

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

GOOGLE LLC,

Petitioner,

v.

JAWBONE INNOVATIONS, LLC,

Patent Owner.

Case No. IPR2022-00623

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

DECLARATION OF DAVID V. ANDERSON

TABLE OF CONTENTS

I.	INTRODUCTION	1
II.	QUALIFICATIONS	2
III.	MATERIALS CONSIDERED	4
IV.	LEGAL STANDARDS	6
V.	LEVEL OF ORDINARY SKILL IN THE ART	10
VI.	CLAIM CONSTRUCTION	11
VII.	BACKGROUND OF THE '058 PATENT	13
	A. The '058 Patent	13
	B. Prosecution History of the '058 Patent	16
VIII.	GROUND OF UNPATENTABILITY.....	17
IX.	CLAIMS 1-5 WOULD HAVE BEEN OBVIOUS TO A PERSON OF ORDINARY SKILL IN THE ART BASED ON THE PRIOR ART IN GROUND 1-3.....	18
	A. [Ground 1] Claims 1-2 and 4-5 are rendered obvious by Meisel in view of Kaneda and Freidin	18
	1. Overview of Meisel.....	18
	2. Overview of Kaneda	24
	3. Overview of Freidin	25
	4. Independent Claim 1	27
	a. [1Preamble] “A system for detecting voiced and unvoiced speech in acoustic signals having varying levels of background noise, comprising:”	27
	b. [1a] “at least two microphones that receive the acoustic signals;”	30

c.	[1b] “at least one voicing sensor that receives physiological information associated with human voicing activity; and”	32
d.	[1c] “at least one processor coupled among the microphones and the voicing sensor,”	34
e.	[1d] “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;”	36
	a. A POSITA would have been motivated to combine Meisel and Freidin.....	42
	a. A POSITA would have reasonably expected to succeed in modifying Meisel with Freidin.....	44
f.	[1e] “identifies 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;”	48
	a. A POSITA would have been motivated to combine Meisel and Freidin.....	52
	a. A POSITA would have reasonably expected to succeed in modifying Meisel with Freidin.....	53
g.	[1f] “generates difference parameters between the acoustic signals received at each of the two receivers, wherein the difference parameters are representative of the relative difference in signal gain between portions of the received acoustic signals;”	54
h.	[1g] “identifies information of the acoustic signals as unvoiced speech when the difference parameters exceed a gain threshold; and”	61

i.	[1h] “identifies information of the acoustic signals as noise when the difference parameters are less than the gain threshold.”	67
i.	A POSITA would have been motivated to combine Meisel and Kaneda	70
ii.	A POSITA would have reasonably expected to succeed in modifying Meisel with Kaneda	74
5.	Independent Claim 2	76
a.	[2Preamble] “A method for removing noise from acoustic signals, comprising:”	76
b.	[2a] “receiving the acoustic signals at two receivers and receiving physiological information associated with human voicing activity at a voicing sensor;”	78
c.	[2b] “generating cross correlation data between the physiological information and an acoustic signal received at one of the two receivers;”	78
d.	[2c] “identifying 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;”	78
e.	[2d] “generating difference parameters between the acoustic signals received at each of the two receivers, wherein the difference parameters are representative of the relative difference in signal gain between portions of the received acoustic signals;”	78
f.	[2e] “identifying information of the acoustic signals as unvoiced speech when the difference parameters exceed a gain threshold; and”	79

g.	[2f] “identifying information of the acoustic signals as noise when the difference parameters are less than the gain threshold.”	79
6.	Dependent Claim 4: “further comprising performing denoising on the identified noise.”	79
7.	Dependent Claim 5: “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.”	80
B.	[Ground 2] Claim 3 is rendered obvious by Meisel in view of Kaneda, Freidin, and Lyons	80
1.	Overview of Lyons.....	81
2.	Dependent Claim 3: “further comprising generating the gain threshold using standard deviations corresponding to the generation of the difference parameters.”	82
a.	A POSITA would have been motivated to combine Lyons with Meisel, Kaneda, and Freidin.....	84
b.	A POSITA would have reasonably expected to succeed in modifying Meisel, Kaneda, and Freidin with Lyons	85
C.	[Ground 3] Claim 4 is rendered obvious by Meisel, in view of Kaneda, Freidin, and Sakoe.....	86
a.	A POSITA would have been motivated to combine Sakoe with Meisel and Kaneda, as modified by Freidin.....	87
b.	A POSITA would have reasonably expected to succeed in modifying Meisel, Kaneda, and Freidin with Sakoe	88
X.	CONCLUSION.....	88

I. INTRODUCTION

1. I have been retained Google LLC (“Google” or “Petitioner”) as an independent expert consultant in this proceeding before the United States Patent and Trademark Office.

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 understand that this proceeding involves U.S. Patent No. 7,246,058 (“the ’058 patent,” Ex. 1001). The application for the ’058 patent was filed on May 30, 2002, as U.S. Patent Application No. 10/159,770 (“the ’770 application”). The ’770 application issued as the ’058 patent on July 17, 2007.

5. The ’058 patent claims priority to eleven U.S. provisional patent applications: U.S. Provisional Patent Application Nos. 60/368,209; 60/368,208; 60/368,343; 60/362,170; 60/361,981; 60/362,162; 60/362,103; 60/362,161; 60/332,202, 60/335,100; and 60/294,383. The ’383 provisional application has the earliest filing date and was filed May 30, 2001.

6. I have been asked by Petitioner to provide my opinion whether the claims of the '058 patent would have been obvious to a person of ordinary skill in the art at the time of the earliest claimed priority date of the '058 patent. In performing my analysis, I have been asked to assume that the priority date is May 30, 2001. My opinions are set forth below.

7. 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

8. I am a professor in the School of Electrical and Computer Engineering at the Georgia Institute of Technology ("Georgia Tech") in Atlanta, Georgia. I received my Ph.D. in Electrical and Computer Engineering from Georgia Tech, and I received my B.S. and M.S. in Electrical Engineering from Brigham Young University. Prior to my time at Georgia Tech, I worked in the field of speech and audio signal processing designing and implementing signal processing methods for hearing aids at Sonic Innovations 1993-1994. From 1994-1999, I performed research in audio signal processing, specializing in speech enhancement. During that time, I also worked as a consultant to multiple companies developing algorithms for general speech enhancement in wireless telephone systems, speech

quality assessment, and speech detection. This work led to the filing and issuance of several patents including one on speech detection in noise and one on speech enhancement (noise removal from speech).

9. In my employment prior to Georgia Tech, as well as in my subsequent studies and research, I have worked extensively in areas related to research, design, and implementation of audio and speech processing algorithms and in speech enhancement. I have designed and taught a graduate course on adaptive signal processing, which includes the topics of estimating transfer functions, correlation, using cross-correlation to estimate signal similarity and to compute transfer functions, and extracting signals of interest from noisy signals. I have also taught graduate and undergraduate level courses at Georgia Tech on the implementation of signal processing theory, algorithms, implementation, and software. For example, I have taught courses on multimedia processing and systems, machine learning for speech, pattern recognition, software design, and signal processing applications (covering topics in audio processing, speech detection, speech processing, and speech recognition). These courses and my research have covered many topics relevant to the subject matter of the '058 patent and the cited prior art, in particular, concepts such as cross correlation, difference parameters, gain and/or detection thresholds. glottal signals, and speech modeling are all topics with which I am intimately familiar.

10. I also have extensive experience with the practical implementation of signal processing algorithms, information theory, signal detection, and related topics through my research and consulting.

11. A copy of my CV, including a list of recent litigations I have testified in, is attached as Exhibit 1004. My compensation is not contingent upon the outcome of this *inter partes* review or any other proceeding.

III. MATERIALS CONSIDERED

12. In forming my opinions, I have reviewed the following documents, and any other document cited in this declaration:

Exhibit	Description
Ex. 1001	U.S. Patent No. 7,246,058 B2 (“the ’058 patent”)
Ex. 1002	File History of U.S. Patent No. 7,246,058
Ex. 1005	U.S. Patent No. 4,718,096 (“Meisel”)
Ex. 1006	U.S. Patent No. 5,208,864 (“Kaneda”)
Ex. 1007	U.S. Patent No. 6,148,313 (“Freidin”)
Ex. 1008	U.S. Patent No. 4,239,936 (“Sakoe”)
Ex. 1009	Signal Processing Toolbox for Use with MATLAB, User’s Guide, Version 4.2 (January 1999) (“MATLAB Manual”)
Ex. 1010	<i>Reserved</i>
Ex. 1011	Steven W. Smith, The Scientist and Engineer’s Guide to Digital Signal Processing (2d ed. 1999) (“Smith”)

Exhibit	Description
Ex. 1012	Richard G. Lyons, <i>Understanding Digital Signal Processing</i> (1996) (“Lyons”)
Ex. 1013	U.S. Patent No. 5,724,485 (“Rainton”)
Ex. 1014	U.S. Patent No. 5,774,847 (“Chu”)
Ex. 1015	U.S. Patent No. 6,694,028 (“Matsuo”)
Ex. 1016	Vinay K. Ingle and John G. Proakis, <i>Digital Signal Processing Using MATLAB</i> (2000) (“Ingle”)
Ex. 1017	U.S. Patent No. 6,219,645 (“Byers”)
Ex. 1018	U.S. Patent No. 6,246,760 (“Makino”)
Ex. 1019	<i>Reserved</i>
Ex. 1020	U.S. Patent No. 6,760,449 (“Matsuo 2”)
Ex. 1021	U.S. Patent No. 6,766,019 (“Benesty”)
Ex. 1022	Scheduling Order, <i>Jawbone Innovations, LLC v. Google LLC</i> , No. 6:21-cv-00985, Dkt. 27 (W.D. Tex. Sep. 23, 2021)
Ex. 1023	Richard O. Duda and Peter E. Hart, <i>Pattern Classification and Scene Analysis</i> (1973) (“Duda”)
Ex. 1024	David M. Green and John A. Swets, <i>Signal Detection Theory and Psychophysics</i> (1988) (“Green”)
Ex. 1025	J. R. Deller et al., “ShortTerm Processing of Speech,” in <i>Discrete-Time Processing of Speech Signals</i> , IEEE (2000) (“Deller”)
Ex. 1026	Docket Control Order, <i>Jawbone Innovations, LLC v. Samsung Electronics Co., Ltd. et al.</i> , No. 2:21-cv-00186, Dkt. 38 (E.D. Tex. Feb. 1, 2022)
Ex. 1027	Declaration of Lauren Gluckman regarding Lyons

Exhibit	Description
Ex. 1028	Online Copyright Office Public Catalog Record of Lyons

13. 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.

14. 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 (in this case, the filing date of the '058 patent): May 30, 2001.

15. 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

1. For purposes of this declaration, I have been asked to opine only on issues regarding obviousness under 35 U.S.C. § 103. I have been informed of the following legal standards, which I have applied in forming my opinions.

2. I have been asked to provide my opinions as to whether the cited prior art teaches or renders obvious claims 1-5 of the '058 patent from the perspective of a person of ordinary skill in the art at the '058 patent's priority date in 2001, as described in more detail below.

3. For purposes of this declaration, I have been informed and understand certain aspects of the law as it relates to my opinions.

4. I have been advised and understand that there are two ways in which prior art may render a patent claim unpatentable. First, I have been advised that the prior art can "anticipate" a claim. Second, I have been advised that the prior art can make a claim "obvious" to a person of ordinary skill in the art. I understand that for an invention claimed in a patent to be patentable, it must not be anticipated and must not be obvious based on what was known before the invention was made.

5. I have been advised and understand the information used to evaluate whether an invention was new and not obvious when made is generally referred to as "prior art." I understand that prior art includes patents and printed publications that existed before the earliest filing date of the patent (which I have been informed is also called the "effective filing date"). I have been informed and understand that a patent or published patent application is prior art if it was filed before the earliest filing date of the claimed invention and that a printed publication is prior art if it was publicly available before the earliest filing date.

6. I have been advised and understand that a dependent claim is a patent claim that refers back to another patent claim. I have been informed and understand that a dependent claim includes all of the limitations of the claim to which it refers.

7. I have been advised and understand that a patent claim may be invalid as obvious under 35 U.S.C. § 103 if the differences between the subject matter claimed and the prior art are such that the claimed subject matter as a whole would have been obvious to a person of ordinary skill in the art at the time the invention was made. I have also been advised that several factual inquiries underlie a determination of obviousness. These inquiries include (1) the scope and content of the prior art, (2) the level of ordinary skill in the field of the invention, (3) the differences between the claimed invention and the prior art, and (4) any objective evidence of non-obviousness (which I have been informed may also be called “secondary considerations”).

8. I have also been advised and understand that, when obviousness is based on a combination of references, that party must identify a reason why a person skilled in the art would have been motivated to combine or modify the asserted references in the manner recited in the claims and to explain why one skilled in the art would have had a reasonable expectation of success in making such combinations or modifications.

9. I have been advised and understand that the law permits the application of “common sense” in examining whether a claimed invention would have been obvious to a person skilled in the art. For example, I have been advised that combining familiar elements according to known methods and in a predictable way may suggest obviousness when such a combination would yield nothing more than predictable results. I understand, however, that a claim is not obvious merely because every claim element is disclosed in the prior art and that a party asserting obviousness must still provide a specific motivation to combine or modify the references as recited in the claims and explain why one skilled in the art would have reasonably expected to succeed in doing so.

10. I have been advised and understand that two references are considered to be in the same field of art when the references are either (1) in the same field of endeavor, regardless of the problems they address, or (2) reasonably pertinent to the particular problem being solved by the inventor in his patent.

11. I am not aware of any evidence of secondary considerations that would support a determination of non-obviousness of the claimed subject matter in the '058 patent.

16. I have been informed that in *inter partes* review proceedings, such as this one, the party challenging the patent bears the burden of proving

unpatentability by a preponderance of the evidence. I understand that a preponderance of the evidence means “more likely than not.”

V. LEVEL OF ORDINARY SKILL IN THE ART

12. In rendering the opinions set forth in this declaration, I have been asked to consider the '058 patent's claims and the prior art through the eyes of a person of ordinary skill in the art (which I may also refer to as “one skilled in the art,” “skilled artisan,” “POSITA,” or similar variation). I have considered factors such as the educational level and years of experience of those working in the pertinent art, the types of problems encountered in the art, the teachings of the prior art, patents and publications of other persons or companies, and the sophistication of the technology.

13. I have been instructed to assume a person of ordinary skill in the art is not a specific real individual, but rather a hypothetical individual having the qualities reflected by the factors discussed above.

14. Taking these factors into consideration, it is my opinion that a person of ordinary skill in the art of the '058 patent as of its filing date, would have at least 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

formal education and additional formal education can substitute for work experience.

15. By May 30, 2001, I was at least a person of ordinary skill in the art, as shown by my qualifications and work experience above.

16. In this declaration, and for all of my opinions herein, I have applied the knowledge of a person of ordinary skill in the art as of May 30, 2001.

VI. CLAIM CONSTRUCTION

17. I have been instructed that the words of a claim are typically given their ordinary and customary meaning, as they would have been understood by a person of ordinary skill in the art at the time of the invention, in view of the intrinsic record (discussed below). In this case, I have been instructed to assume that the “time of the invention” for purposes of claim construction is May 30, 2001, which is the earliest claimed priority filing date for the ’058 patent. The opinions herein pertain to that time frame, except where expressly stated otherwise.

18. I have been instructed that the “intrinsic record” includes the patent itself, including the claims, description, and figures (Ex. 1001), and the patent’s prosecution history—i.e., the record of proceedings at the U.S. Patent and Trademark Office (“Patent Office”) concerning the patent (Ex. 1002). I understand that, like the claims and written description, the prosecution history provides evidence to a person of ordinary skill in the art of how the inventor intended his

patent to be understood, and how the Patent Office understood the patent. I understand that the inventor is permitted to apply a special definition to the terms or to limit the scope of claim terms in his patent claims, which may differ from the term's plain and ordinary meaning. That special definition or limitation on scope may be provided in the patent's written description, the patent's prosecution history, or both.

19. I understand that claim interpretation may also be informed by “extrinsic evidence” (that is, evidence outside of the patent record itself). I have been informed that extrinsic evidence may include dictionaries, technical treatises, and other materials evidencing the meaning of a claim term and the understanding held by a POSITA in the relevant time period.

20. I have been asked for purposes of this declaration to apply the plain and ordinary meaning of the claim terms as they would have been understood by a person of ordinary skill in the art at the earliest claimed priority date of the '058 patent (May 30, 2001).

21. I understand that no claim term's construction is in dispute at this time. Based on my review of these materials and my personal knowledge and experience, I have considered each term of the '058 patent as it would have been understood by one skilled in the art as of May 30, 2001.

VII. BACKGROUND OF THE '058 PATENT

A. The '058 Patent

22. The '058 patent describes a system and a method of detecting voiced and unvoiced speech using both acoustic and non-acoustic sensors, depicted in Figures 1 and 3. Ex. 1001, Abstract. The patent describes characterizing speech signals as voiced and unvoiced and, in certain embodiments, distinguishing the speech signals from noise signals. Ex. 1001, 4:36-41. Such methods may be used in speech recognition. Ex. 1001, 4:36-41.

23. As shown in annotated Figures 1 and 3, below, the system includes at least one **processor (30)**, which is connected to two **acoustic microphones 10 (MIC 1 and MIC 2)** that receive acoustic signals ($m_1(n)$ and $m_2(n)$) and at least one **voicing sensor 20 (VAD)** that receives physiological information associated with human voicing activity, such as vocal fold vibrations. Ex. 1001, 2:30-37, 3:64-4:13, 5:26-29. The '058 patent discloses that **voicing sensor 20** may be a radio frequency device, an electroglottograph, an ultrasound device, an acoustic throat microphone, or an airflow detector. Ex. 1001, 5:26-29, Claim 5.

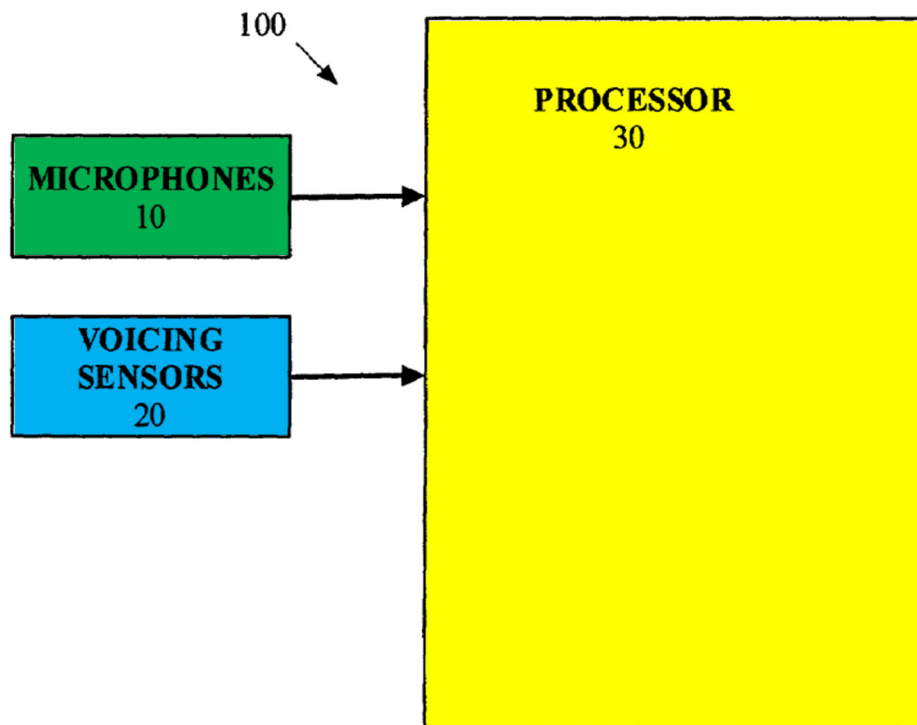


Figure 1

Ex. 1001, FIG. 1 (modified, annotated).

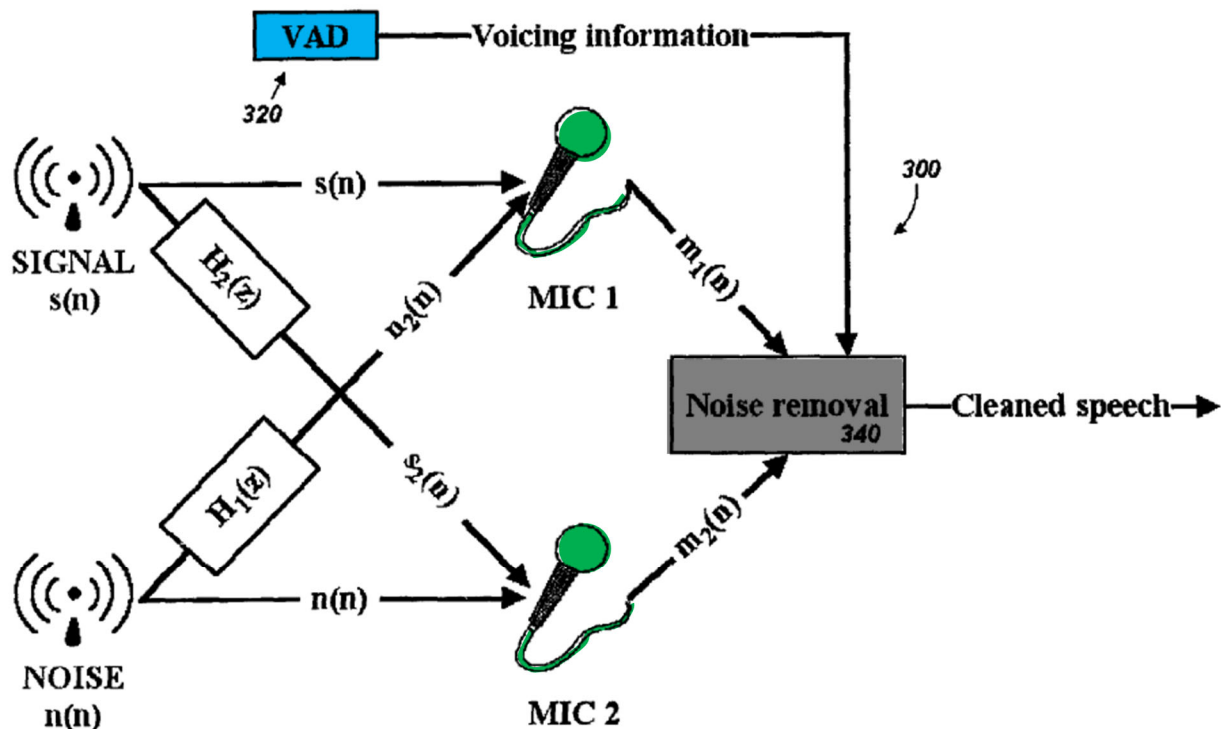


Figure 3

Ex. 1001, FIG. 3 (annotated).

24. **Processor 30** generates cross-correlation data correlating the physiological information (“voicing information” in Figure 3) received from the **voicing sensor** and the acoustic signal received from one of the **microphones**. See Ex. 1001, 4:64-5:3, 5:11-25. **Processor 30** then compares the cross correlation data to an empirical correlation threshold. Ex. 1001, 5:39-45. If the correlation data exceeds the correlation threshold, the processor identifies information in the acoustic signal as voiced speech. Ex. 1001, 5:39-45. **Processor 30** also compares the acoustic signals received at the two microphones to generate

difference parameters, which represent a relative difference in signal gain between portions of the acoustic signals. Ex. 1001, 5:39-45, 7:52-67. The processor compares the difference parameters to a predetermined gain threshold, that is “a running average of the standard deviations and the standard deviation noise floor.” Ex. 1001, 8:43-48. If the difference parameters exceed the gain threshold, the processor identifies information in the acoustic signals as unvoiced speech. Ex. 1001, 8:21-41. If the difference parameters are less than the gain threshold, then the processor determines that the information in the acoustic signals is noise. Ex. 1001, 8:41-42. The processor may then **remove the identified noise (noise removal 340)**. Ex. 1001, 7:18-19.

B. Prosecution History of the '058 Patent

25. The '058 patent was filed as U.S. Patent Application No. 10/159,770 on May 30, 2002. After the Office rejected all claims as obvious over U.S. Patent No. 5,590,241 (“Park”) in view of U.S. Patent No. 6,233,551 (“Cho”) (Ex. 1002, 132-134), the applicant argued that Park fails to disclose using two microphones and voicing activity detection. Ex. 1002, 124. The applicant further argued that Cho fails to disclose using two microphones and detecting voiced and unvoiced speech, as claimed. *Id.* at 125. The Office maintained the rejection of claims 1-5 as obvious over Park and Cho, but withdrew the rejection of original claims 6 and 7, indicating these claims were allowable. Ex. 1002, 108-110. The applicant then

canceled claims 1-5 and added claims 8-10. Ex. 1002, 100. Prosecution claims 6-10 correspond to issued claims 1-5, respectively. The '770 application issued as the '058 patent on July 17, 2007. Ex. 1001, cover.

VIII. GROUNDS OF UNPATENTABILITY

26. Based on my review of the materials set forth above, including my application of the knowledge of a person of ordinary skill in the art, it is my opinion that claims 1-5 of the '058 patent would have been obvious to a person of ordinary skill in the art.

27. In particular, it is my opinion that claims 1-5 of the '058 patent would have been obvious to a person of ordinary skill in the art based on the combinations shown below:

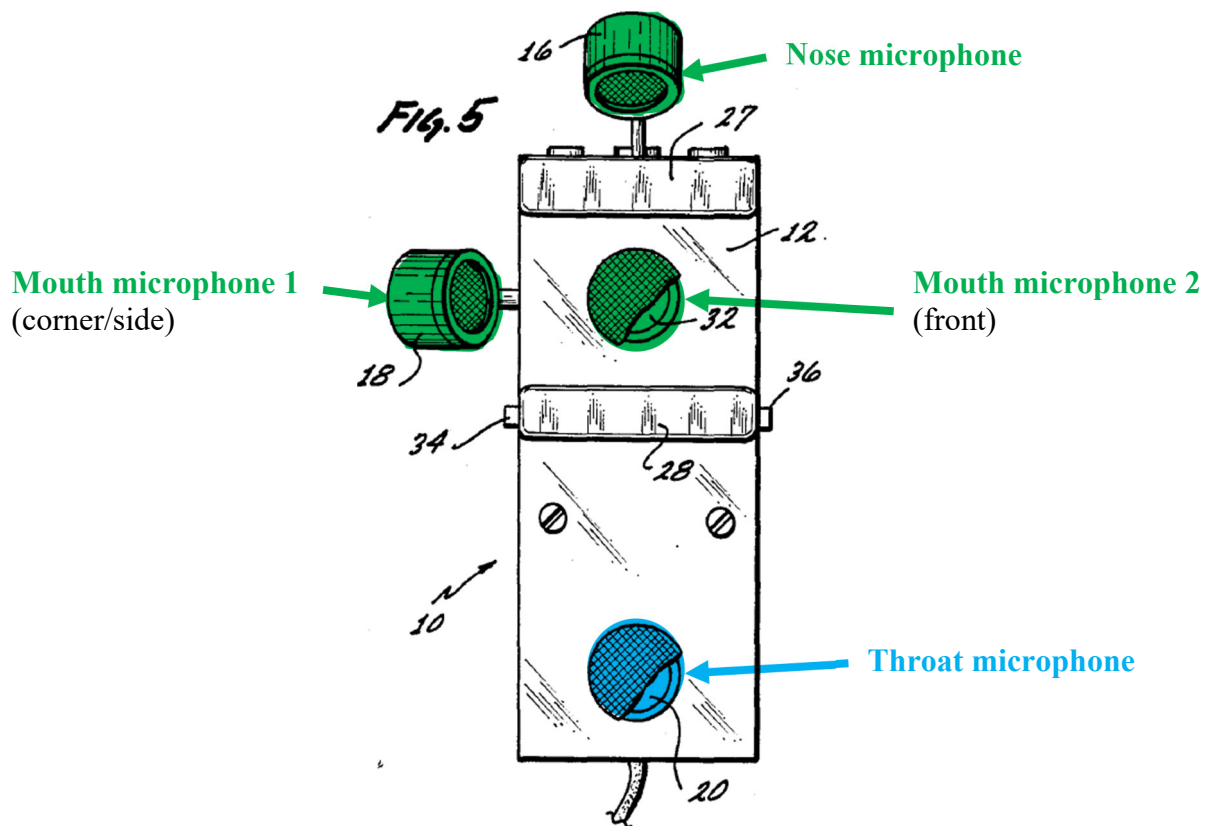
Ground 1	Claims 1-2 and 4-5 are rendered obvious by Meisel in view of Kaneda and Freidin
Ground 2	Claim 3 is rendered obvious by Meisel in view of Kaneda, Freidin, and Lyons
Ground 3	Claim 4 is rendered obvious by Meisel, in view of Kaneda, Freidin, and Sakoe

IX. CLAIMS 1-5 WOULD HAVE BEEN OBVIOUS TO A PERSON OF ORDINARY SKILL IN THE ART BASED ON THE PRIOR ART IN GROUNDS 1-3**A. [Ground 1] Claims 1-2 and 4-5 are rendered obvious by Meisel in view of Kaneda and Freidin**

28. In my opinion, claims 1, 2, 4, and 5 would have been obvious to a POSITA based on Meisel in view of Kaneda and Freidin.

1. Overview of Meisel

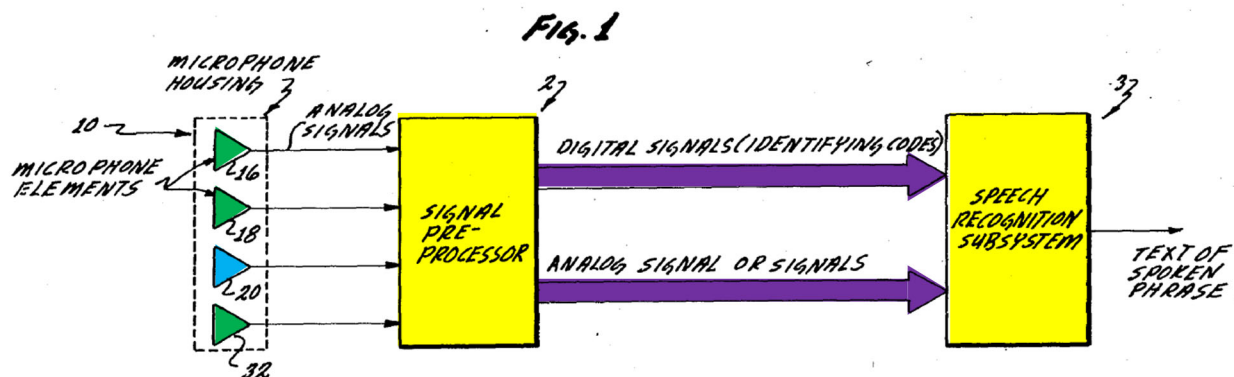
29. Meisel discloses a speech detection system comprising a plurality of microphones and a voice activity detector. Ex. 1005, Abstract. Meisel's system includes a **nose microphone (16)** and two **mouth microphones (18, 32)** (one microphone on the corner of a speaker's mouth (18) and another microphone in front of the speaker's mouth (32)). Ex. 1005, 2:11-30, 4:56-59, FIG. 5. Nose microphone 16 and mouth microphones 18, 32 are acoustic microphones that detect acoustic speech signals. *See* Ex. 1005, 8:49-59 (explaining that the nose and mouth microphones detect the "M" acoustic sound signals). Meisel's system includes a **throat microphone (20)** that detects physiological information associated with human voicing activity, such as "vibration of the vocal cords, or 'voicing.'" Ex. 1005, 2:33-35.



Ex. 1005, FIG. 5 (annotated).

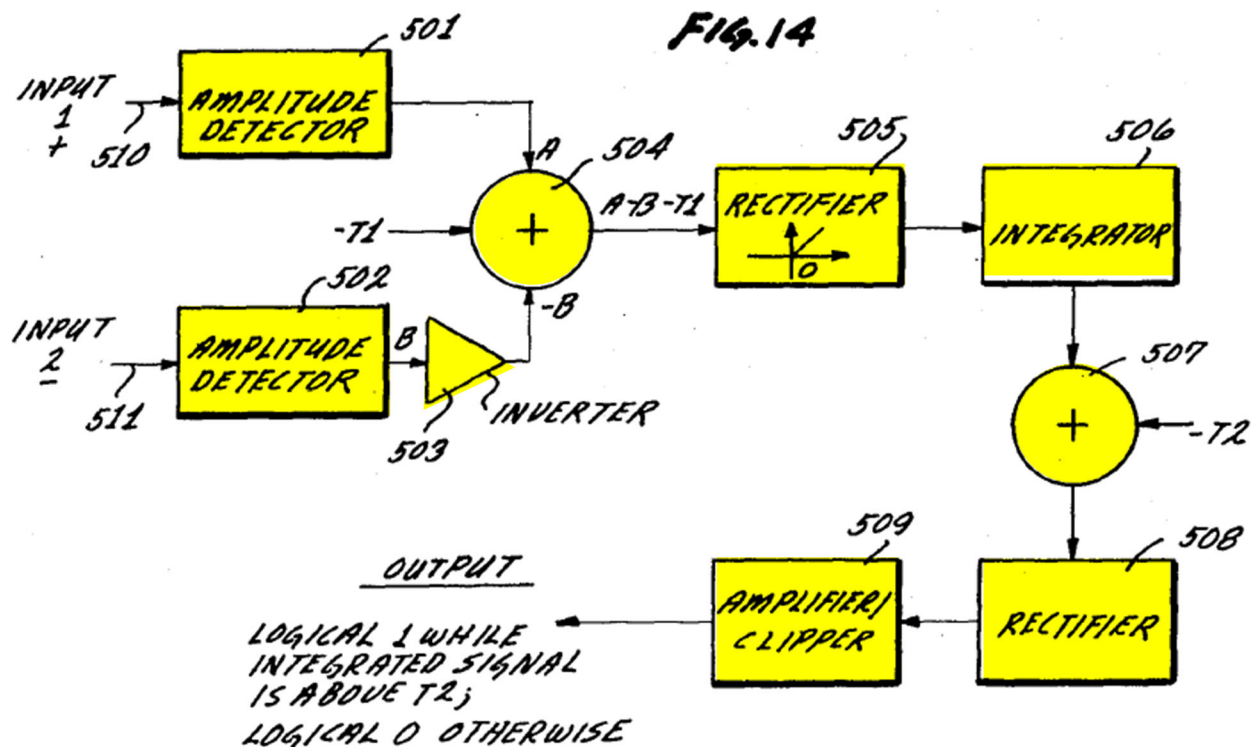
30. Meisel's speech recognition system includes at least one **processor**, such as **signal pre-processor (2)** and **speech recognition subsystem (3)**.

Ex. 1005, 4:27-34. The signals from the microphones are sent to the **signal pre-processor (2)**, which compares and detects imbalances (differences) in signal amplitudes between any two microphones. Ex. 1005, 4:35-42. The differences are coded as Boolean variables, which are used to classify speech as voiced or unvoiced by distinguishing, for example, voiced and unvoiced stop consonants and voiced and unvoiced fricatives. Ex. 1005, 2:31-48.



Ex. 1005, FIG. 1 (annotated).

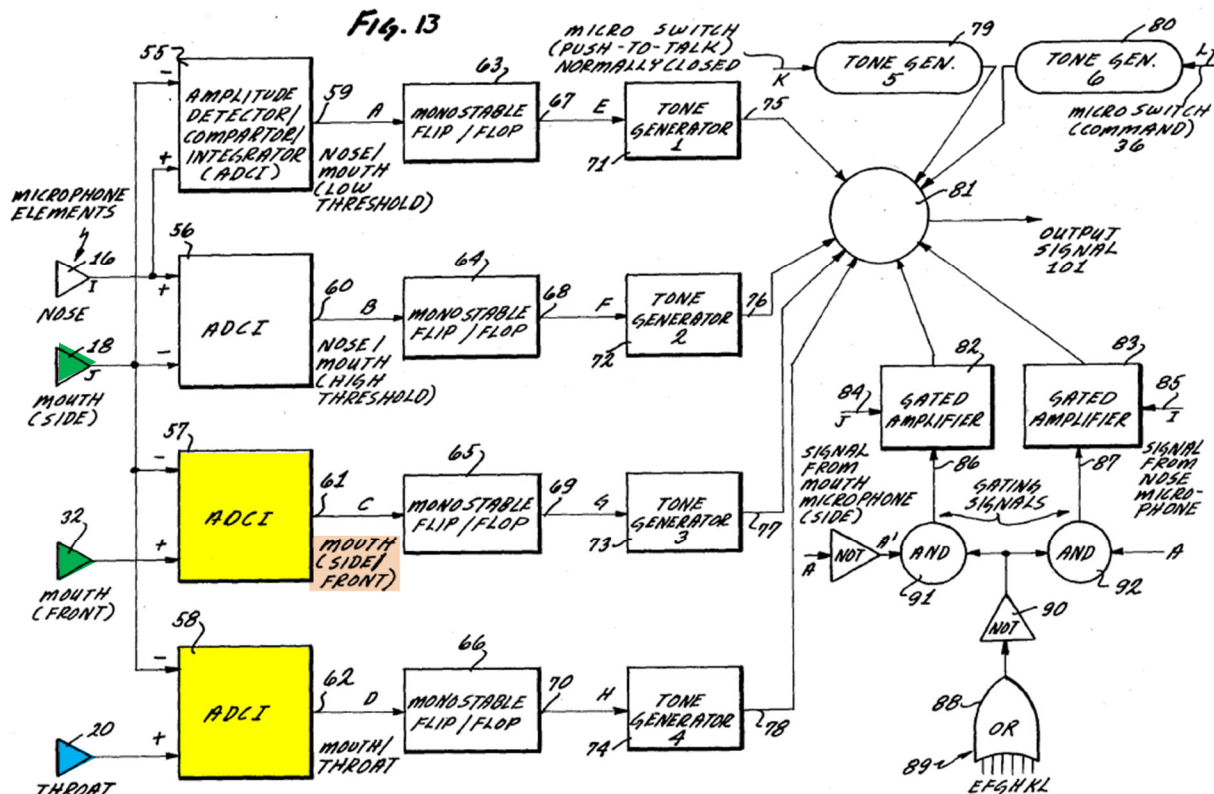
31. The operation of Meisel's signal pre-processor (2) uses amplitude detector/comparator/integrator units (ADCI) to distinguish between voiced and unvoiced speech using the differences in the two signals, as shown below.



Ex. 1005, FIG. 14 (annotated).

32. For example, Meisel's signal pre-processor (2) calculates a difference between two input signals (A-B) and compares that difference to a predetermined threshold (T1). Ex. 1005, 9:12-29. The signal pre-processor (2) further processes this information using an integrator (506), which integrates signals over a period of time, and compares the output data to a second predetermined threshold (T2). Ex. 1005, 9:38-42. For example, the signal pre-processor (2) compares the signals from the first mouth microphone (18) and the throat microphone (20) to distinguish voiced from unvoiced stop consonants and fricatives. Ex. 1005, 2:31-48. As one example, Meisel's speech detection system measures the difference in signal amplitudes (gains) between the received signals for the side mouth microphone (18) and the front mouth microphone (32). Ex. 1005, 9:38-50, 10:4-11. If the difference between the signals is greater than the threshold T1, the signal pre-processor (2) determines that the speech is voiced or unvoiced. Ex. 1005, 9:38-50. Meisel provides one example of ADCI operation in Figure 14 where "the occurrence of a nasal when the input 510 is from the nasal microphone 16 and the input 511 is from the mouth microphone 18." Ex. 1005, 9:47-50. A POSITA would understand that, because Figure 14 describes the ADCI operation generally, the signal amplitude comparison step applies equally to the inputs from the side mouth microphone (18) and the front mouth microphone (32) for ADCI unit (57) in view of Figure 13. Moreover, a POSITA would understand that the output of the

comparison of the side mouth microphone (18) and the front mouth microphone (32) in ADCI unit (57) would be combined with the output of the throat microphone (20) to distinguish voiced and unvoiced speech, further refining the speech recognition process. Ex. 1005, 2:31-48 (explaining that the throat microphone (20) would be used in combination with other microphones to distinguish voiced and unvoiced speech). In particular, Meisel explains that “[a] similar argument applies to distinguishing voiced fricatives (v, z, and th) from unvoiced fricatives (f, s, and sh).” Ex. 1005, 2:46-48. Thus, a POSITA would recognize that Meisel’s voicing detection (throat microphone (20)) and fricative detection (ADCI unit (57)) would be combined to specify whether the fricative is voiced or unvoiced.

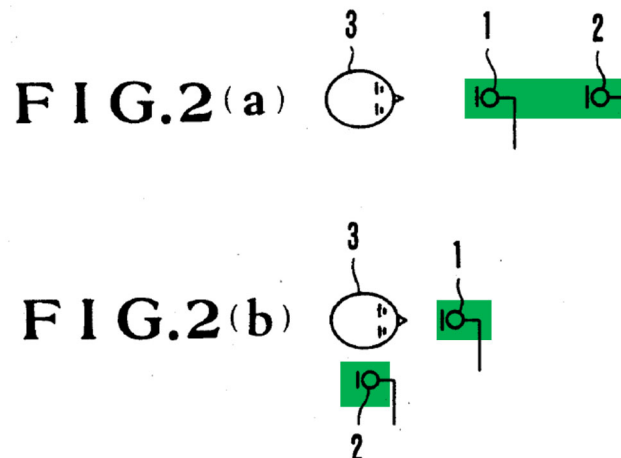


Ex. 1005, FIG. 13 (annotated).

33. Meisel further discloses that its speech detection system operates in conditions with varying levels of background noise, such as “under conditions where there is background noise and where the speaker can continue to perform other duties,” suggesting that it operates in environments where noise is present and should be removed in processing. Ex. 1005, 1:10-15, 1:46-50. In Figure 14, if signal A does not exceed signal B by a threshold amount, T1 (“A-B-T1” in Figure 14 is greater than zero), the summing unit 504 does not provide an output, rejecting the signals as noise. Ex. 1005, 9:12-25.

2. Overview of Kaneda

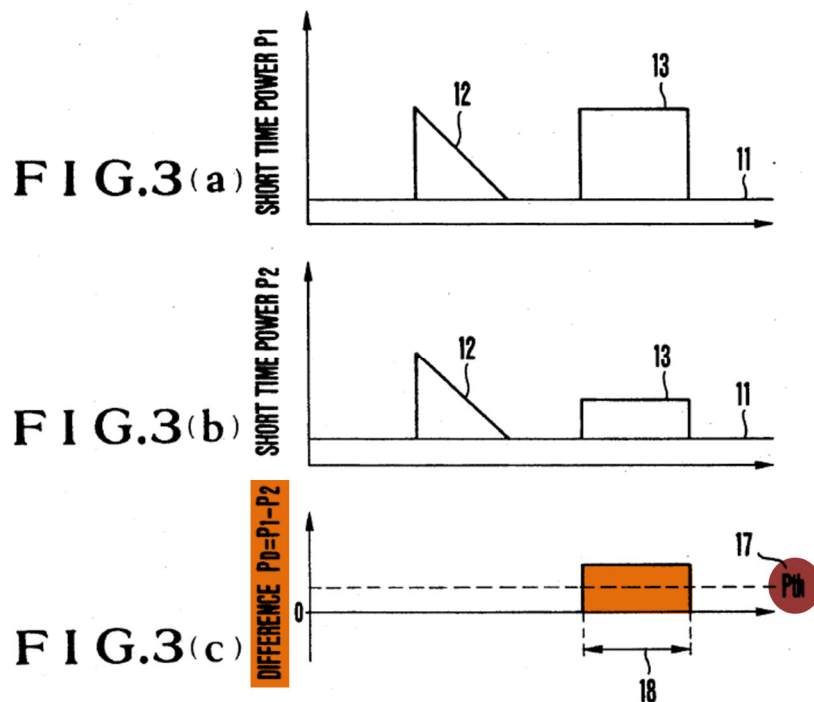
34. Kaneda discloses a conventional speech recognition system including **two acoustic microphones** that are located at different positions relative to the speaker; for example, one microphone is located near a speaker's mouth (microphone 1) and the other microphone is located away from the speaker (microphone 2). Ex. 1006, 1:55-57, FIG. 2(a). Kaneda discloses that speech signals received at microphone 1 result in a higher power output than those received at microphone 2. Ex. 1006, 1:61-63. This is shown in Figures 3(a) and 3(b), in which the speech signal power (13) of microphone 1 (P_1) is greater than that of microphone 2 (P_2). Ex. 1006, 2:1-12.



Ex. 1006, FIGS. 2(a)-2(b) (annotated).

35. To determine the presence of a speech, Kaneda discloses that the “conventional” speech recognition system calculates a difference between the power outputs of the two microphones, “PD ($=P_1-P_2$).” Ex. 1006, 2:12-15. “When

the **power difference PD** is larger than a given **threshold value Pth17**, a corresponding time period 18 is detected as a speech period.” Ex. 1006, 2:12-18. Similarly, a POSITA would understand that when PD is less than Pth17, the corresponding time period is noise. Ex. 1006, 2:18-22.



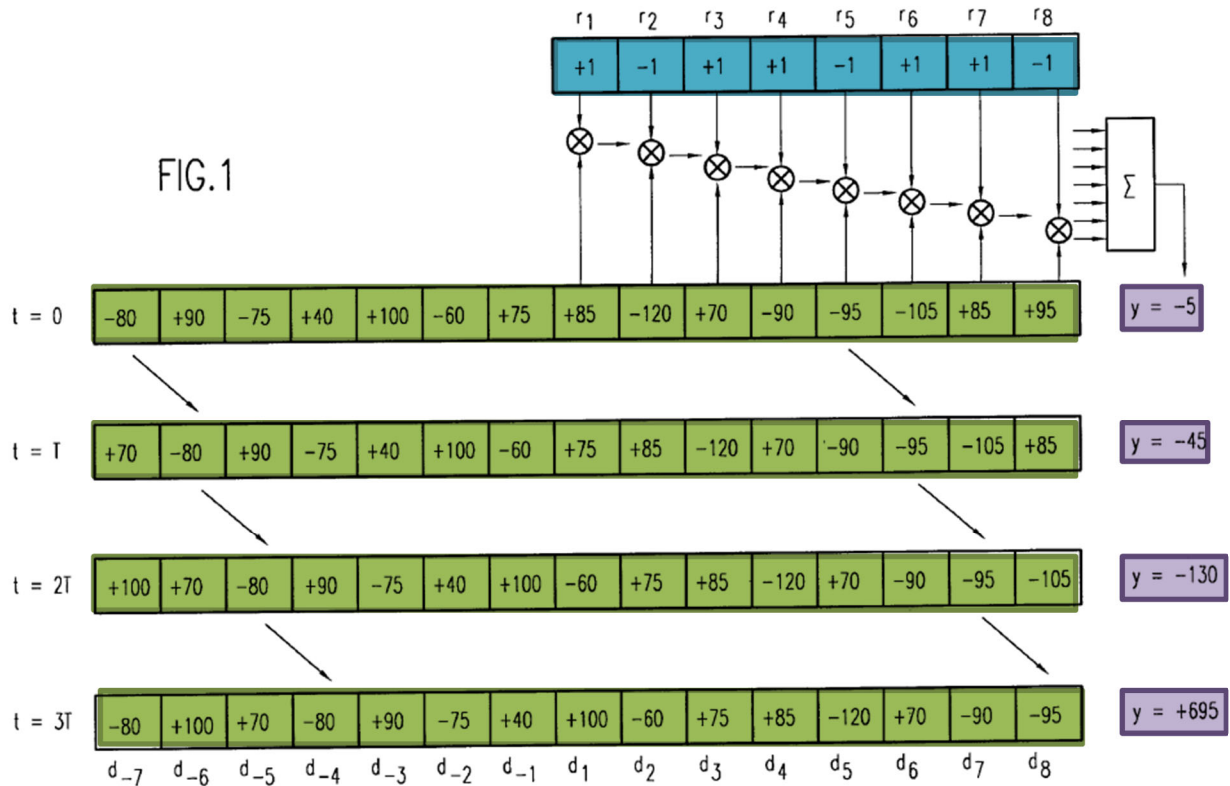
Ex. 1006, FIGS. 3(a)-3(c) (annotated).

3. Overview of Freidin

36. Freidin discloses a correlator used for determining a correlation between digital signals for use in “a number of applications including speech recognition[.]” Ex. 1007, 1:7-26. Freidin explains that Figure 1 shows, as an example, “a communication receiver correlator operating with a **correlation reference sequence**, which is eight (8) bits in length, to produce a correlation peak

at the time when the **received data sequence** matches the reference sequence.”

Ex. 1007, 1:37-41.



Ex. 1007, FIG. 1.

37. Freidin discloses that “[a]t each clock period, the last eight samples of input data d_i ($i=1,8$) are multiplied sample by sample against the reference sequence bits r_i ($i=1,8$) to produce a **correlation value y** .” Ex. 1007, 1:43-46.

Freidin also discloses that the “correlation value y is compared against a **correlation threshold**, y_{TH} , and whenever y exceeds y_{TH} , a correlation match is declared indicating that the correlation sequence has been received and the receiver is thereby synchronized to the transmitter.” Ex. 1007, 1:48-52. Freidin also

explains that “[w]hile preferred embodiments are disclosed herein, many variations are possible which remain within the concept and scope of the invention. Such variations would become clear to one of ordinary skill in the art after inspection of the specification, drawings and claims herein. The invention therefore is not to be restricted except within the spirit and scope of the appended claims.” Ex. 1007, 11:49-12:4.

4. Independent Claim 1

38. In my opinion, Meisel teaches or suggests each element of claim 1. Kaneda confirms and renders obvious calculating difference parameters between the acoustic signals received at each of the two receivers, as recited in [1a]. Freidin confirms and renders obvious generating cross correlation data between physiological information and acoustic signals in [1d], and the cross correlation data exceeding a correlation threshold, as recited in [1e].

a. [1Preamble] “A system for detecting voiced and unvoiced speech in acoustic signals having varying levels of background noise, comprising:”

39. Meisel discloses a system for detecting voiced and unvoiced speech in acoustic signals having background noise levels. For example, Meisel discloses using a throat microphone, a nose microphone, and one or more mouth microphones to detect and classify a speaker’s speech as voiced or unvoiced when

the signals are processed by the pre-processing system. Ex. 1005, 2:11-48, 2:61-3:4, 10:6-11, 13:10-22, 14:57-61.

40. Meisel's system compares "the signal from the throat microphone [20] to that of the mouth microphone [18]" to "distinguish, for example, voiced from unvoiced stop consonants" and "voiced fricatives (v, z, and th) from unvoiced fricatives (f, s, and sh)." Ex. 1005, 2:31-53. A POSITA would understand that distinguishing voiced from unvoiced stop consonants and voiced from unvoiced fricatives are examples of distinguishing voiced from unvoiced speech. For example, Meisel provides examples for distinguishing voiced from unvoiced speech:

All vowels and some consonants are voiced. By comparing the signal from the throat microphone to that of the mouth microphone, one can distinguish, for example, voiced from unvoiced stop consonants. 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. For unvoiced stops (p, t, and k), the vocal cords do not vibrate and little sound comes from the throat as well. For voiced stops (b, d, and g), however, the vocal cords vibrate during the closure.

Ex. 1005, 2:35-45.

41. Meisel further discloses that the sounds received from the nose and mouth microphones are acoustic signals. Ex. 1005, Claim 6 (“a first microphone disposed adjacent said first end for detecting acoustic sounds substantially emanating from a corner of a speaker’s mouth. . . one of said second and third microphones for detecting acoustic sounds substantially emanating from the speaker’s nose.”), 10:22-29.

42. Meisel suggests that the acoustic signals have varying levels of background noise. For example, Meisel explains that “close-talking” microphones, which receive speech signals as well as noise (e.g., background noise), were known in the prior art, suggesting that conventional microphones known at the time received different levels of noise along with speech. Ex. 1005, 2:54-60. In the Background, Meisel identifies several problems with “background noise” making it difficult for prior art systems to distinguish speech. Ex. 1005, 1:7-50. A POSITA would, therefore, have understood that Meisel’s system operates in environments having varying background noise.

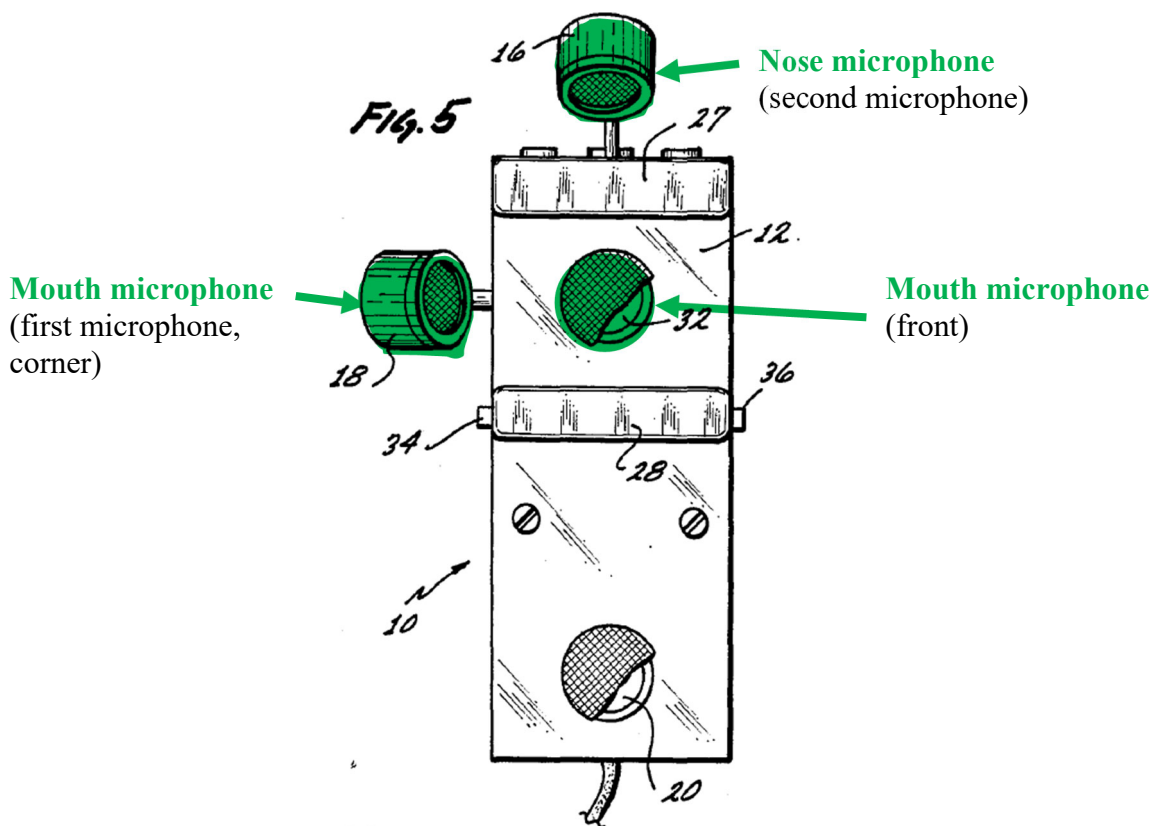
43. Moreover, in my opinion, “varying levels of background noise” as recited in the claim would have been obvious in view of Kaneda.

44. Kaneda, like Meisel, is directed to a speech detection system including two microphones. *Supra* §§ IX.A.1, 0. As such, a POSITA would have

looked to Kaneda to understand the operation of the acoustic microphones in Meisel's system (e.g., nose and mouth microphones). Ex. 1005, 1:7-16, 2:11-53. Kaneda discloses that microphone signals "generally contain[] stationary noise 11 (noise having almost a constant power, such as air-conditioning noise or fan noise of equipment), unstationary noise 12 (noise whose power is greatly changed, such as a door closing sound and undesired speech), and desired speech 13." Ex. 1006, 1:30-35. A POSITA would understand that stationary and unstationary noise refers to varying levels of background noise that are commonplace in environments for Meisel's system, such as offices.

b. [1a] "at least two microphones that receive the acoustic signals;"

45. Meisel discloses three microphones that receive the acoustic signals. For example, Meisel discloses a **nose microphone (16)** that detects acoustic sounds emanating from the speaker's nose and two **mouth microphones (18, 32)** that detect acoustic sounds emanating from the speaker's mouth. Ex. 1005, FIG. 5, Abstract, 5:29-36. Meisel explains that a **first microphone (mouth microphone 18)** "is disposed so as to detect sounds substantially emanating from a corner of the speaker's mouth" and a **second microphone (nose microphone 16)** "is disposed so as to detect sounds substantially emanating from the nose of the speaker." Ex. 1005, 2:11-17.



Ex. 1005, FIG. 5 (annotated).

46. The nose microphone (16) and the mouth microphones (18, 32) are acoustic microphones that receive acoustic signals from the speaker's nose and mouth, respectively. Meisel discloses that the mouth microphones 18, 32 are a "first means for detecting *acoustic sounds* substantially emanating from the mouth of the speaker and for producing signals representing such sounds," and that the nose microphone 16 is the "second means for detecting *acoustic sounds* substantially emanating from the nose of the speaker and for producing signals representing such sounds." Ex. 1005, Claim 1 (emphases added), 5:29-36; Meisel explains that the nose microphone 16 detects the nasal acoustic signals or "the

nasal component in order to provide a reliable recognition of speech content.”

Ex. 1005, 8:49-59. Meisel further explains that the throat microphone (20) and the second mouth microphone (32) detect “the sounds emanating from the throat or immediately in front of the mouth for certain words in order to provide a reliable recognition of such spoken words.” Ex. 1005, 8:55-59.

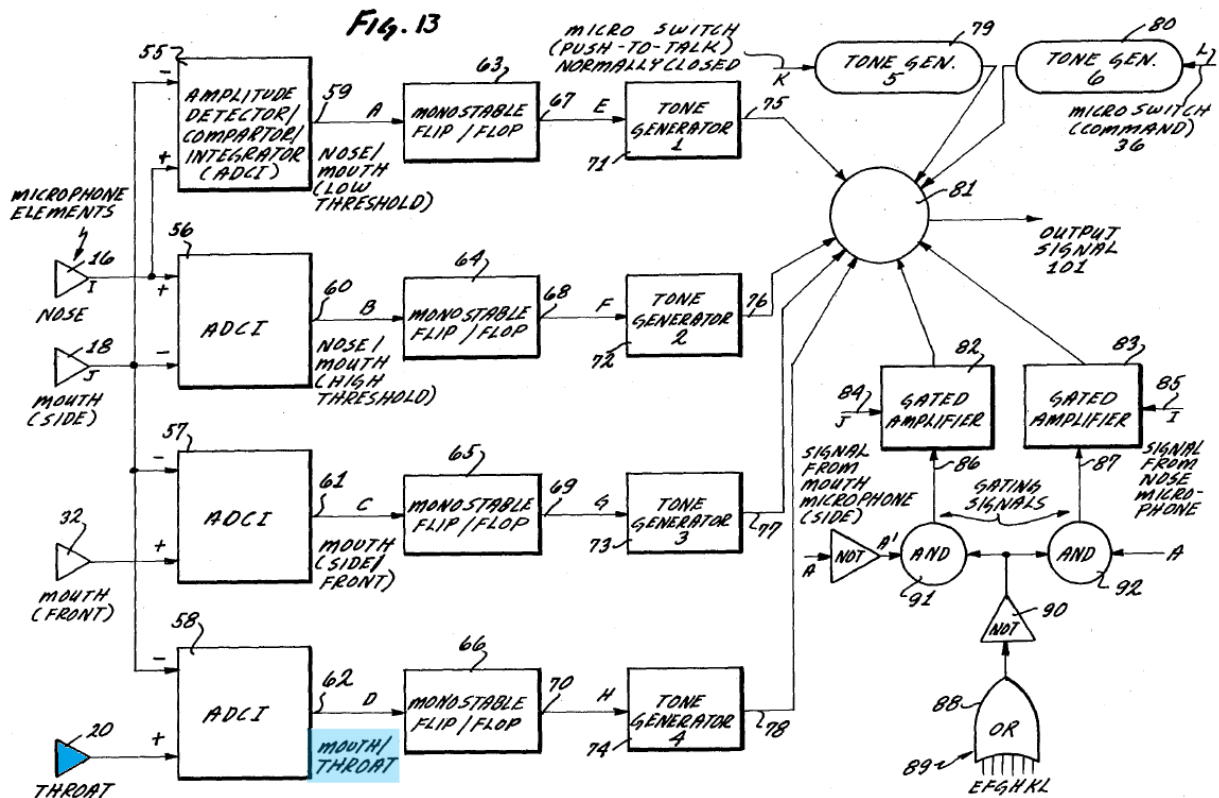
c. [1b] “at least one voicing sensor that receives physiological information associated with human voicing activity; and”

47. Meisel discloses at least one **voicing sensor** that receives physiological information associated with human voicing activity, such as vibrations from the user’s throat via acoustic throat microphone 20.

48. The ’058 patent explains that voicing sensors include “any sensor that is able to detect human tissue motion associated with speech,” (Ex. 1001, 3:61-63), including an “acoustic throat microphone[],” such as Meisel’s throat microphone. Ex. 1001, Claim 5. Meisel states that its speech recognition system is “usable in the normal work environment of the speaker, particularly under conditions where there is background noise and where the speaker can continue to perform other duties.” Ex. 1005, 1:10-15.

49. Meisel discloses an **acoustic throat microphone (20)**, the same as the ’058 patent, which detects vocal cord vibrations associated with voicing activity (e.g., voiced speech) through sounds emanating from the speaker’s throat

(physiological information). Ex. 1005, 2:31-35; In particular, Meisel explains that “[t]he throat microphone, if used, detects the vibration of the vocal cords, or ‘voicing,’” and “the [throat] microphone 20 is disposed to detect vibration of the vocal cords through sounds emanating from the throat of a speaker.” Ex. 1005, 2:31-35, 5:29-36. A POSITA would have recognized that detecting “vibration of the vocal cords” is receiving physiological information about “voicing” activity of the speaker. Ex. 1005, 2:33-35.



Ex. 1005, FIG. 13 (annotated).

d. [1c] “at least one processor coupled among the microphones and the voicing sensor,”

50. Meisel discloses at least one processor coupled among the microphones and the voicing sensor. For example, Meisel discloses the same system architecture as the '058 patent for coupling a processor among the microphones and the voicing sensor. Figure 1 of the '058 patent shows microphones (10) and voicing sensors (20) connected to the same processor (30).

Ex. 1001, FIG. 1.

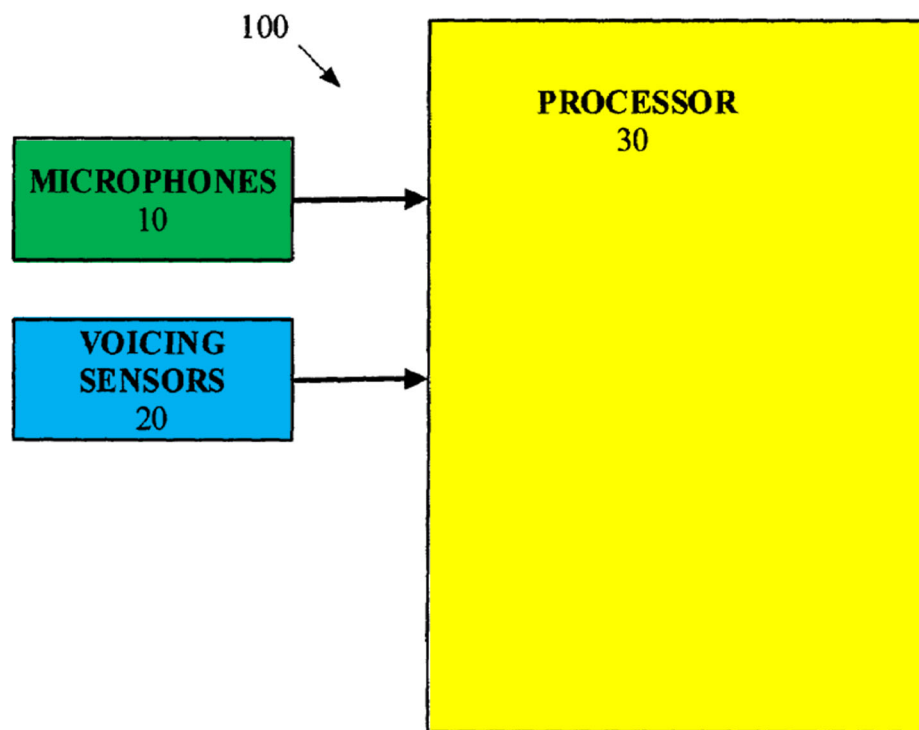
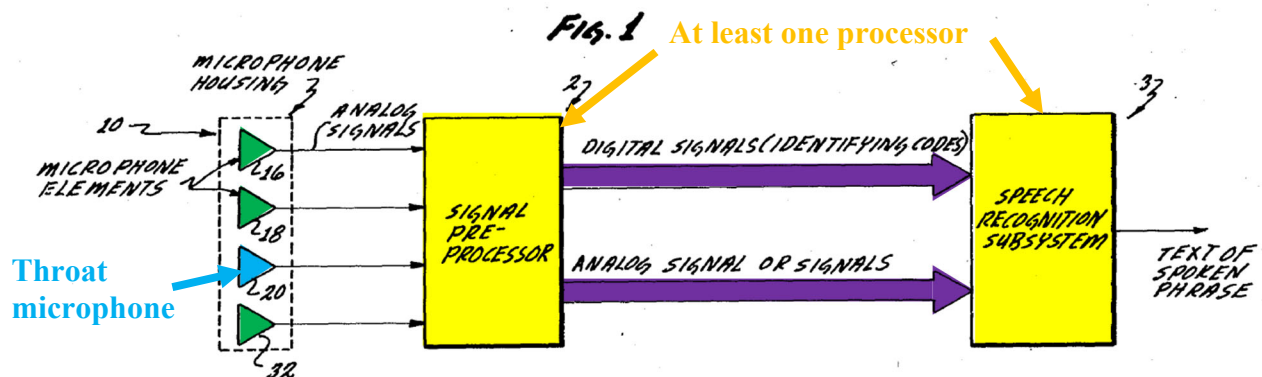


Figure 1

Ex. 1001, FIG. 1 (modified, annotated).

51. Meisel's Figure 1 similarly shows that all microphones, including the **nose microphone (16)** and **two mouth microphones (18, 32)**, and the **voicing sensor (throat microphone (20))** are all connected, or coupled, to the same **signal pre-processor (2)** and then to the **speech recognition subsystem (3)**. Ex. 1005, FIG. 1, 4:27-34.



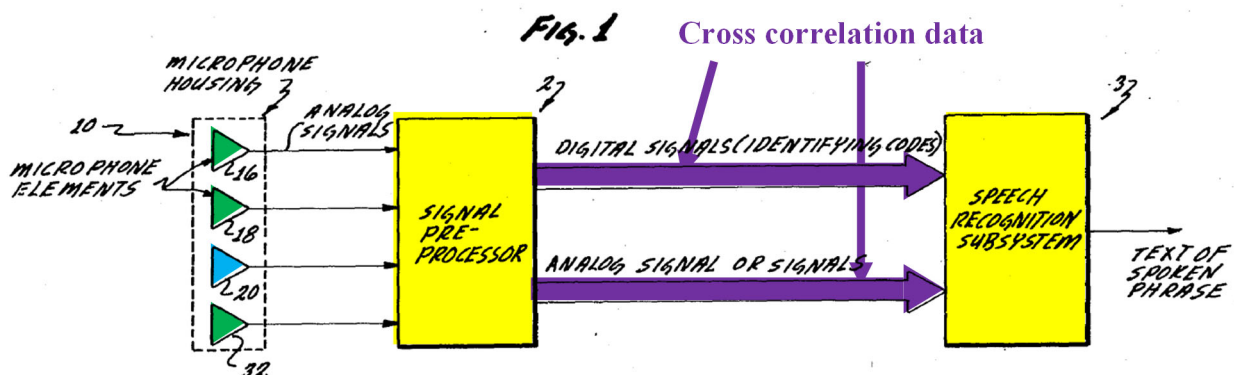
Ex. 1005, FIG. 1 (annotated).

52. Meisel discloses that the signals from the **acoustic microphones (nose microphone (16) and mouth microphones (18, 32))** and the **voicing sensor (20)** are sent to the **signal pre-processor (2)** to detect differences in the signal amplitudes. Ex. 1005, 4:35-42 (“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. These imbalances are indicative of the production of certain speech sounds.”). The differences in the signal amplitudes are coded as Boolean variables and then sent to the speech recognition subsystem (3), which processes the codes to detect and classify speech. Ex. 1005, 4:39-53 (“The

detected imbalances are coded as Boolean variables in a conventional manner to create a digital code corresponding to the pattern of imbalances.”), 3:33-41 (“The selected signals and the codes are passed to the recognition subsystem. The selected signals and the codes may be processed to detect and classify syllables of speech, and the identified syllables may be combined into phrases or sentences.”).

- e. [1d] “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;”

53. Meisel discloses generating cross correlation data between the physiological information and an acoustic signal received at one of the two microphones. The **nose and mouth microphones (16, 18, 32)** receive acoustic signals, as explained in [1a] above (*supra* § IX.A.4.b), and the **throat microphone (20)** receives physiological information, as explained in [1b] above (*supra* § IX.A.4.c). Meisel further discloses that the processor (e.g., **signal pre-processor (2)**) generates **cross correlation data** between the physiological information and an acoustic signal received at one of the two microphones.



Ex. 1005, FIG. 1 (annotated).

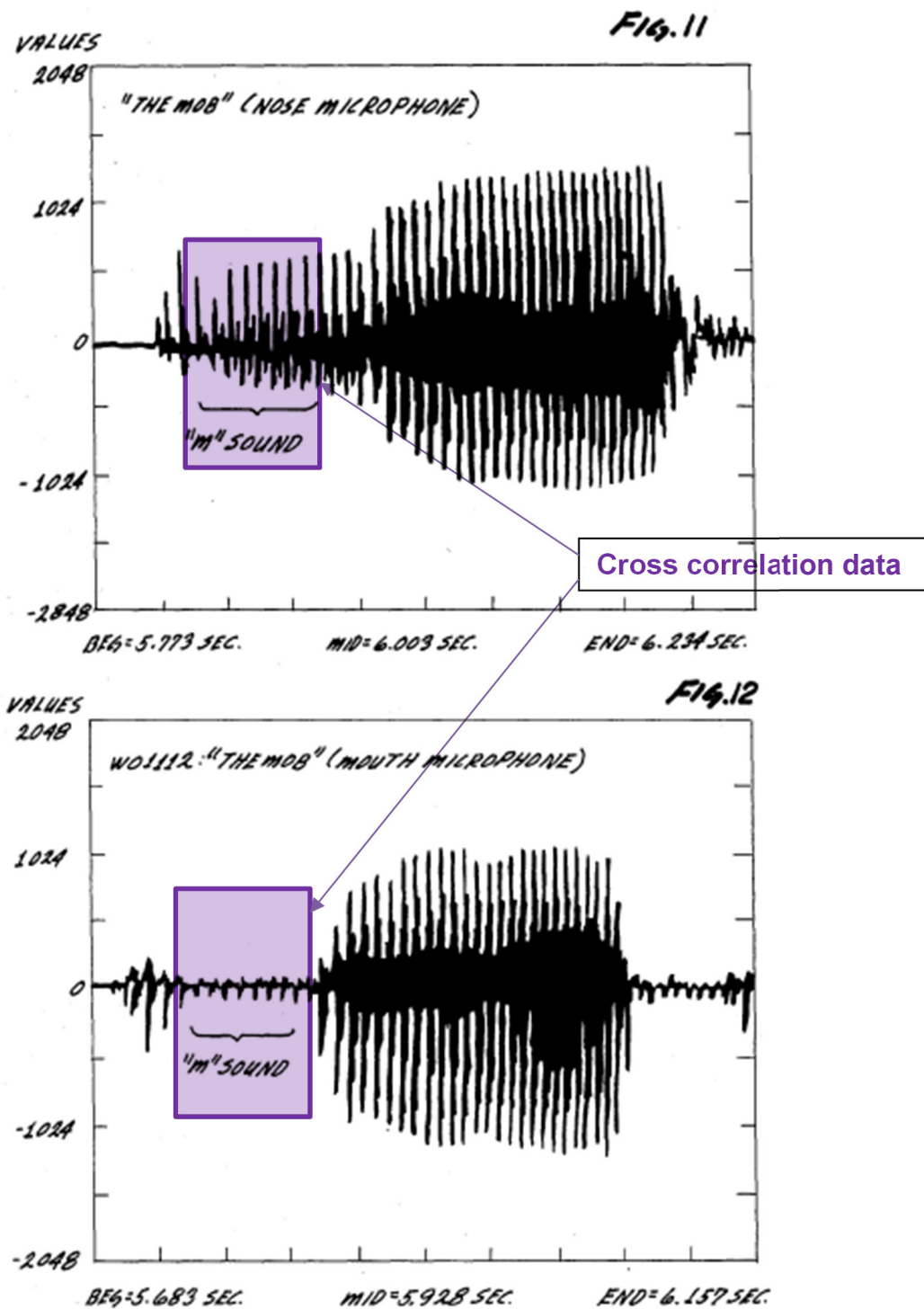
54. Meisel discloses correlating the acoustic information from at least one of the mouth microphones (e.g., **the first mouth microphone (18)** located to the side of the speaker's mouth) with the physiological information from the **throat microphone (32)** by comparing their signals to detect differences in their signal amplitudes. Ex. 1005, 4:35-42.

55. As explained above in [1c] (*supra* § IX.A.4.d), Meisel discloses sending signals from the **nose and mouth microphones (16, 18, 32)** and the **throat microphones (20)** to the **signal pre-processor (2)**, which compares these signals and determines “imbalances” (differences) in the amplitude of signals from different microphones. Ex. 1005, 4:35-42, FIG. 1. These differences in signal amplitudes are then coded as “the pattern of imbalances” and then sent to the **speech recognition subsystem (3)**. Ex. 1005, 4:35-44, FIG. 1. The speech recognition subsystem (3) uses the coded differences in signal amplitudes to complete speech recognition. Ex. 1005, 4:43-53. A POSITA would understand that the pattern of imbalances is a type of **cross correlation data** because it contains, or is based on, a correlation between signal amplitudes of the acoustic microphones (e.g., the **side mouth microphone 18**) and the **throat microphone 20**.

56. Meisel discloses that its speech recognition system compares, for example, “the signal from the throat microphone to that of the mouth microphone”

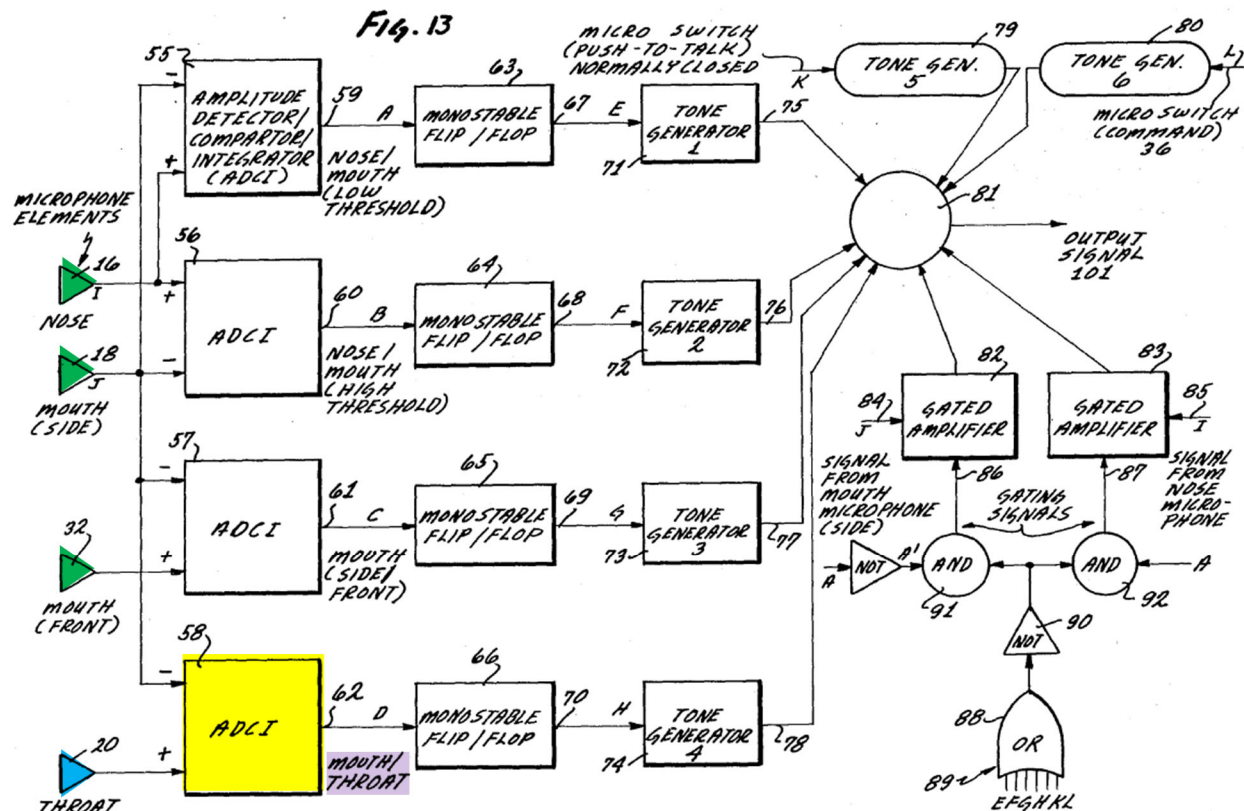
to “distinguish, for example, voiced from unvoiced stop consonants” and “voiced fricatives (v, z, and th) from unvoiced fricatives (f, s, and sh).” Ex. 1005, 2:31-48. As explained above in [1a] (*supra* § IX.A.4.b), the nose microphone (16) and the mouth microphones (18, 32) receive acoustic signals emanating from the speaker’s nose and mouth, respectively. Ex. 1005, 2:11-17.

57. In Figures 11 and 12, Meisel discloses that the signals from the nose microphone (16) and the mouth microphones (18, 32) are correlated with each other to determine whether a portion of the signals is nasal. Ex. 1005, 8:24-59, FIGS. 11-12. While Figures 11 and 12 illustrate cross correlation of the signals from the nose microphone (16) and the first mouth microphone (18), Meisel explains that “[s]imilar proof can be provided of the importance of recording the sounds emanating from the throat or immediately in front of the mouth for certain words in order to provide a reliable recognition of such spoken words.” Ex. 1005, 8:55-59. Indeed, Meisel’s Figure 13 shows an ADCI for comparing different signals, such as ADCI unit (58) for comparing the throat microphone (20) with the side mouth microphone (18). Ex. 1005, FIG. 13, 9:60-66. Therefore, a POSITA would understand that the same cross correlation process illustrated in Figures 11 and 12 applies equally to the signals from the throat microphone (20) and, for example, the side mouth microphone (18).



Ex. 1005, FIGS. 11-12 (annotated).

58. In Figure 13, Meisel explains that the first mouth microphone (18) and the throat microphone (20) are both connected to ADCI unit (58), which is part of the signal pre-processor (2). Ex. 1005, 9:60-66. The output (62) from the ADCI unit (58) “will become logically true when a voiced stopped sound is being pronounced.” Ex. 1005, 10:9-11. In other words, if the output of the ADCI unit (58) (the comparison of a throat signal and side mouth signal) is non-zero, then the signal pre-processor (2) determines that a voiced speech is present. Ex. 1005, 9:12-22; A POSITA would understand that the throat microphone (20) generally indicates whether the corresponding speech types (e.g., fricative, plosive, stopped) produced at any of the other microphones (16, 18, 32) are voiced or unvoiced.



Ex. 1005, FIG. 13 (annotated).

59. Moreover, generating the claimed cross correlation data would have been obvious to a POSITA by cross-correlating Meisel's throat microphone (20) and sound received at one or more of the acoustic microphones (16, 18, 32) to identify voiced speech, including because cross correlation was a well-known and routine process in the art, as taught by Freidin. A POSITA would have found this cross correlation obvious to identify sounds at the throat and mouth microphones as their respective voiced or unvoiced type based on whether the throat microphone has a high signal amplitude, rather than rejecting them as noise. Such

signal comparison would have been understood by a POSITA to allow for more accurate characterization of Meisel's speech.

60. Freidin discloses a correlator that correlates an input signal sequence, d_i , with a reference signal sequence, r_i , to calculate a correlation value, y , using the formula $y=d_i*r_i$ for the sequence $i=(1,n)$. Ex. 1007, 1:10-26, 1:37-52. A POSITA would have recognized that Freidin's correlation formula is a cross correlation formula. *see, e.g.*, Ex. 1016, 20 ("Given two real-valued sequences $x(n)$ and $y(n)$ of finite energy, the *crosscorrelation* of $x(n)$ and $y(n)$ is a sequence $r_{xy}(l)$ defined as

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

."). A POSITA would have found it obvious to

apply Freidin's correlator to the input signals from Meisel's side mouth microphone (18) and throat microphone (20) to calculate a correlation value (cross correlation data) between the two input signals to identify voiced speech when the signals are both present at both microphones (e.g., the difference would be small because a signal is present both at the throat and mouth microphones).

a. A POSITA would have been motivated to combine Meisel and Freidin

61. In my opinion, a POSITA would have been motivated to combine Meisel's speech processing system with Freidin's correlator to calculate cross correlation data of the signals received at Meisel's microphones. Like Meisel (*infra*

§ IX.A.4.g), Freidin discloses comparing two received signals (a reference signal sequence and an input signal sequence) to calculate cross correlation data and determine whether the signals are correlated. Ex. 1007, 1:43-46. A POSITA would know that determining whether two signals are correlated was a well-understood concept at the '058 patent's earliest priority date. Computing the cross correlation, summing the squared or absolute differences, or finding correlation coefficients are fundamental approaches taught in undergraduate classes to POSITAs and are all part of the same family of techniques. Meisel determines correlation between two signals by integrating (summing) the differences between two signals and comparing the result against a threshold; however, a POSITA would readily understand that the signal correlation would be found using any of the techniques in the aforementioned family, including the cross correlation function, and then subsequently applying a threshold comparison.

62. Freidin describes the cross correlation function to produce the cross correlation data, y , to identify a particular result, such as whether Meisel's speech is voiced speech. Ex. 1007, 1:48-52. A POSITA would have looked to Freidin to implement the cross correlation and speech identification performed by Meisel's ADCI units and signal pre-processor (2) and speech recognition subsystem (3) because Meisel describes a system in which the signals are compared and correlated with each other and produce a result, but Freidin provides a well-known

cross correlation framework to render Meisel's identification more objective and accurate. A POSITA also would have been motivated to use cross correlation, as disclosed in Freidin, to 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. Thus, a POSITA would have been motivated to use Freidin's cross correlation calculation to improve the accuracy of voiced and unvoiced speech identification already performed by Meisel. *see, e.g.*, Ex. 1005, 2:31-48.

63. 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." Ex. 1007, 1:7-26. 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.

a. A POSITA would have reasonably expected to succeed in modifying Meisel with Freidin

64. In my opinion, a POSITA would have reasonably expected to succeed in combining Meisel's speech recognition system and Freidin's correlator for

correlating received speech signals because the combination uses only known techniques of standard signal processing functions to achieve predictable results. A POSITA would have recognized, and it would have been well within a POSITA's skill, to implement Meisel's processing in digital processors; in fact, it would have been most straightforward to process the received signals using digital processors. It was well-known at the time to cross-correlate received signals to determine signal correlation. Ex. 1013 (Rainton), Abstract ("a cross correlator calculates a cross correlation value by using a predetermined cross correlation function based on the filtered first and second signals"); Ex. 1015 (Matsuo), 7:63-8:26 (discussing using a cross correlation and cross correlation coefficient (R) to determine the correlation between two signals); Ex. 1014 (Chu), Abstract, 2:6-19; Ex. 1016 (Ingle), 20, 29; Ex. 1017 (Byers), 7:40-48, 12:33-56; Ex. 1018 (Makino), Abstract, 7:2-9; Ex. 1015 (Matsuo), 3:10-15; Ex. 1020 (Matsuo 2), 17:1-16; Ex. 1021 (Benesty), Abstract, 4:4-13, 6:16-9:35. Indeed, cross correlation calculations on two received signals was so conventional that it was a built-in function for signal

processing programs, such as the XCORR function of Matlab.¹ *See, e.g.*, Ex. 1009, 6-367 through 6-372.

65. Indeed, cross correlation calculations date back decades to at least 1966 with a book stating that “[t]he optimal detection scheme for the signal specified exactly is therefore cross correlation between the expected signal and the received waveform.” Ex. 1024 (Green), 163; *see also* Ex. 1023 (Duda), 279-80 (1973 reference explaining “template matching” using cross correlation to determine similarity in a template and a region of interest in an image.). In fact, cross correlation calculations were standard calculations for undergraduate signal processing classes at the earliest claimed priority date of the ’058 patent. Thus, applying Freidin’s cross correlation function to compare Meisel’s acoustic and throat microphone signals (or any other received microphone signals) to correlate the signals and identify, for example, voiced speech, was nothing more than the application of conventional, decades-old processing techniques applied in their ordinary and expected manner to achieve the expected result. A POSITA would have found that implementing Freidin’s cross correlation in Meisel’s speech

¹ Cross correlation functions were so conventional that the ’058 patent identifies the built-in conventional XCORR function rather than explaining cross correlation. Ex. 1001, 4:64-5:3.

recognition system was combining prior art elements according to known methods (digital signal processing of speech) to yield predictable results (improving the accuracy of speech detection). Further, combining Meisel and Freidin would have been applying a known technique (a correlator) to a known device (a speech recognition system) ready for improvement to yield predictable results (improving the accuracy of speech detection).

66. The '058 patent discloses using “cross-correlation (XCORR) between acoustic signal from microphone 1 and the sensor data,” and that the thresholds are empirically determined. Ex. 1001, 4:66-5:3, 5:42-44. “XCORR” was a well-known signal abbreviation and is a common function name for cross correlation functions in signal processing software, such as MATLAB, at least as early as 1999. Ex. 1009, 3-2 through 3-3. For example, the MATLAB Manual for The MathWorks’ Signal Processing Toolbox explains that “[t]he functions xcorr and xcov estimate the cross-correlation and cross-variance sequences for random processes.” Ex. 1009, 3-2, 3-4. It also provides that “the true cross-correlation sequence is a statistical quantity defined as $\gamma_{xy}(m) = E\{x_n y_{n+m}^*\}$ where x_n and y_n are stationary random processes, $-\infty < n < \infty$, and $E\{\}$ is the expected value operator.” Ex. 1009 at 3-2, 3-4. A POSITA would understand that speech signals are modeled as a series of stationary random processes over a period such as about 20ms. Ex. 1025 (Deller), 231. A POSITA would have known how to apply cross

correlation equations because they are conventional functions, such as one built into MATLAB, using the signals from the acoustic microphones (e.g., first mouth microphone (18) and the throat microphone (20)) as the correlation inputs to arrive at the cross correlation between the two microphones. For example, as explained above, Figure 13 in Meisel shows ADCI unit (58), which compares the signals received at the throat microphone (20) with the signals received at the side mouth microphone (18) for correlation. Ex. 1005, FIG. 13, 9:60-66. Based on this disclosure, a POSITA would have been motivated to look at the MATLAB Manual because the Signal Processing Toolbox was a commonly used program in the art for speech signal processing at the time. Ex. 1016, 29. Given that both Meisel and Kaneda use signal processing in their speech recognition systems, a POSITA would have had a reasonable expectation of success in implementing the cross correlation (e.g., xcorr) function as a conventional signal processing operation to determine cross correlation data between Meisel's throat microphone signal and one or more of the mouth or nose microphone signals.

- f. **[1e] “identifies 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;”**

67. Meisel discloses that its processor (**signal pre-processor (2)** and **speech recognition subsystem (3)**) identifies 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.

68. As explained in [1d] (*supra* § IX.A.4.e), Meisel discloses **a signal pre-processor (2)** that receives acoustic signals from the **nose microphone (16)** and the **mouth microphones (side (18) and/or front (32))** (Ex. 1005, 2:11-17), and physiological information from the **throat microphone (20)**. Ex. 1005, 2:31-53. The received signals are then compared with each other to distinguish speech as, for example, voiced and unvoiced speech. Ex. 1005, 2:31-53. A POSITA would understand that Meisel describes this comparison process using cross correlation data between the acoustic information (e.g., side mouth microphone (18)) and the physiological information (e.g., throat microphone (20)). Meisel further discloses that the cross correlation data is sent to a data buffer (302), and then to a feature extractor (303) to examine a particular time period, such as sixty (60) milliseconds, following the first rise in the amplitude signal from the microphone (a portion of the acoustic signal received at the one receiver) “to determine if any of the logical signals 109, 110, 111 or 112 is true.” Ex. 1005, 13:5-13. If the feature extractor (303) recognizes “a voiced stop at the start of the word as indicated by a true signal on the line 112, a line 308 is set true.” Ex. 1005, 13:16-18. A POSITA would understand that classifying speech as a voiced stop involves determining whether the portion of the acoustic signal analyzed exceeds a

correlation threshold. For example, as explained in § IX.A.4.e ([1d]), Meisel's signal pre-processor (2) compares the "m" sound portion of an acoustic signal from the nose microphone (16) with the "m" sound portion of an acoustic signal from the side mouth microphone (18) to detect a nasal speech. Ex. 1005, 8:24-59, FIGS. 11-12. A POSITA would have understood from Meisel's disclosure that the signals from the throat microphone (20) would be used to further classify the speech as voiced or unvoiced. Ex. 1005, 2:31-48.

69. Moreover, identifying the acoustic signal 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** would have been obvious in view of Freidin. Freidin discloses a correlator that correlates two received signals, an input data sequence and a reference sequence, to calculate a **correlation value y**. Ex. 1007, 1:43-46. A POSITA would recognize that correlation value y is cross correlation data because it is calculated "by first multiplying each data sample d_i by a corresponding reference value r_i ," which is the same as the cross correlation equation that cross-correlates the input data samples to the reference values. Ex. 1007, 2:37-40. This **correlation value y** results in cross correlation data because it creates a correlation value for each data sample i. Ex. 1007, 1:14-22.

70. Freidin further discloses that the "**correlation value y** [cross correlation data] is compared against a **correlation threshold y_{TH}** , and whenever y

exceeds the **correlation threshold y_{TH}** , a correlation match is declared indicating that the correlation sequence has been received” Ex. 1007, 1:48-52.

71. In my opinion, a POSITA would have known that Freidin’s cross correlation calculation and comparison to a threshold value are applicable to Meisel’s speech recognition because Freidin discloses that “[c]orrelation is used in a number of applications including speech recognition.” Ex. 1007, 1:23-26. Freidin expressly states that the correlator operation “is not to be restricted except within the spirit and scope of the appended claims” because many variations of the disclosed embodiments “are possible which remain within the concept and scope of the invention,” which “would become clear to one of ordinary skill in the art after inspection of the specification, drawings and claims.” Ex. 1007, 11:51-12:4. A POSITA would have found it obvious to apply Freidin’s cross correlator and threshold comparison to Meisel’s comparison of, for example, the side mouth microphone (18) and physiological throat microphone (20) because Meisel teaches correlating the received signals and using the result of the correlation to recognize speech. Ex. 1005, Abstract. A POSITA would have found it obvious to use cross correlation because it was a conventional technique for comparing signals, which technique was a standard built-in function for signal processing software, such as MATLAB. A POSITA also would have found it obvious to compare the cross correlation data to a correlation threshold to provide an objective determination of

the data to identify the desired speech type, such as voiced speech. Ex. 1005, 2:31-53.

a. A POSITA would have been motivated to combine Meisel and Freidin

72. 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. While Meisel and Freidin find a correlation between signals, Freidin describes the cross correlation function to produce the cross correlation data y and explains how to use that cross correlation data to objectively compare it to a threshold y_{TH} to identify a particular result, such as Meisel's voiced speech. Ex. 1007, 1:48-52.

73. A POSITA would have looked to Freidin to implement the cross correlation and speech identification performed by Meisel's ADCI and signal pre-processor (2) because Meisel describes a system in which the signals are compared and correlated with each other and produce a result, and Freidin provides a well-known cross correlation and threshold comparison framework to render Meisel's identification more objective. For example, a POSITA would have been motivated to use the cross correlation function because it identifies voiced speech when there is a speech signal at both microphones (e.g., throat microphone and

mouth microphone), and thus the difference (calculated by Meisel's ADCI) may be small. Thus, because cross correlation allows for calculating the presence of voiced speech when the signal is present at both microphones (e.g., small difference, but high multiplication of signal values), it provides additional accuracy to Meisel's speech type determinations and expands to correlating other voiced speech types than just voiced stops.

74. A POSITA would have been motivated to use Freidin's threshold comparison 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 speech recognition. Ex. 1007, 1:7-26.

a. A POSITA would have reasonably expected to succeed in modifying Meisel with Freidin

75. A POSITA would have been motivated to combine Meisel and Freidin for the reasons discussed in § IX.A.4.e above. It was also well known in the art to compare the cross correlation value to a predetermined threshold value to determine whether the signals sufficiently correlate with each other. Ex. 1007, 1:48-52; *e.g.*, Ex. 1014 (Chu), Abstract ("The signal is declared to be stationary in the current time interval when the cross correlation exceeds a threshold value, and is declared to be non-stationary in the current time interval when the

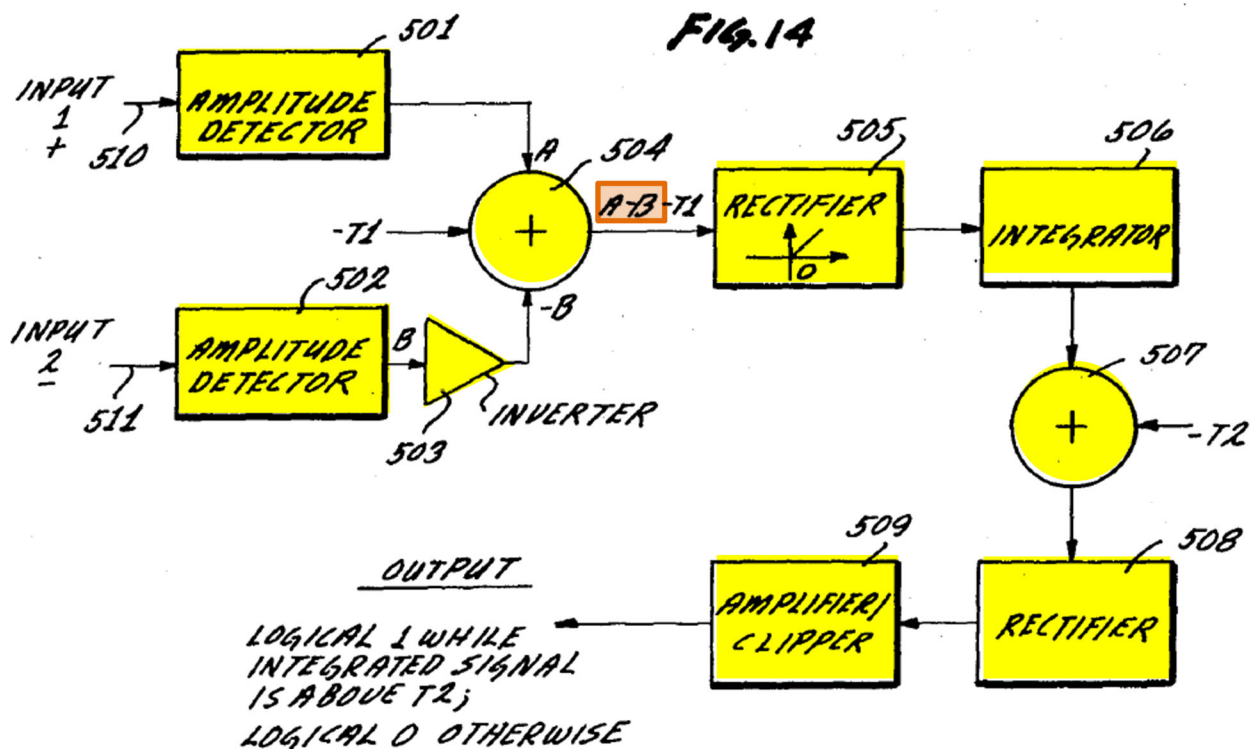
cross-correlation is less than the threshold value.”). Accordingly, a POSITA would have reasonably expected to succeed in using Freidin’s cross correlation threshold comparison in Meisel’s speech recognition system to determine whether the signals represent a particular speech type, such as voiced speech, because the use of these techniques is nothing more than the application of conventional decades-old functions applied in their ordinary and expected manner. Therefore, a POSITA would have found that implementing Freidin’s cross correlation threshold comparison in Meisel’s speech recognition system was combining prior art elements according to known methods (digital signal processing) to yield predictable results (improving the accuracy of speech detection). Further, combining Meisel and Freidin would have been applying a known technique (a correlator employing a correlation threshold) to a known device (a speech recognition system) ready for improvement to yield predictable results (improving the accuracy of speech detection).

- g. [1f] “generates difference parameters between the acoustic signals received at each of the two receivers, wherein the difference parameters are representative of the relative difference in signal gain between portions of the received acoustic signals;”**

76. Meisel discloses that the processor (e.g., **signal pre-processor (2)**) generates difference parameters between the acoustic signals received at each of the two receivers (e.g., microphones 16, 18), wherein the difference parameters are

representative of the relative difference in signal gain between portions of the received acoustic signals.

77. As explained in [1c] above (*supra* § IX.A.4.d), different microphones in Meisel (e.g., **nose microphone (16)** and **mouth microphone (18)**) generate different signal amplitudes for the received signals, and the **signal pre-processor (2)** compares those signals to determine “instant imbalances in the relative amplitude of each microphone.” Ex. 1005, 4:35-39. Meisel’s “imbalances” refer to differences in the signal amplitudes received at different microphones, which a POSITA would recognize as the signal gain. Meisel explains that “[t]hese imbalances are indicative of the production of certain speech sounds” received at the microphones. Ex. 1005, 4:35-39. A POSITA would understand that the comparison of the signal amplitudes of the two microphones to detect instant imbalances, which a POSITA would understand determines or generates the difference (difference parameters) between the acoustic signals received at each of the two microphone receivers. As shown below in Figure 14, the input signals (A, B) from two microphones are compared using the summing unit (504), which subtracts the amplitude of input signal B from the amplitude of input signal A, as shown by the “A-B” indicator below. Ex. 1005, 9:12-22.

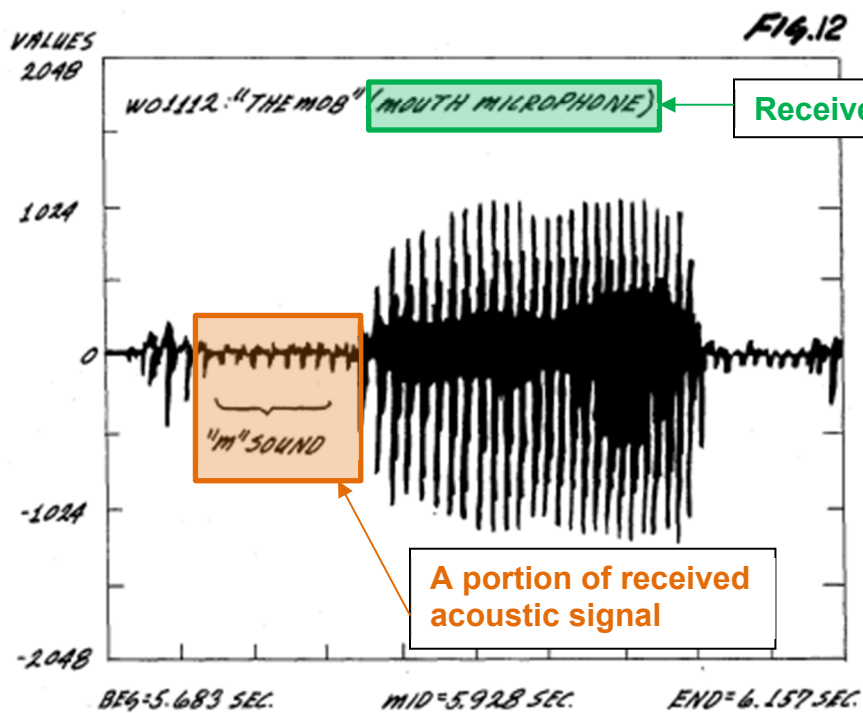
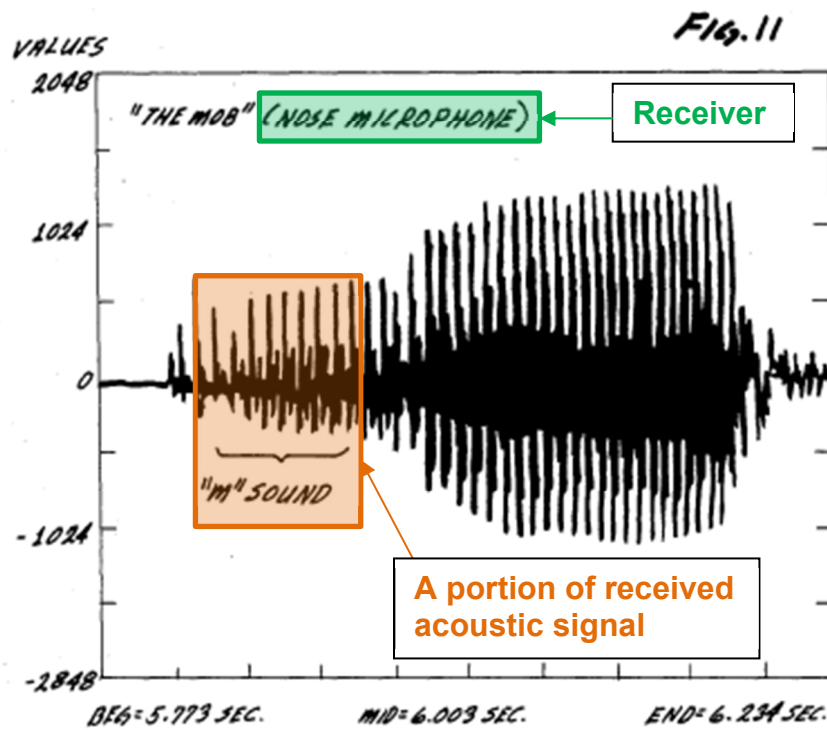


Ex. 1005, FIG. 14 (annotated).

78. As explained in § IX.A.4.g ([1g]), *infra*, after the summing unit (504) calculates the difference between the amplitudes of input signals A and B, it compares the difference in the input signal amplitudes to a predetermined threshold T_1 . Ex. 1005, 9:12-22. If input signal A is greater than input signal B by at least T_1 , then the signals proceed to the next processing step to identify a speech type. Ex. 1005, 9:23-37. The integrator (506) processes the signals, and its output of the integrated amount (which measures by how much input signal amplitude A exceeds the input signal amplitude B) is fed to the comparator circuit (507) to compare the difference parameter (difference between input signal A and input

signal B) to a second predetermined threshold T2. Ex. 1005, 9:38-42. The output of the comparator circuit (507) determines the type of speech, for example, the occurrence of a nasal speech. Ex. 1005, 9:47-50.

79. Meisel further discloses that the difference parameters are representative of the **relative difference in signal gain between portions of the received acoustic signals**. For example, Meisel explains that Figures 11 and 12 show the amplitude of acoustic signals from the nose microphone (16) and the first mouth microphone (18), respectively. Ex. 1005, 4:3-8. The values of the signals in Figures 11 and 12 represent the amplitude patterns recorded by the nose microphone (16) and the first mouth microphone (18) and are compared to determine whether the speech is nasal. Ex. 1005, 8:24-59. A POSITA would understand that amplitude signals are representative of the signal gain for each microphone, and the difference in gain is used to determine, for example, voiced speech.



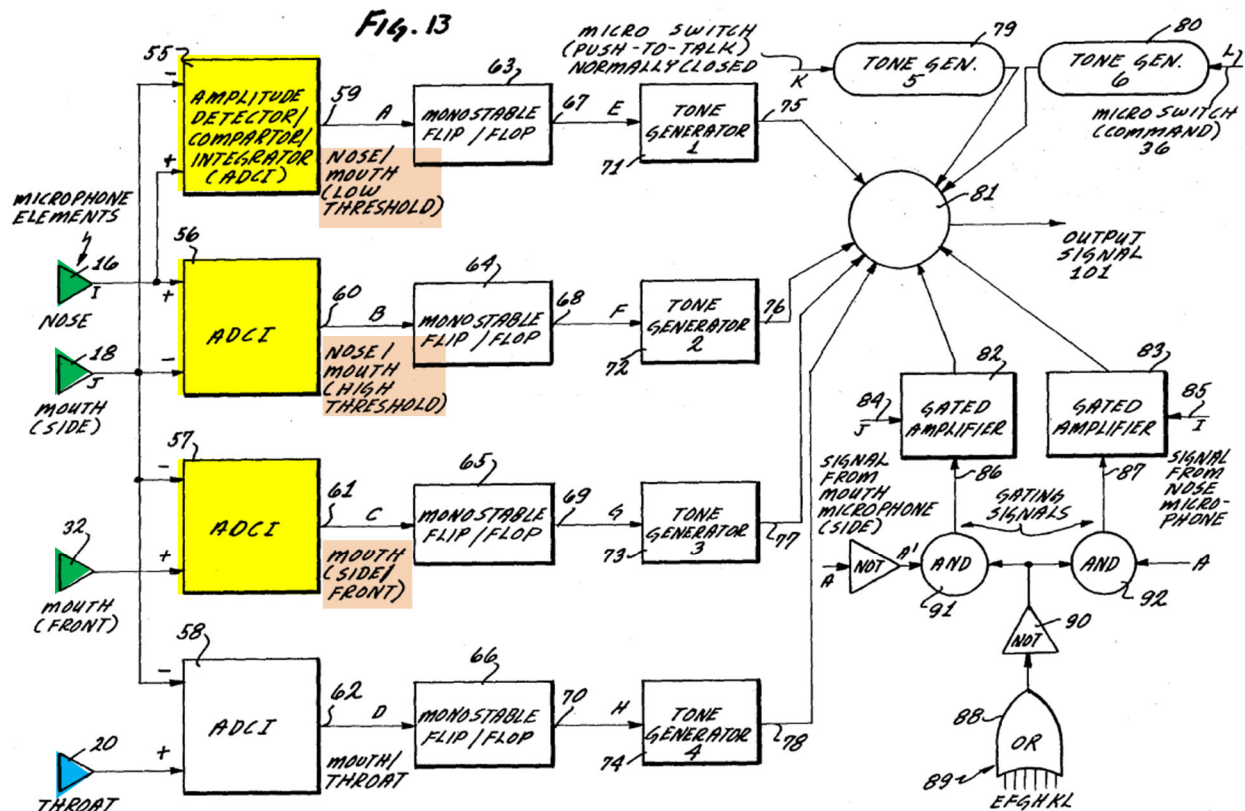
Ex. 1005, FIGS. 11-12 (annotated).

80. As one example in Meisel, the processor (**signal pre-processor (2)** and **speech recognition subsystem (3)**) recognizes the “M” sound in “The Mob” as a nasal speech because it recognizes that the signals received at the nose microphone (16) exhibit “a relatively large amplitude” during a period when the signals received at the first mouth microphone (18) exhibit “a relatively low amplitude.” Ex. 1005, 8:24-59; *see also* Ex. 1005, 2:25-30, 5:51-55, 10:1-3. A POSITA would understand that comparing a “large amplitude” to a “low amplitude” via subtraction is determining a difference in the signals received at the two microphones and constitutes generating the difference parameters between the nose microphone (16) and the first mouth microphone (18). In other words, the relative difference in signal gain between the nose microphone (16) and the first mouth microphone (18) in the “M” portion of the acoustic signals allows the processor to determine the type of speech (e.g., nasal). A POSITA would understand that the signal amplitude comparison in Figures 11 and 12 would be applied to the speech signals when a speaker’s mouth is closed, partially open, or open, each of which imparts different gains to the speech. Ex. 1005, 9:66-10:7.

81. As explained further in [1g] below (*infra* § IX.A.4.h), the summing unit (504), which acts as a comparator, generates difference parameters by inverting the amplitude of the second microphone signal (e.g., the nose microphone (16)), and summing the inverted signal with the amplitude of the first

acoustic signal. Ex. 1005, 9:12-22. Stated differently, the second acoustic signal amplitude is subtracted from the first acoustic signal amplitude to generate the difference parameter between the two microphones. Ex. 1005, 9:12-22.

82. A POSITA would understand that the same process would be applied to other microphones, for example, between the first mouth microphone (18) located at the corner of the speaker's mouth and the second mouth microphone (32) located in front of the speaker's mouth. In fact, Meisel expressly states that "[s]imilar proof can be provided of the importance of recording the sounds emanating from the throat or immediately in front of the mouth for certain words in order to provide a reliable recognition of such spoken words." Ex. 1005, 8:55-59. Based on this disclosure, a POSITA would understand that the acoustic signals from the first mouth microphone (18) can be compared with those of the second mouth microphone (32) to determine a difference in their signal gains and generate difference parameters between the two mouth microphones.



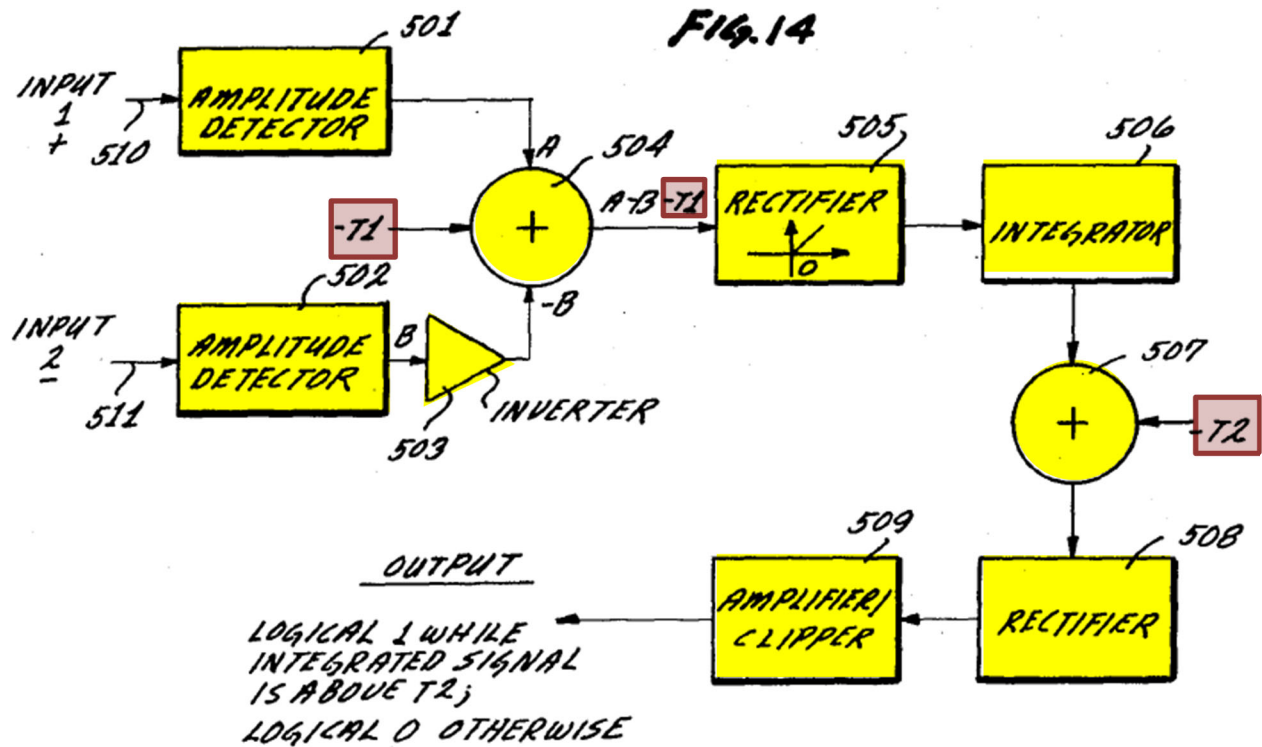
Ex. 1005, FIG. 13 (annotated).

- h. [1g] “identifies information of the acoustic signals as unvoiced speech when the difference parameters exceed a gain threshold; and”

83. Meisel discloses that the processor (**signal pre-processor (2)**) identifies information of the acoustic signals as unvoiced speech when the difference parameters exceed a **gain threshold**.

84. Meisel discloses that the acoustic signals received at a pair of microphones in the microphone assembly (comprising the nose microphone (16) and two mouth microphones (18, 32)) are input into the ADCI units. The ADCI

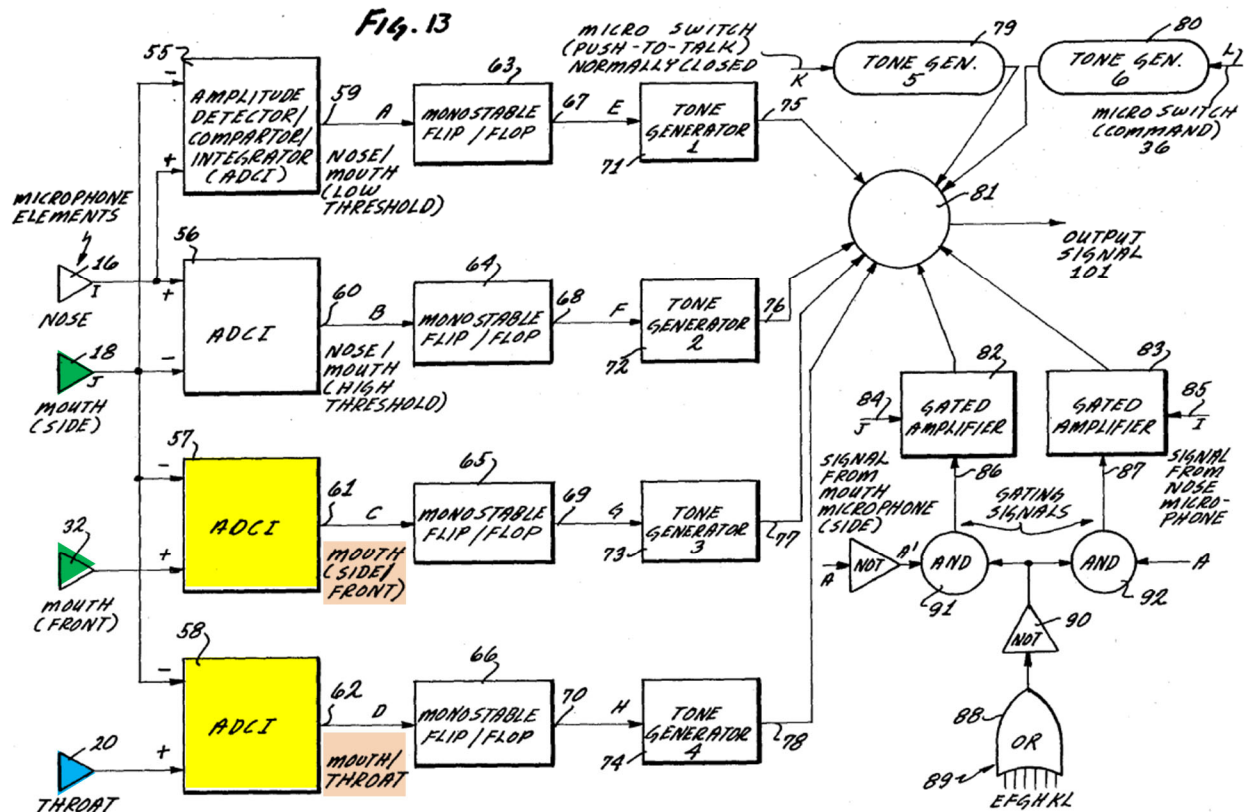
units are part of the signal pre-processor (2) and include two gain thresholds, T1 and T2, to which the differences are compared. Ex. 1005, 8:60-9:2, 9:12-42.



Ex. 1005, FIG. 14 (annotated).

85. As explained in [1e] above, Meisel discloses the feature extractor (303), which distinguishes voiced and unvoiced stop consonants. *Supra* § IX.A.4.f. Meisel explains that this is accomplished when each of the amplitude detectors (501, 502) receives inputs (acoustic signals, the first input and the second input) from a pair of microphones. Ex. 1005, 8:63-9:11 The amplitude detectors (510, 511) determine the amplitudes of the input signals. Ex. 1005, 9:4-11. The inverter (503) inverts the signal amplitude (Input 2) from the second microphone (e.g., nose microphone (16) or second mouth microphone (32)), such that it becomes a

negative value (-B). Ex. 1005, 9:12-14. The summing unit (504) adds the signal amplitude of the first input and inverted signal amplitude of the second input, and therefore generates difference parameters between the two signals. Ex. 1005, 9:14-20.



Ex. 1005, FIG. 13 (annotated).

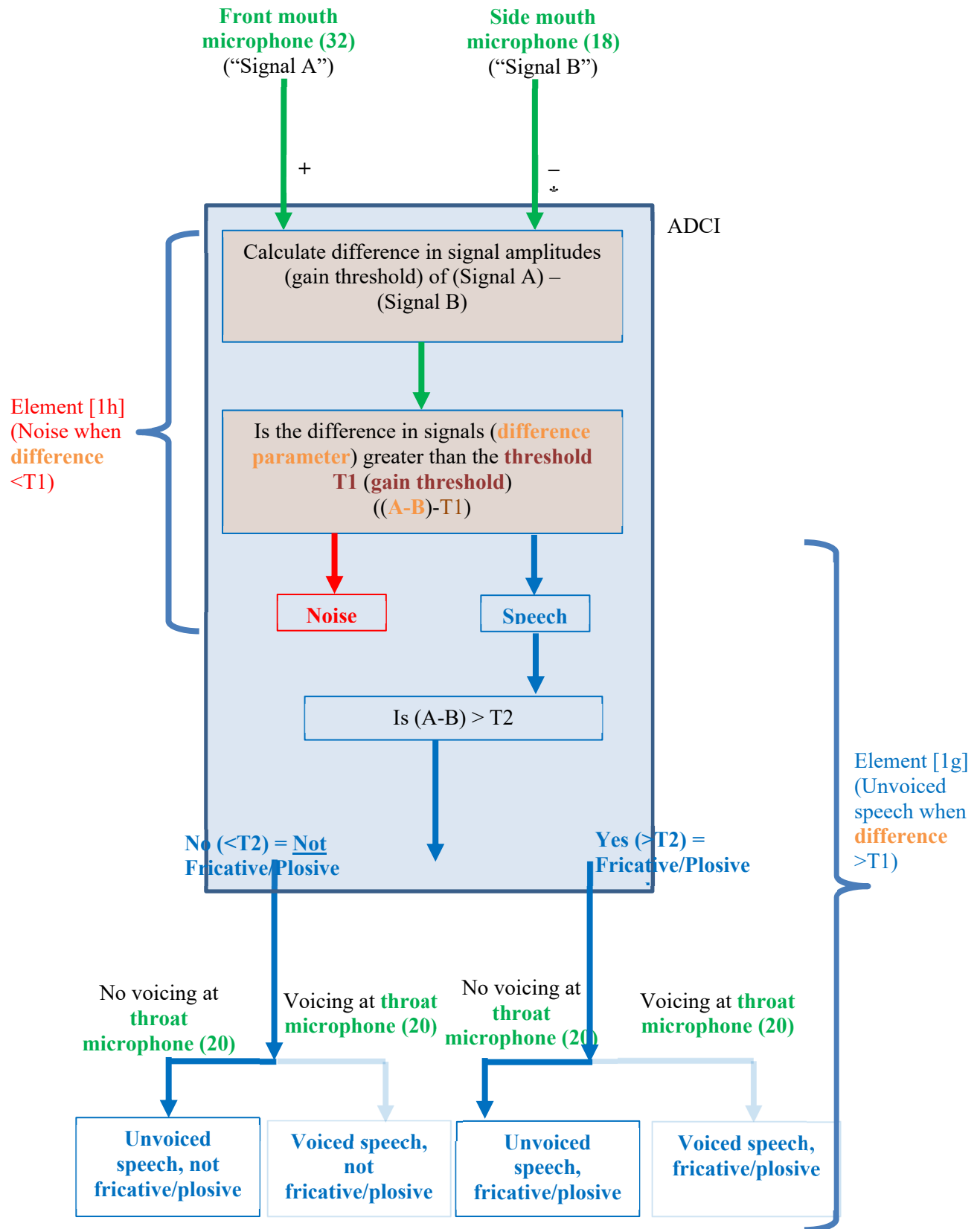
86. Meisel discloses that this signal comparison step would be applied to the side mouth microphone (18) and the front mouth microphone (32) to determine unvoiced speech and noise. Ex. 1005, 8:60-9:50. The signal inputs from both microphones are fed to the summing unit (504) of ADCI unit (57), which compares the amplitudes of the signals and determines whether the difference between the

two signals is above a predetermined threshold T1 (a **gain threshold**). Ex. 1005, 9:12-37. The output of the summing unit (504) is fed to the rectifier (505), which ensures “that there is no output from the rectifier 505 unless the signal at the input is positive.” Ex. 1005, 9:23-25. A positive signal input to the rectifier (505) happens only when the first input (e.g., front mouth microphone (32)) is greater than the second input (e.g., side mouth microphone (18)), and the difference between the first input and the second input (A-B) exceeds the predetermined threshold (T1). Ex. 1005, 9:23-37. In view of Figure 13, a POSITA would understand that the output (61) would also be correlated or compared with the signals from the throat microphone (20) to distinguish the speech as voiced and unvoiced because the throat microphone (20) indicates when speech is voiced or not. Ex. 1005, 2:31-48.

87. If the difference between the signal inputs from the front mouth microphone (32) and the side mouth microphone (18) is greater than the **gain threshold T1**, Meisel’s system proceeds to compare the difference (A-B) to a second threshold to determine the type of speech, for example, whether or not the speech is a fricative or a plosive. Ex. 1005, 9:38-50, 10:7-9. T1 in Meisel is a gain threshold because it is a threshold for acoustic signal amplitudes designed to distinguish voiced and unvoiced speech. A POSITA would have recognized that Meisel’s disclosure of the throat microphone (20) would also increase the accuracy

of speech recognition when combined or compared with various speech types (e.g., fricative/plosive identification) to distinguish voiced and unvoiced speech.

Ex. 1005, 10:9-11. A POSITA would have found it obvious to combine the output (61) of ADCI unit (57) with the cross correlation data explained in § IX.A.4.e ([1d]) above to distinguish voiced speech from unvoiced speech. A POSITA would also understand that a positive (non-zero) output of the summing unit (504) in ADCI unit (57) denotes the difference parameters exceeding a **gain threshold (T1)**. This process is summarized below in Flow Chart 1.



88. A POSITA would recognize that Meisel performs generating the difference parameters and comparing the difference parameters to a gain threshold in two steps. Meisel discloses that signal amplitude A is summed with the inverted signal amplitude B and compared to the gain threshold T1 (Ex. 1005, 9:14-20), which Figure 14 describes as “A-B-T1” (Ex. 1005, FIG. 14). A POSITA would understand that, based on the simple order of operation, A-B must precede -T1 as a separate step, and these steps occur separately. Such multiple subtractions were conventionally resolved through multiple steps, rather than a single double subtraction. Furthermore, a POSITA would recognize that it was less expensive and easier to calculate the A-B and -T1 steps separately when using logic circuits and also negligible in processing time when using digital signal processing. Therefore, a POSITA would understand that Meisel discloses generating difference parameters (A-B) as a separate step from comparing the difference parameter to the gain threshold (-T1).

- i. **[1h] “identifies information of the acoustic signals as noise when the difference parameters are less than the gain threshold.”**

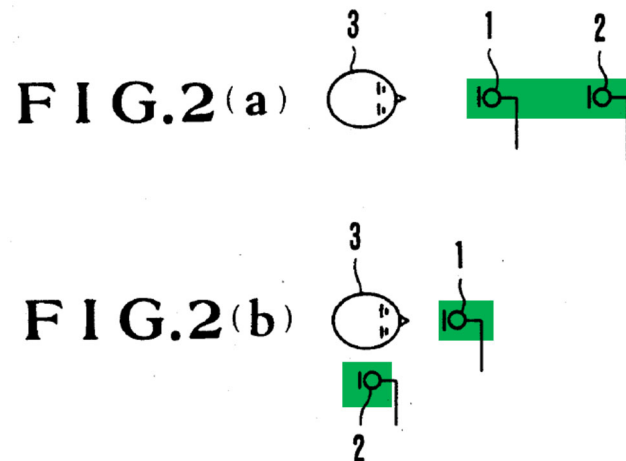
89. Meisel discloses that its speech detection system operates when there is background noise or other types of noise associated with the speaker performing other duties, suggesting that the noise is detected to minimize its effects in accurate speech recognition. Ex. 1005, 1:10-15 (“More particularly, the invention relates to

a speech recognition system usable in the normal work environment of the speaker, particularly under conditions where there is background noise and where the speaker can continue to perform other duties.”). Meisel cites Sakoe, which teaches how “to detect background noise with a view toward minimizing its effect on recognition accuracy.” Ex. 1005, 1:46-50.

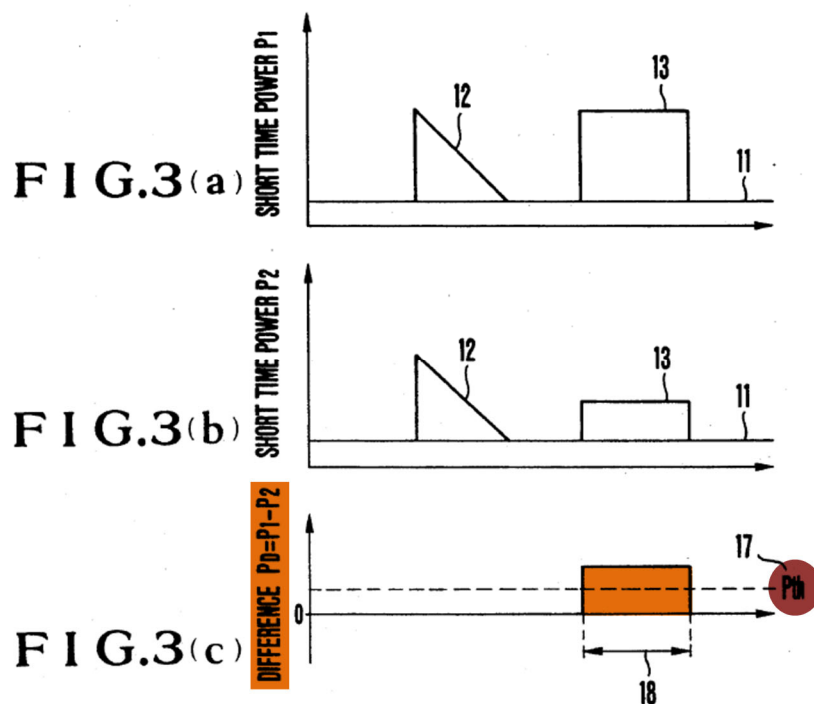
90. Meisel discloses that its comparison process illustrated in Figure 13 above, explained in § IX.A.4.h ([1g]), distinguishes speech from noise. When the difference between the first mouth microphone input and the second mouth microphone input (A-B) is below the threshold T1, then the information of the acoustic signals from the side mouth microphone (18) and the front mouth microphone (32) is identified as noise. *supra* § IX.A.4.h (Flow Chart 1).

91. Moreover, “identif[ying] information of the acoustic signals as noise when the difference parameters are less than the gain threshold” would have been obvious to a POSITA in view of Kaneda. Kaneda discloses a “conventional” speech recognition system comprising **two acoustic microphones**, in which one of the microphones is located near a speaker (microphone 1) and another microphone is located away from the speaker (microphone 2). Ex. 1006, 1:51-57, FIG. 2(a). During a speech period, the speech signals received at microphone 1 would have a higher power output than those received at microphone 2 because of the relative distance of each microphone from the speaker. Ex. 1006, 1:61-63; *compare* the

speech power output (13) in Figure 3(a) *with* the speech power output (13) in Figure 3(b).



Ex. 1006, FIGS. 2(a)-2(b) (annotated).



Ex. 1006, FIGS. 3(a)-3(c) (annotated).

92. This “conventional” speech recognition system then calculates a difference between the speech power outputs of the two microphones “PD (=P1-P2)” (**difference parameters**). Ex. 1006, 2:12-15. “When the power difference PD is larger than a given threshold value Pth17, a corresponding time period 18 is detected as a speech period.” Ex. 1006, 2:15-18. The **threshold value (Pth17)** is a gain threshold because it is a speech signal amplitude threshold used to discriminate speech and noise. If the **power difference PD (difference parameters)** is less than **Pth17 (a gain threshold)**, then the conventional speech recognition system identifies the acoustic signals as noise. Ex. 1006, 2:18-22.

93. A POSITA would have found it obvious to combine Meisel’s speech recognition system and Kaneda’s conventional speech detection system to determine whether it indicates speech or noise.

i. A POSITA would have been motivated to combine Meisel and Kaneda

94. A POSITA would have been motivated to combine Meisel’s speech processing system with Kaneda’s conventional speech and noise determination using difference parameters between two acoustic input signals. Ex. 1006, 1:51-2:22. Kaneda, like Meisel, is directed to a speech recognition system using two or more acoustic microphones. *Supra* §§ IX.A.1, 0, IX.A.4.b. Meisel discloses that the acoustic signals from two acoustic microphones would be compared with each

other to determine the difference in their signal amplitudes. Ex. 1005, 4:35-39, 8:24-59, FIGS. 11-12. Kaneda, like Meisel, uses a difference in signal amplitudes, and explains how the difference would be used to distinguish between noise and speech instead of determining speech type. *Compare* Ex. 1005, 9:12-50, *with* Ex. 1006, 2:12-22.

95. While Meisel does not expressly disclose identifying acoustic signals as noise, Meisel discloses that its speech recognition system is “usable in the normal work environment of the speaker, particularly under conditions where there is background noise and where the speaker can continue to perform other duties.” Ex. 1005, 1:10-15. Furthermore, a POSITA would recognize that Kaneda’s teachings are directly implementable in Meisel’s system because both involve finding signal amplitudes, finding the differences in those amplitudes, and comparing the differences to a threshold. These are all functions of Meisel’s ADCI units. Incorporating the results of Kaneda’s conventional speech detection into Meisel’s system would be trivial to a POSITA because Meisel already discloses a set of signals that conveys the detected characteristics of the speech to a processor for speech recognition. Furthermore, Kaneda discloses that correctly identifying speech periods and noise periods in a speech recognition system is important for improving the accuracy and discloses methods for doing so. Ex. 1006:9-21. A POSITA would have been motivated to look to Kaneda’s noise determination to

remove noise from Meisel's acoustic detection to better identify speech signals in Meisel's speech recognition. Therefore, a POSITA would have looked to Kaneda to identify noise periods to mitigate errors in speech recognition.

96. Although Kaneda discloses three conditions that must be present for the conventional speech recognition system to accurately distinguish speech and noise in practical applications, Ex. 1006, 2:23-27, a POSITA would recognize that Meisel meets all three conditions. First, Kaneda discloses that “[a]n S/N ratio difference in two signals must be present,” which Kaneda acknowledges is satisfied in conventional speech recognition systems, such as Meisel's. Ex. 1006, 2:28-29. Indeed, as explained in Sections IX.A.1 (Overview of Meisel) and IX.A.4.b (element [1a]), Meisel's speech recognition system includes at least two acoustic microphones (a nose microphone (16), a first mouth microphone (18), and a second mouth microphone (32)), which provide different acoustic signal strengths that Meisel uses to calculate the power differences in the ADCIs. Ex. 1005, 2:11-30, 8:37-59, FIGS. 11-12. Second, Kaneda discloses that “[n]oise and speech periods of the two signals must be matched with each other as a function of time.” Ex. 1006, 2:30-32. A POSITA would understand that Meisel's system satisfies this condition because its microphones are located close to each other in a small, unitary multi-microphone structure. Ex. 1005, FIG. 5. Due to the close proximity of the microphones, any temporal difference in noise and speech

periods of the two signals would be minimal and negligible.² Third, Kaneda discloses that “[a] variation in S/N ratio difference caused by various factors is small (stability of the S/N ratio difference).” Ex. 1006, 2:33-35. Given the close proximity between the microphones in Meisel’s system, a POSITA would understand that the noise levels received at the microphones would be similar, leading to a stable speech to noise (S/N) ratio difference between the microphones. As such, a POSITA would have recognized that Meisel’s system was appropriate for using Kaneda’s “conventional” speech detection method of $PD = P1 - P2$, where PD is compared to a threshold value to differentiate speech from noise.

97. While Kaneda discloses that “variation factors”³ in conventional speech recognition systems may cause an unstable S/N ratio, Ex. 1006, 3:13-48, a POSITA would have recognized that the effects of the variation factors are

² A POSITA would understand that the temporal difference for sound arriving at microphones mounted as disclosed by Meisel (within several inches of each other) would be much less than 1 msec, which is negligible compared to the short-term power calculation which Kaneda disclosed occurs over a 30 msec window. Ex. 1006, 13:59-60.

³ These factors include the relative position of the noise source, movement of the speaker, and room echoes, which vary as a function of time. Ex. 1006, 2:13-48.

minimal in Meisel's system because of the close proximity of the microphones to each other. Indeed, a POSITA would have been motivated to implement Kaneda's conventional speech detection method in Meisel's speech recognition system to minimize the effects of the variation factors in accurate speech recognition.

98. Although Kaneda acknowledges that, in some circumstances, "the second conventional method has [a] decisive drawback," referring to the variation factors discussed above (Ex. 1006, 3:49-51), a POSITA would understand that this drawback does not apply to Meisel's speech recognition system for the reasons explained above. Meisel's microphones are located very close to each other, making the effects of the variation factors trivial among the microphones. Meisel's closely spaced microphone arrangement, therefore, meets the conditions for the conventional system in Kaneda to operate accurately *Compare* Ex. 1005, FIG. 5 *with* Ex. 1006, FIGS. 6(a)-6(b).

ii. A POSITA would have reasonably expected to succeed in modifying Meisel with Kaneda

99. A POSITA would have reasonably expected to succeed in combining Kaneda's acoustic signal processing to identify noise by generating difference parameters from acoustic microphones with Meisel's system.

100. Meisel describes a speech recognition system that includes three acoustic microphones (a nose microphone (16) and two mouth microphones (18,

32)). Ex. 1005, 2:11-17, 2:31-33. Kaneda's conventional speech recognition system similarly includes a microphone array having two acoustic microphones. Ex. 1006, 1:51-2:22. Also, Kaneda, like Meisel, discloses comparing acoustic signal amplitudes from the microphones to calculate differences in received signal amplitudes. Ex. 1006, 2:12-18. Given the similarities in the system architecture and the general approach of comparing signal amplitudes, a POSITA would have had a reasonable expectation of success in combining Meisel's speech recognition system with Kaneda's conventional signal processing technique of generating difference parameters. Kaneda's conventional signal processing for detecting noise uses the signals received and difference parameters already calculated in Meisel's system, adding an additional signal processing step to detect noise. Therefore, implementing Kaneda's noise detection in Meisel's system would not inhibit Meisel's signal processing functions, which are used to determine the types of speech for speech recognition. For example, Kaneda's conventional noise detection would be implemented in parallel with Meisel's speech detection.

101. Adding Kaneda's conventional signal processing technique for detecting noise using difference parameters merely combines known prior art elements according to known and expected functions, such as a known speech signal processing technique to calculate a difference between acoustic signals by subtracting one signal amplitude from another and comparing the difference to a

predetermined threshold. This combination would yield predictable results to accurately determine the type of speech or noise. Also, it would have been applying a known technique (a known speech signal processing technique of calculating a difference between acoustic signals by subtracting one signal amplitude from another and comparing the difference to a predetermined threshold) to a known device ready for improvement (a speech recognition system) to yield predictable results (accurately determine the type of speech).

5. Independent Claim 2

102. In my opinion, Meisel teaches or suggests each element of claim 2.

a. [2Preamble] “A method for removing noise from acoustic signals, comprising:”

103. As explained in [1h] above (*supra* § IX.A.4.i), Meisel discloses that detecting noise in acoustic signals to minimize its effects in speech detection was well known in the art. Moreover, it would have been obvious in view of Kaneda.

104. Notably, neither “removing noise” nor “denoising” is a recited step of independent claim 2. It is, instead, recited in dependent claim 4, discussed in § IX.A.6, *infra*. As such, removing noise is not a required element of claim 2. And, removing noise would have been obvious for the reasons discussed for dependent claim 4.

105. A POSITA would understand that identifying noise disclosed in Meisel would be beneficial to suppress or mitigate the impact of noise on speech recognition. Removing noise would therefore have been obvious to provide better speech recognition in environments where noise exists. Kaneda explains that filter control for minimizing the power of a noise component is a “well-known minimization problem” in speech recognition. Ex. 1006, 8:29-32. As such, a POSITA would have been motivated to mitigate this problem using denoising methods known in the art.

106. For example, Sakoe—cited in the background section of Meisel—describes one such noise-removing method. Meisel cites Sakoe and explains that Sakoe’s system “detect[s] background noise with a view toward minimizing its effect on recognition accuracy.” Ex. 1005, 1:46-50. Sakoe discloses a speech recognition system, which includes a “rejection unit which, when the intensity of said noise exceeding a predetermined standard value is detected in any of said speech signal durations, rejects the recognition results determined by said recognition unit.” Ex. 1008, 2:17-33. A POSITA would understand that rejecting the recognition results is removing noise from acoustic signals.

- b. **[2a] “receiving the acoustic signals at two receivers and receiving physiological information associated with human voicing activity at a voicing sensor;”**

107. This element is substantially similar to element [1a]-[1c]. As discussed above for [1a] through [1c], Meisel discloses these elements. *Supra* §§ IX.A.4.b-IX.A.4.d.

- c. **[2b] “generating cross correlation data between the physiological information and an acoustic signal received at one of the two receivers;”**

108. This element is identical to element [1d]. As discussed above for [1d], Meisel discloses this element. *Supra* § IX.A.4.e.

- d. **[2c] “identifying 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;”**

109. This element is substantially identical to [1e]. As explained in § IX.A.4.f, Meisel discloses this element.

- e. **[2d] “generating difference parameters between the acoustic signals received at each of the two receivers, wherein the difference parameters are representative of the relative difference in signal gain between portions of the received acoustic signals;”**

110. This element is substantially identical to element [1f]. As discussed above for [1f], Meisel renders this element obvious. *Supra* § IX.A.4.g.

- f. **[2e] “identifying information of the acoustic signals as unvoiced speech when the difference parameters exceed a gain threshold; and”**

111. This element is substantially identical to element [1g]. As discussed above for [1g], Meisel renders this element obvious. *Supra* § IX.A.4.h.

- g. **[2f] “identifying information of the acoustic signals as noise when the difference parameters are less than the gain threshold.”**

112. This element is substantially identical to element [1h]. As discussed above for [1h], Meisel renders this element obvious. *Supra* § IX.A.4.i.

6. Dependent Claim 4: “further comprising performing denoising on the identified noise.”

113. Claim 4 depends from claim 2. In my opinion, Meisel and Kaneda render obvious claim 4.

114. Meisel discloses performing denoising on the identified noise. As explained in [1h], Meisel and Kaneda disclose a process of identifying noise. *Supra* § IX.A.4.i. A POSITA would understand that the identification and minimization of noise disclosed in Meisel and Kaneda correspond to “performing denoising on the identified noise.” As explained above in [2Preamble] (*supra* § IX.A.5.a), Sakoe further discloses a noise removal process, indicating that denoising or noise removal was well known at the time of the invention.

115. Like Meisel and Kaneda, Sakoe is directed to “[a] speech recognition system adaptable to noisy environments,” which includes “a recognition unit for

recognizing input speech signals and a noise measuring unit for measuring the intensity of ambient noises,” and “a rejection unit responsive to a rejection standard controlled by the intensity of the measured noise” Ex. 1008, Abstract. Therefore, a POSITA would have been motivated to combine Sakoe’s noise removal process with the speech recognition systems disclosed in Meisel and Kaneda. Kaneda also discloses detecting noise and minimizing its effect on speech recognition. Ex. 1006, 4:59-63. Therefore, a POSITA would have had a reasonable expectation of success in combining Sakoe with Meisel and/or Kaneda.

7. Dependent Claim 5: “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.”

116. As explained in [1b] (*supra* § IX.A.4.c), Meisel discloses using an acoustic “throat microphone” (20) to “detect[] the vibration of the vocal cords, or ‘voicing.’” Ex. 1005, 2:31-35. A POSITA would understand that Meisel’s throat microphone (20) is an acoustic throat microphone because it “is disposed to detect vibration of the vocal cords through sounds emanating from the throat of a speaker.” Ex. 1005, 5:29-36.

B. [Ground 2] Claim 3 is rendered obvious by Meisel in view of Kaneda, Freidin, and Lyons

117. In my opinion, claim 3 would have been obvious to a POSITA based on Meisel in view of Kaneda, Freidin, and Lyons.

1. Overview of Lyons

118. Lyons discloses averaging digital signals using “statistical measures known as the mean, variance, and standard deviation” to minimize effects from unwanted random disturbances (noise) when measuring a parameter. Ex. 1012, 319, 477. Lyons teaches that the standard deviation of digital signals is “defined as the positive square root of the variance, or

$$\sigma = \sqrt{\frac{1}{N} \sum_{n=1}^N [x(n) - x_{ave}]^2} . \quad (8-3)$$

Ex. 1012, 320.

119. “If the sequence $x(n)$ represents a time series of signal samples, we can say that x_{ave} specifies the constant, or DC, value of the signal, [and] the standard deviation σ reflects the amount of the fluctuating, or AC, component of the signal” Ex. 1012, 320, 479.

120. Lyons’s digital signal processing has many applications, including “voice recognition[and] data communication” (Ex. 1012, 1), and cites to 22 publications directed to digital signal processing used in speech recognition (Ex. 1012, 21, 47, 127, 128, 155, 156, 217, 218, 295, 317, 318, 441, 442). A POSITA would understand that the standard deviation disclosed in Lyons is applicable to the speech recognition processes disclosed in Meisel.

121. Lyons would have been publicly available to a POSITA at the earliest priority date of the '058 patent. In fact, prior to May 30, 2001, I personally read portions of Lyons from a publicly available copy in the late 1990s and I personally know that Richard Lyons, including the text of Exhibit 1012, was highly regarded as a source of useful digital signal processing information for practitioners in the field. POSITAs would have sought out Lyons because it provides technical materials for the relevant undergraduate training and reference materials for post-graduate students. I know from personal experience that Lyons was available at university libraries in the late 1990s, prior to the earliest claimed priority date of the '058 patent. Therefore, Lyons would have been publicly accessible to a POSITA in the art before May 30, 2001.

2. Dependent Claim 3: “further comprising generating the gain threshold using standard deviations corresponding to the generation of the difference parameters.”

122. Claim 3 depends from claim 2. In my opinion, Lyons confirms and renders obvious using standard deviations corresponding to the generation of the difference parameters, as recited in claim 3.⁴

⁴ It is my opinion that Claim 3 is not described in the '058 patent. Moreover, a POSITA would not be able to understand the scope of “standard deviations corresponding to the generation of the difference parameters” recited in claim 3.

123. Meisel discloses the gain threshold for the reasons explained for [1g]. *Supra* § IX.A.4.h. Meisel does not expressly disclose generating the gain threshold using standard deviations corresponding to the generation of the difference parameters, but it would have been obvious to a POSITA based on well-known statistical techniques in the art. A POSITA would understand that applying this well-known statistical technique to calculate standard deviations of the differences in the received signals (alternatively, the difference of the standard deviations of each received signal) represents the general variation that corresponds to the meaningful threshold difference in the signals (gain thresholds). Ex. 1001, 8:41-47.

124. Indeed, it was well known at the time that calculating a standard deviation is a common statistical technique used for digital signal processing. Ex. 1012, 479 (explaining that a “useful measure of a signal sequence is the square root of the variance known as the *standard deviation*”); Ex. 1011, 14 (explaining that for most acquired signals that “do not show a well defined peak-to-peak value, but have a random nature . . . [a] more generalized method must be used in these cases, called the **standard deviation**, denoted by σ (a lower case Greek *sigma*)”).

To the extent that this element refers to applying a generic standard deviation calculation to the difference parameters or received signals, it would have been obvious to a POSITA as explained below.

Lyons discloses that the standard deviation of digital signals is “defined as the positive square root of the variance, or

$$\sigma = \sqrt{\frac{1}{N} \sum_{n=1}^N [x(n) - x_{ave}]^2} . \quad (8-3)$$

Ex. 1012, 320.

125. Moreover, a computer algorithm for calculating standard deviation in speech signal processing was also well known at the time. A POSITA would recognize that applying standard deviations to generating the gain threshold for speech signal processing would have been obvious because the standard deviation represents the background power in signal processing that corresponds to the threshold at which the difference in signals is meaningful.

a. A POSITA would have been motivated to combine Lyons with Meisel, Kaneda, and Freidin

126. A POSITA would have been motivated to apply standard deviation to Meisel’s speech recognition system as modified by Kaneda and Freidin, based on what was known in the art at the time. It was well known at the time that standard deviation is a common statistical measure to calculate an average deviation of input digital signals and minimize the effects of noise in measuring the true value of a parameter. Ex. 1012, 319, 476. A POSITA would understand that the standard deviation is applicable to the processing of speech signals disclosed in Meisel and

Kaneda because they are a type of digital signal. Moreover, Lyons cites to 22 different publications related to speech signal processing, demonstrating that the digital processing techniques disclosed in Lyons, including the standard deviation calculation, are applicable to speech processing. Ex. 1012, 21, 47, 127, 128, 155, 156, 217, 218, 295, 317, 318, 441, 442. A POSITA would understand that speech recognition techniques are derived from digital signal processing techniques disclosed in Lyons.

b. A POSITA would have reasonably expected to succeed in modifying Meisel, Kaneda, and Freidin with Lyons

127. A POSITA would have had reasonable expectation of success in applying standard deviation to the speech recognition processes disclosed in Meisel and Kaneda as modified by Freidin because doing so would have been combining prior art elements (difference parameters and a standard deviation) according to known methods (using any one of known methods of calculating standard deviations for speech signals) to yield predictable results (obtaining a standard deviation of the speech signals to generate a gain threshold). Also, it would have been obvious to try because standard deviation is only one of three statistical methods (mean, variance, and standard deviation) for speech signal processing. Ex. 1012, 477. Therefore, generating a gain threshold using standard deviations corresponding to the generation of the difference parameters would

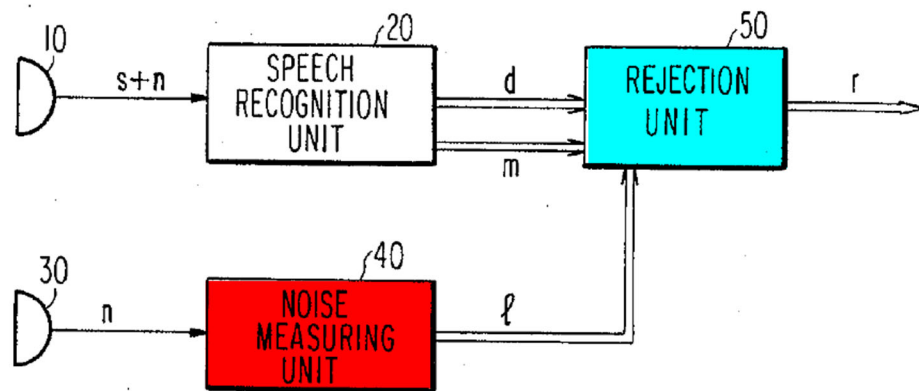
have been choosing from a finite number of identified, predictable solutions, with a reasonable expectation of success.

C. [Ground 3] Claim 4 is rendered obvious by Meisel, in view of Kaneda, Freidin, and Sakoe

128. In my opinion, claim 4 would have been obvious to a POSITA based on Meisel in view of Kaneda, Freidin, and Sakoe.

129. As explained in § IX.A.6, Meisel, Kaneda, and Freidin render claim 4 obvious. Claim 4 is further rendered obvious by Meisel in view of Kaneda, Freidin, and Sakoe.

130. Like Meisel and Kaneda, Sakoe is directed to “[a] speech recognition system adaptable to noisy environments,” which includes “a recognition unit for recognizing input speech signals and a noise measuring unit for measuring the intensity of ambient noises,” and “a rejection unit responsive to a rejection standard controlled by the intensity of the measured noise” Ex. 1008, Abstract. Therefore, Sakoe discloses “performing **denoising** on the **identified noise**,” and when combined with Meisel, Kaneda, and Freidin, renders claim 4 obvious.

FIG. 1A

Ex. 1008, FIG. 1A (annotated).

- a. **A POSITA would have been motivated to combine Sakoe with Meisel and Kaneda, as modified by Freidin**

131. A POSITA would have been motivated to combine Sakoe's noise removal process with the speech recognition systems disclosed in Meisel and Kaneda. As explained in § IX.A.4.i ([1h]), *supra*, Meisel discloses a process of detecting speech signals. Meisel's background section explains that Sakoe "detect[s] background noise with a view toward minimizing its effect on recognition accuracy." Ex. 1005, 1:46-50. Like Sakoe, Kaneda discloses detecting and minimizing noise effects on speech recognition. Ex. 1006, 4:59-63. Therefore, both Meisel and Kaneda, as modified by Freidin, provide some teaching, suggestion, or motivation to remove noise to improve the accuracy of Meisel's speech recognition system.

b. A POSITA would have reasonably expected to succeed in modifying Meisel, Kaneda, and Freidin with Sakoe

132. A POSITA would have reasonably expected to succeed in combining Sakoe's denoising process to remove noise. Combining Sakoe with Meisel and Kaneda, as modified by Freidin, would have been combining prior art elements (speech recognition system and denoising system) according to known methods (known denoising process) to yield predictable results (remove noise to increase the accuracy of speech detection). A POSITA also would have expected the use of a known technique (denoising) to improve similar devices (speech recognition systems) to obtain predictable results (removing noise to increase the accuracy of speech detection).

X. CONCLUSION

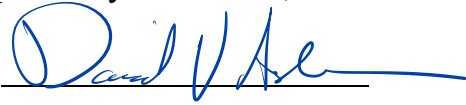
133. For the reasons set forth in Section IX it is also my opinion that one skilled in the art would have found claims 1-5 of the '058 patent obvious to the person of ordinary skill in the art.

134. 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: February 28, 2022

By: 

David V. Anderson