely acoustic sensors in the form of two acoustic microphones "to detect both voiced and unvoiced speech." Ex. 1001, 9:44-46. The PSAD system may "operate independently" of the NAVSAD, or as the sole voiced speech detection if the NAVSAD fails. Ex. 1001, 2:48-49. The "PSAD system supports an acoustic approach that functions in most low-to medium-noise environments[.]" Ex. 1001, 3:12-13. Figure 2 of the '058 patent shows the acoustic PSAD system being used to determine voicing activity, particularly "in low acoustic noise environments" where it is "highly sensitive." Ex. 1001, 2:45-49.

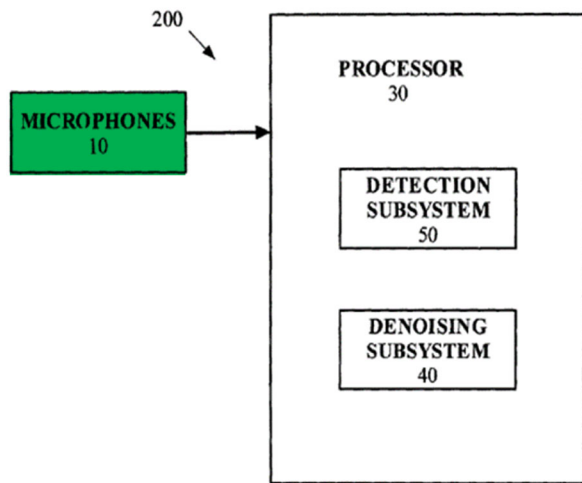


Figure 2

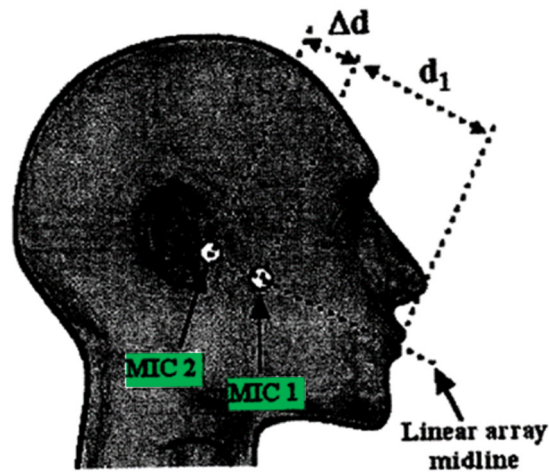


Figure 7

Ex. 1001, FIGS. 2, 7.

16. U.S. Provisional Application No. 60/368,209 (“the ’209 provisional”), which is incorporated by reference in the ’058 patent, confirms that “voicing sensor” includes acoustic voicing sensors, such as acoustic throat microphones. The ’209 provisional explains that “[t]he external VAD [(voicing activity detectors)] device may include ... acoustic microphones and associated software,” that “[t]wo-microphone acoustic VADs” can “operate without any external hardware, just using two microphones,” and that the “Pathfinder VAD” can “determine when voicing is occurring” “with any of the microphone configurations above with no modifications.” Ex. 1038, 8, 15, 16.

17. This is consistent with Dr. Sayeed’s testimony during deposition that “[s]ounds are changes in air pressure or air molecules; vibrating, if you will, and -- and typical microphone, it’s diaphragm, senses those changes in pressure.” Ex. 1039,

38:11-14; *see also* 38:15-17 (confirming an acoustic signal can travel through a solid). A POSITA would understand that acoustic throat microphone does not need to touch the skin of a user to pick up vibrations. For example, as Dr. Sayeed acknowledged during deposition, “you can detect the vibrations of the skin, if you will, or the throat, in this case, through a microphone pointed at it.” Ex. 1039, 41:18-20. Thus, it is my opinion, and POSITA would have understood when reading the specification, that the ’058 patent discloses an embodiment that uses only acoustic microphones as voicing sensors.

18. The GEMS sensor relied on by Dr. Sayeed also does not meet the criteria that Dr. Sayeed describes as a voicing sensor. For example, contrary to Dr. Sayeed’s opinion, GEMS may be affected by environmental noise in the form of non-acoustic environmental noise. In reviewing Dr. Sayeed’s deposition transcript, I note that he acknowledged that noise can be non-acoustic. Ex. 1039, 24:11-17. As I state above, GEMS is affected by non-acoustic noise, such as noise caused by movement near the vocal cord during speech. The ’058 patent acknowledges this, explaining that “[a]nything that moves near the [GEMS] sensor will cause a change in phase of the reflected wave that will be amplified and displayed as a change in voltage output from the sensor” and appear as noise signals. Ex. 1001, 4:6-9. In reviewing the inventor’s thesis, I note that he further confirms that GEMS is affected by substantial environmental noise, including “the jaw and tongue moving, the blood coursing through arteries and veins, the vertical motion of glottis as speech begins and ends, and

any motion of the GEMS itself with respect to the neck and trachea.” Ex. 1040, 14.

This confirms that the GEMS sensor on which Dr. Sayeed relies to allege that “voicing sensor” must be a non-acoustic sensor fails to support his position. Dr. Sayeed further confirmed during deposition that the GEMS sensor in the ’058 patent does not directly touch the skin. *See* Ex. 1039, 43:16-44:8. Rather, the GEMS sensor is merely one type of voicing sensor, and as I explain above, the ’058 patent discloses acoustic microphone voicing sensor in the PSAD embodiment.

19. Dr. Sayeed asserts that his construction of “voicing sensor” comes directly from the specification. Ex. 2004, ¶ 38. I disagree that the specification provides a definition of “voicing sensor” that is different from the ordinary meaning of “voicing sensor” as understood by a POSITA. The passage that uses the language of Dr. Sayeed’s construction does not reference “voicing sensor,” refers to “[a]lternative embodiments,” and states that such embodiments “may use” a “sensor that is able to detect human tissue motion associated with speech and is unaffected by environmental acoustic noise.” A POSITA reading this passage would not understand this passage to be defining a “voicing sensor.”

20. Dr. Sayeed also opines that “[e]very embodiment disclosed in the ’058 Patent, including those using ... acoustic throat microphones ... utilizes a non-acoustic

sensor.” Ex. 2004, ¶ 39. But this is unsupported by any passage in the ’058 patent, and ignores “acoustic” from “acoustic throat microphones.”

21. Dr. Sayeed contrasts a “throat microphone” to a conventional “acoustic microphone,” to support his opinion that a “voicing sensor” is unaffected by environmental noise. Ex. 2004, ¶ 43. But again, Dr. Sayeed ignores that the ’058 patent recites an “*acoustic* throat microphone,” not just a “throat microphone.”

B. Physiological Information

22. I have considered Dr. Sayeed’s opinions regarding Jawbone’s proposed construction for the term “physiological information” as “human tissue information.” Ex. 2004, ¶ 47. I disagree with this construction because it is more limited than a POSITA’s understanding of this term, including Jawbone’s own expert.

23. At deposition, Dr. Sayeed testified that “physiology is much bigger than human tissue motion.” Ex. 1039, 48:10-18. I agree. Physiological information is not limited to human tissue motion information. Instead, physiological information can also include information from radio frequency vibration, airflow, or acoustic signals.

24. I note that another patent by the same inventor confirms that a POSITA would understand that “physiological information” is not limited to “human tissue information.” In U.S. Patent No. 9,196,261 (“the ’261 patent”), the inventor explains:

The VAD 204 uses physiological information to determine when a speaker is speaking. In various embodiments, the VAD can include at least one of an accelerometer, a skin surface microphone in physical contact with skin of a user,

a human tissue vibration detector, a radio frequency (RF) vibration and/or motion detector/device, an electroglottograph, an ultrasound device, an acoustic microphone that is being used to detect acoustic frequency signals that correspond to the user's speech directly from the skin of the user (anywhere on the body), an airflow detector, and a laser vibration detector.

Ex. 1041, 3:40-50. The description of “physiological information” in the ’261 patent is consistent with the understanding of a POSITA.

25. Dr. Sayeed opines that the term “physiological information” does not have its ordinary meaning to a POSITA in the context of the ’058 patent. I disagree. There are no statements in the ’058 patent that would narrow the meaning of the term.

26. The term “physiological information” only appears once in the specification. Specifically, the term appears in the Abstract in a discussion of certain embodiments of the system: “Further, embodiments of the system include non-acoustic sensors that receive physiological information to aid in identifying voiced speech.” Ex. 1001, Abstract. This description only discusses certain embodiments of the patent among many embodiments in the patent. Further, the Abstract makes no mention of human tissue motion. Thus, a POSITA reading this passage would not understand that “physiological information” has been limited to “human tissue motion.”

27. Dr. Sayeed also cites to the same passage in the ’058 patent that he opines provides a construction of “voicing sensor.” Ex. 2004, ¶ 48 (citing 3:61-63). But this passage does not reference a “voicing sensor” nor “physiological

information,” and only states that “[a]lternative embodiments, however, *may use* any sensor that is able to detect human tissue motion.” The passage does not require human tissue motion.

28. Dr. Sayeed further opines that “the human tissue motion described as being sensed by a voice sensor through the specification is the claimed physiological information” and that the claim language in claim 1 and 2 imply that “physiological information” is distinct from “acoustic signals.” But Dr. Sayeed does not reconcile this opinion with the language in claim 5 that recites that the “voicing sensor” includes an “acoustic throat microphone.”

29. The language of claim 5 confirms that “physiological information” can include acoustic information corresponding to speech. Further, claims 1 and 2 do not exclude “physiological information” from being an “acoustic signal.” Dr. Sayeed ignores the full language of the claim 1 which says “generating cross correlation date between the physiological information and an acoustical signal *received at one of the two microphones.*” Similarly, the full language of claim 2 says “generating cross correlation date between the physiological information and an acoustical signal *received at one of the two receivers.*” A POSITA would have understood that the voicing sensor could receive acoustic signals, non-acoustic signals, or both, including signals from an acoustic throat microphone. There is no restriction in the claim language.

30. At deposition, Dr. Sayeed also acknowledged that “at a very basic level,

you can argue that [an acoustic signal produced by a person speaking contains physiological information].” Ex. 1039, 50:24-51:16.

C. Cross Correlation Data

31. I have considered Dr. Sayeed’s opinions regarding Jawbone’s proposed construction for the term “cross correlation data” as “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.” Ex. 2004, ¶ 51. I disagree with this overly narrow construction because it is unsupported by the intrinsic record and is more limited than a POSITA’s understanding of this term.

32. In my opinion, “cross correlation data” is a well-known term in the art. Meisel’s difference function performs cross-correlation by determining differences and similarities between two signals.

33. However, if the Board finds that Meisel does not disclose the claimed “cross correlation data,” Freidin discloses it as explained in my previous declaration, and a POSITA would have found it obvious in view of the combination of Freidin and Meisel for the reasons explained in my previous declaration. Ex. 1003, ¶¶ 67-75.

34. Dr. Sayeed contends that a “POSITA would know the definition of cross-correlation as” the mathematical formula provided in ¶ 53 of his declaration. I disagree. The ’058 patent’s specification does not define “cross correlation.” Instead, the ’058 patent states that “cross-correlation (XCORR)” is calculated “between the

acoustic signal from microphone 1 and the sensor data “[t]o further differentiate sensor noise from tissue motion.” Ex. 1001, 5:2-3, 5:11-12. In my opinion, Dr. Sayeed’s narrow construction of “cross correlation” does not appear to be supported by the ’058 patent’s specification, and is also inconsistent with how a POSITA would have understood the term, as I explained in my initial declaration. Ex. 1003, ¶¶ 53-75.

35. It is also my opinion that the extrinsic evidence Dr. Sayeed relies on does not support Jawbone’s narrow construction of “cross correlation data.” Dr. Sayeed states that Jawbone’s construction is correct because similar examples are found in “multiple exhibits” of Petitioner. Ex. 2004, ¶¶ 55-57. I disagree. Although these exhibits include descriptions of one example of cross correlation, that does not mean that cross correlation is limited to that one example. In my review of the intrinsic record (patent and prosecution history), I did not find any description that limits the meaning of cross correlation data to this narrow definition. As I explained in my previous declaration, a POSITA would not have understood the term cross correlation data to be so limited. Ex. 1003, ¶¶ 53-75.

36. Further, Dr. Sayeed acknowledged during deposition that the equations in his declaration are not windowed and contradict his construction. Ex. 1039, 65:16-23. Dr. Sayeed also acknowledged that cross correlation can be computed in the analog domain and the signals are not required to be digitized as required by his construction.

Ex. 1039, 63:8-11. Thus, Jawbone's construction is inconsistent with the intrinsic record and the POSITA's understanding of the term.

37. Dr. Sayeed contends that “the '058 Patent explains that its use of cross-correlation (XCORR) requires the ability to shift the samples in time,” and thus the claimed cross correlation must be limited that way. Ex. 2004, ¶ 59. I disagree because the intrinsic record does not indicate that the patentee intended to narrow the meaning of cross correlation data this way. Dr. Sayeed's citation to '058 patent, column 5, lines 11-14, is merely an exemplary application, not a requirement for time shifting. But, even if the Board concludes that “shift[ing] in time” is required by this claim term, in my opinion Meisel discloses it. Meisel discloses that input speech signals are adjusted “to follow the envelope of the speech signal” and the “output of each of the amplitude detectors 501 and 502 thus represents the amplitude of the input signal on an instant-by-instant basis,” which a POSITA would have understood refers to shifting signals to evaluate the same speech segment (window). Ex. 1005, 9:6-11, claim 5 (delaying incoming signal for comparison). A POSITA would have recognized the need to adjust Meisel's signals to account for delays in its multi-microphone speech recognition system because different microphones positioned at different locations would receive speech signals at different times and should be shifted in time so that the signals can be properly compared. A POSITA would have understood that Meisel extracts the envelope that represents a windowed signal over a period of time to correlate different

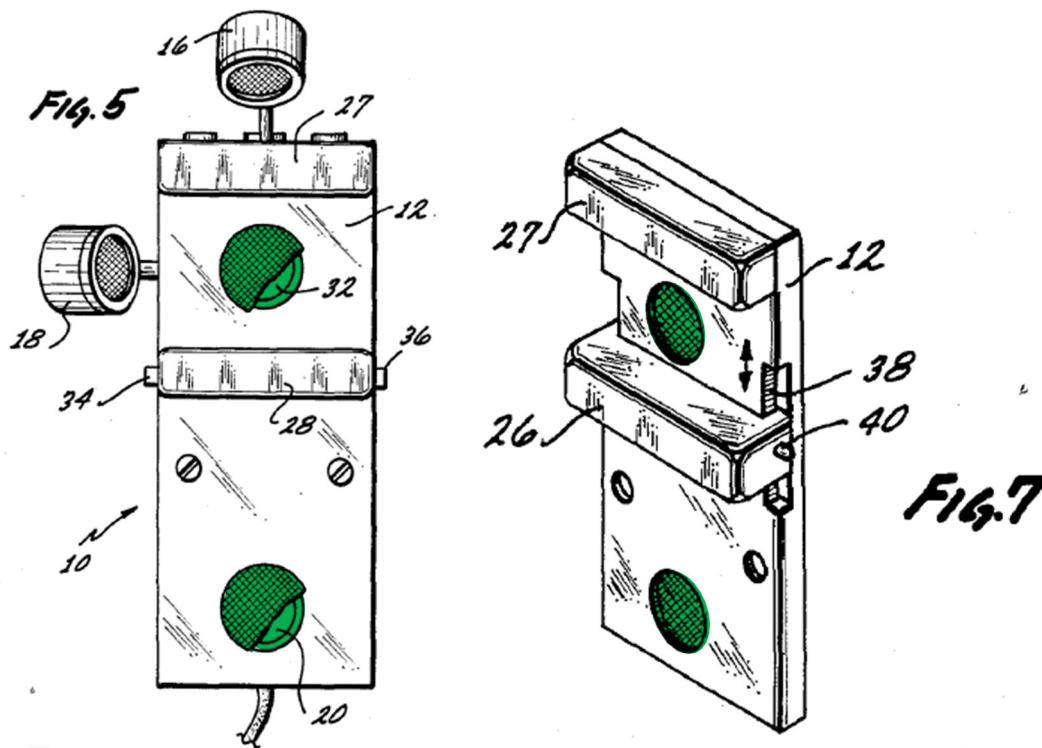
signals that may be misaligned in time by several milliseconds, similar to the '058 patent.

VI. GROUND 1

A. Meisel Discloses a “Voicing Sensor”

38. Dr. Sayeed argues that a “POSITA would understand that the throat microphone in Meisel is affected by environmental noise.” Ex. 2004, ¶ 68. I disagree because, to the extent Meisel’s microphone receives any noise, Meisel’s setup mitigates that noise to prevent it from hindering the operation.

39. Meisel’s microphones are embedded in the housing, which has a ring around where the microphone would be. Meisel’s ring around the throat microphone is described as “a solid sound absorbing substance” that prevents sound “from entering the microphones” and “attenuates sound vibrations in the housing.” Ex. 1005, 8:9-21. Meisel’s Figure 7 shows where the acoustic throat microphone is also recessed to help protect the microphone from sounds emanating from different directions, which mitigates the effects of any noise signal the microphone may receive and improves the signal-to-noise ratio when the user speaks and the throat microphone detects the vibrations of the user’s vocal cords. These disclosures confirm that Meisel’s microphones are insulated from environmental noise.



Ex. 1005, FIGS. 7 (annotated).

40. Dr. Sayeed also contends that Meisel's throat microphone must touch the skin to meet the claimed voicing sensor. Ex. 2004, ¶ 42. I disagree. I do not see any language in claim 1 or claim 5 of the '058 patent that would limit the construction or placement of the acoustic throat microphone. Moreover, the specification does not describe, let alone require, skin contact for the acoustic microphones. To the contrary, the '209 provisional explains that acoustic microphones perform best "when the [unidirectional microphone] is far enough away from the user's mouth (approximately 10-15 centimeters depending on the microphone)" (Ex. 1038 10-11), which indicates to a POSITA that an acoustic throat microphone, which is a unidirectional microphone, should not contact the speaker's skin. Meisel's throat microphone detects vibrations of

the skin created by the movement of the vocal cords when those vibrations cause vibrations in the air near the microphone, which is confirmed by Meisel’s teaching that the “throat microphone ... *detects the vibration of the vocal cords*, or ‘voicing,’” without contacting the skin. Ex. 1005, 2:33-35 (emphasis added). Meisel thus, in my opinion, discloses the claimed “voicing sensor.”

41. Dr. Sayeed argues “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” and states that Meisel only refers to “an acoustic microphone pointed at a throat.” But Dr. Sayeed does not explain the difference between an “*acoustic* throat microphone” in contrast to a “throat microphone.” As I discussed above, Dr. Sayeed ignores the “acoustic” in the “acoustic throat microphone,” and argues that an “acoustic throat microphone” uses a “non-acoustic sensor” (Ex. 2004, ¶ 39) without explanation of what “acoustic” then references in an “acoustic throat microphone.”

B. Meisel Discloses a “Physiological Information”

42. Dr. Sayeed argues that “Meisel’s throat microphone does not receive physiological information in the form of ‘human tissue motion,’ as required.” Ex. 2004, ¶ 66. I disagree because Meisel’s microphone detects vibrations from a speaker’s throat.

43. Even in Dr. Sayeed’s description of Meisel, Dr. Sayeed concedes that “Meisel’s throat microphone is described as a conventional microphone pointed at a

speaker's throat ... and is described as receiving *sounds* (*i.e.*, acoustic signals). This is the acoustic throat microphone recited in claim 5.

44. A POSITA would understand that to detect human tissue motion, it is not necessary for the sensor to touch the human tissue that is moving. Nor is it required by Dr. Sayeed's construction. During deposition, Dr. Sayeed acknowledged that to detect the vibration of the vocal cord, it is not necessary for a sensor to touch the vocal cord itself. Ex. 1039, 40:1-12. Instead, the sensor, such as a microphone, can detect the vibrations cause by the moving human tissue traveling through solids (such as the throat) or the air. Ex. 1039, 38:8-25.

45. Thus, to the extent that the Board construes "physiological information" as "human tissue motion information," Meisel discloses this limitation.

C. Meisel in View of Freidin Discloses the Cross Correlation Data Recited in Claims 1 and 2

46. Dr. Sayeed argues that Meisel's imbalances "are a comparison but not a cross correlation as required" (Ex. 2004, ¶ 76), but he does not explain how a "comparison" is different from a "correlation." As explained above, in my opinion, a POSITA would have understood that Meisel's signal comparison evaluating their similarities and differences is a type of cross correlation data.

47. Furthermore, a POSITA would have recognized that Freidin's cross correlation is compatible with Meisel's system. It was well known that "XCORR is more susceptible to errors from acoustic noise, and in high (<0 dB SNR) environments

is almost useless,” and Meisel’s system functions in environments with little “background noise, such as may occur in an office environment” where functions like XCORR perform best. Ex. 1001, 5:21-25; Ex. 1005, 2:58-60. Applying cross correlation data, as construed by Jawbone and disclosed in Freidin, to Meisel’s system would have involved nothing more than using a known technique (cross correlation) to improve devices (Meisel’s system applying a simply comparison of two signals) in the same way (signal processing for speech detection/recognition). It would have involved only combining prior art elements (cross correlation and speech detection/recognition) according to known methods (signal processing for speech detection/recognition) to yield predictable results (more accurate speech detection/recognition).

48. In my opinion, Dr. Sayeed’s contrary arguments are misplaced. Dr. Sayeed argues that “[d]etecting similarities goes against the actual inventive concept in Meisel.” Ex. 2004, ¶ 83. Dr. Sayeed incorrectly assumes that cross correlation only detects similarities and Meisel only detects differences between two signals. That is incorrect. Although a high value of cross correlation using Dr. Sayeed’s assumed equation indicates similarities, a low value indicates differences. When Meisel compares two signals this includes determining similarities and differences between the signals because a low difference indicates that the signals are similar and a high

difference indicates that they are different. Both thus achieve the same function to compare two signals.

49. Dr. Sayeed does not appear to dispute that Freidin discloses the claimed cross correlation data. Instead, Dr. Sayeed argues that a POSITA would not have been motivated to combine Meisel’s speech recognition system with Freidin’s cross correlation because “Meisel outputs a 1 from ADCI 58 when a sound is voiced, in which case the throat microphone amplitude is greater than the side mouth microphone amplitude.” Ex. 2004, ¶ 79. I disagree. Meisel’s system recognizes speech as a voiced sound even where the throat microphone and the side mouth microphone both have high signal amplitudes, so long as the throat microphone signal is sufficiently higher than the side mouth microphone signal to exceed the threshold value.

50. Dr. Sayeed also argues that “no portion of Meisel uses, or could use, data suggesting that the throat microphone and mouth microphone both have high amplitudes” because Meisel would have a low cross correlation when it recognizes a “voiced consonant” using ADCI 58. Ex. 2004, ¶ 85. Dr. Sayeed’s analysis is incomplete and incorrect because he looks at one edge case for detecting only *one* type of voiced speech—voiced stop consonants. In my opinion, a POSITA would have recognized that Meisel’s detection of vowels, which are also voiced speech, produces high cross correlation values under Dr. Sayeed’s interpretation of cross correlation data. Ex. 1005, 2:35-36, 12:25-33. Meisel discloses that vowels are “relatively easy to distinguish” due to the large signal amplitudes at both throat and side mouth

microphones. Ex. 1005, 15:49-51; Ex. 1039, 12:21-25. A POSITA would have found it obvious to implement Freidin's cross correlation data in Meisel's system to detect vowels, where a high cross-correlation value indicates that a vowel is being spoken, for the reasons I explain in my previous declaration. Ex. 1003, ¶¶ 67-75.

51. I also disagree with Dr. Sayeed's argument because the '058 patent itself does not produce a high cross correlation value between a mouth microphone signal and a GEMS signal for voiced stop consonants. The mouth microphone of the '058 patent would have a low signal amplitude and GEMS will have a high signal amplitude, leading to a low cross correlation value. Dr. Sayeed's focus on voiced stop consonants leads to an absurd result where the disclosed system itself would not practice the claim.

D. Meisel in View of Kaneda Discloses Identifying Information of Acoustic Signals as Noise When Difference Parameters are Less Than the Gain Threshold

52. Dr. Sayeed states that a POSITA would not have been motivated to combine Meisel with Kaneda because Meisel does not disclose any noise detection process. Ex. 2004, ¶ 97. I disagree.

53. Meisel discloses noise detection because its system is designed for use "in the normal work environment of the speaker, particularly under conditions where there is background noise." Ex. 1005, 1:10-15. As explained in my previous declaration, Meisel's system generates a difference between two microphone signals (A-B) and compares the difference to a threshold (T1); and if the

difference is greater than T1, Meisel's system recognizes the signals as speech. Ex. 1003, ¶¶ 89-101. If the difference is less than T1, Meisel's system recognizes the signals as noise. Ex. 1003, ¶¶ 89-101.

54. In my opinion, a POSITA would have understood that two mouth microphones placed at different locations relative to the speaker's mouth (e.g., side and front of the mouth) produce speech signals with different amplitudes. The front microphone would have a higher signal amplitude than the side microphone because the front microphone is located directly in front of the speaker's mouth whereas the side microphone is positioned at an angle relative to the speaker's mouth. If the side mouth microphone (signal B) has a signal amplitude greater than that of the front mouth microphone (signal A), a POSITA would understand that the signal is noise. Also, if the front mouth microphone has a signal amplitude greater than the side mouth microphone, but the difference between the signals is less than T1, including a situation where the difference is zero because the two signals are identical, a POSITA would have understood that the signal is noise. Thus, in my opinion, Meisel's processing step using T1 detects noise.

55. Dr. Sayeed also argues that Meisel does not detect noise because it uses a "Boolean output 61" to "determine if a plosive or fricative is being pronounced." Ex. 2004, ¶¶ 101-102. Dr. Sayeed misunderstands Meisel. As I explained in my previous declaration, Meisel's system uses two gain thresholds,

T1 and T2, for speech recognition. Ex. 1003, ¶ 83-101. The T1 threshold is set to filter signals associated with background noise before proceeding to the next step. Ex. 1003, ¶ 83-101. It would have been well within a POSITA's skill to set T1 appropriately to filter low level background noise while leaving speech signals intact. Ex. 1003, ¶¶ 83-88. For example, each of Meisel's ADCIs is designed to pick up signals with different amplitudes, and thus, a POSITA would have set different T1 for each ADCI to accurately distinguish noise and speech. Ex. 1003, ¶¶ 83-88. The speech signals then proceed to the next comparison against T2, which is designed to detect different types of speech, for example, voiced or unvoiced speech. Ex. 1003, ¶¶ 83-88. Thus, each ADCI unit detects noise at the T1 step before it outputs a Boolean variable. Ex. 1003, ¶¶ 83-88.

56. Dr. Sayeed argues that Meisel "faults Sakoe for using one of its two microphones for non-speech purposes." Ex. 2004, ¶ 99. I disagree. Sakoe's purported shortcoming would have been one of the reasons why a POSITA would have been motivated to add another speech microphone to achieve more accurate speech recognition. Meisel explains that, in Sakoe, "a single microphone yields no direct clue as to the physiological source (e.g., mouth, nose, or throat) of the sound, even though this information is valuable in recognition." Ex. 1005, 1:58-61. This would have motivated a POSITA to

combine Meisel and Kaneda to solve this known problem and improve the accuracy of speech recognition.

57. Dr. Sayeed argues that “under Petitioner’s proposed system, if there are similar amplitude signals from two microphones, then that portion of the entire output would be classified as noise.” Ex. 2004, ¶ 105. This is not true. As explained in paragraph 35 above, Meisel would classify signals as noise only if two signals have sufficiently similar amplitudes (i.e., $A-B$ is less than $T1$), or if signal B is greater than signal A. A POSITA would have understood that, during speech, microphones at different locations, for example, front mouth microphone (signal A) and side mouth (signal B) microphone, should have different speech signal amplitudes because they are positioned differently with respect to the sound source. Furthermore, during speech, the side mouth microphone would have lower signal amplitudes because it is at an angle with respect to the speaker’s mouth, unlike the front mouth microphone, which is directly in front of the speaker’s mouth. If these criteria are not met, Meisel properly determines the signals as noise.

VII. GROUNDS 2-3

58. Dr. Sayeed does not separately opine on the patentability grounds 2 and 3, except for his reliance relative to ground 1. Ex. 2004, ¶¶ 108-109. Thus, it is still my opinion that claims 3 and 4 are also unpatentable.

VIII. CONCLUSION

59. For the reasons set forth above, it is my opinion that one skilled in the art would have found claims 1-5 of the '058 patent obvious.

60. In signing this declaration, I understand 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 acknowledge that I may be subject to cross-examination in this 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.

Respectfully submitted,

Date: October 2, 2024

By: 
David V. Anderson