

UNITED STATES PATENT AND TRADEMARK OFFICE

---

BEFORE THE PATENT TRIAL AND APPEAL BOARD

---

**AMAZON.COM, INC.**  
Petitioner,

v.

**JAWBONE INNOVATIONS, LLC,**  
Patent Owner.

---

IPR2023-00222  
U.S. Patent No. 7,246,058

---

**DECLARATION OF RICHARD M. STERN, Ph.D.  
IN SUPPORT OF PETITION FOR *INTER PARTES* REVIEW  
OF U.S. PATENT NO. 7,246,058**

## TABLE OF CONTENTS

|       |                                                                                                            |    |
|-------|------------------------------------------------------------------------------------------------------------|----|
| I.    | BACKGROUND -----                                                                                           | 1  |
| A.    | Experience and Qualifications-----                                                                         | 1  |
| B.    | Topics of Opinions-----                                                                                    | 3  |
| C.    | Materials Considered -----                                                                                 | 3  |
| II.   | APPLICABLE LEGAL STANDARDS -----                                                                           | 5  |
| A.    | Claim Construction -----                                                                                   | 6  |
| B.    | Obviousness -----                                                                                          | 7  |
| III.  | PERSON OF ORDINARY SKILL IN THE ART -----                                                                  | 11 |
| IV.   | TECHNOLOGY BACKGROUND -----                                                                                | 13 |
| A.    | Using Microphones and Tissue Sensors Together to<br>Detect Voiced and Unvoiced Speech Was Well-Known ----- | 13 |
| B.    | Using Cross Correlation to Determine Whether a<br>Desired Signal Is Present Was Well-Known -----           | 16 |
| C.    | Using Two Microphones and Differences in Signal<br>Strength to Detect Speech Was Well-Known -----          | 16 |
| V.    | THE '058 PATENT -----                                                                                      | 21 |
| VI.   | SUMMARY OF OPINIONS AND LISTING OF GROUNDS -----                                                           | 24 |
| VII.  | THE REFERENCES ARE PRIOR ART -----                                                                         | 25 |
| VIII. | CLAIMS 1-5 OF THE '058 PATENT WOULD HAVE BEEN<br>OBVIOUS -----                                             | 26 |
| A.    | Ground 1a: Holzrichter, Takano, and Kinsler -----                                                          | 26 |
| 1.    | Claim 1 -----                                                                                              | 26 |

|    |                                                                                                                 |    |
|----|-----------------------------------------------------------------------------------------------------------------|----|
| a. | Preamble -----                                                                                                  | 26 |
| b. | At Least Two Microphones -----                                                                                  | 29 |
| c. | Voicing Sensor That Receives Physiological<br>Information Associated With Human<br>Voicing Activity -----       | 38 |
| d. | Processor Coupled Among the Microphones<br>and the Voicing Sensor -----                                         | 39 |
| e. | Generating Cross Correlation Data Between<br>the Physiological Information and an<br>Acoustic Signal -----      | 40 |
| f. | Identifying Information as Voiced Speech -----                                                                  | 46 |
| g. | Generating Difference Parameters Between<br>the Acoustic Signals Received at Each of<br>the Two Receivers ----- | 48 |
| h. | Identifying Information as Unvoiced Speech<br>When the Difference Parameters Exceed a<br>Gain Threshold-----    | 49 |
| i. | Identifying Information as Noise When the<br>Difference Parameters Are Less Than the<br>Gain Threshold-----     | 54 |
| 2. | Claim 2 -----                                                                                                   | 55 |
| 3. | Claim 3 -----                                                                                                   | 55 |
| 4. | Claim 4 -----                                                                                                   | 56 |
| 5. | Claim 5 -----                                                                                                   | 57 |
| B. | Ground 1b: Holzrichter, Takano, Kinsler, and Lyons -----                                                        | 58 |
| 1. | Claim 3 -----                                                                                                   | 58 |

|     |                                                                                                                 |    |
|-----|-----------------------------------------------------------------------------------------------------------------|----|
| C.  | Ground 2a: Holzrichter, Torick, Kaneda, and Kinsler -----                                                       | 62 |
| 1.  | Claim 1 -----                                                                                                   | 62 |
| a.  | At Least Two Microphones -----                                                                                  | 62 |
| b.  | Generating Difference Parameters Between<br>the Acoustic Signals Received at Each of the<br>Two Receivers ----- | 66 |
| c.  | Identifying Information as Unvoiced Speech<br>When the Difference Parameters Exceed a<br>Gain Threshold-----    | 68 |
| d.  | Identifying Information as Noise When the<br>Difference Parameters Are Less Than the<br>Gain Threshold-----     | 69 |
| 2.  | Claim 2 -----                                                                                                   | 70 |
| 3.  | Claim 4 -----                                                                                                   | 71 |
| 4.  | Claim 5 -----                                                                                                   | 71 |
| D.  | Ground 2b: Holzrichter, Torick, Kaneda, Kinsler, and<br>Lyons -----                                             | 71 |
| 1.  | Claim 3 -----                                                                                                   | 71 |
| IX. | SECONDARY CONSIDERATIONS -----                                                                                  | 72 |
| X.  | CONCLUSION -----                                                                                                | 72 |

I, Richard M. Stern, Ph.D., do hereby declare:

1. I am making this Declaration at the request of Amazon.com, Inc. and Amazon.com Services LLC. (“Amazon”). I have been retained by Amazon as a technical expert in this matter.

2. I am being compensated for my work on this case. My compensation does not depend on the content of this Declaration or the outcome of these proceedings. I do not own any stock in Amazon and, to my knowledge, I have no financial interest in Amazon.

3. I understand that Amazon is filing a Petition for *Inter Partes* Review of claims 1-5 of U.S. Patent No. 7,246,058 (“the ’058 patent”). I understand that the ’058 patent is purportedly owned by Jawbone LLC (“Patent Owner” or “Jawbone”). I have been asked to provide technical expert opinions, including an opinion concerning the patentability of claims 1-5 of the ’058 patent.

## **I. BACKGROUND**

### **A. Experience and Qualifications**

4. I am a Professor at Carnegie Mellon University in the Department of Electrical and Computer Engineering, the Department of Computer Science, and the Language Technologies Institute. I have been on the faculty of Carnegie Mellon since 1977.

5. I received a S.B. degree from the Massachusetts Institute of Technology (MIT) in 1970, a M.S. degree from the University of California, Berkeley, in 1972, and a Ph.D. from MIT in 1977, all in electrical engineering.

6. I am a fellow of the Institute of Electrical and Electronics Engineers (IEEE), the Acoustical Society of America, and the International Speech Communication Association (ISCA). I was the ISCA 2008-2009 Distinguished Lecturer, a recipient of the Allen Newell Award for Research Excellence in 1992, and I served as the General Chair of Interspeech 2006. Interspeech is the world's largest technical conference focused on speech processing and application.

7. Much of my current research is in spoken language systems, where I have been particularly concerned with the development of techniques with which automatic speech recognition can be made more robust with respect to changes in environment and acoustical ambience.

8. I have actively worked on the theory and application of systems to accomplish voice activity detection for more than twenty years. Most recently, my students developed the system that provided best performance worldwide in the 2021 Fearless Steps Phase-03 Speech Activity Detection (SAD) Challenge, which was organized and run by the University of Texas, Dallas (Vuong *et al.* 2021). My relevant publications, including those cited above, are available on Carnegie Mellon's web site at <http://www.cs.cmu.edu/afs/cs/user/robust/www/papers.html>.

9. I understand a copy of my current curriculum vitae, which lists my publications for the last ten years, is being submitted as Exhibit 1017.

**B. Topics of Opinions**

10. I offer opinions in this Declaration on the following general topics:

- the subject matter described and claimed in the '058 patent;
- the level of ordinary skill in the art pertaining to the '058 patent;
- the teachings of the prior art; and
- whether claims 1-5 of the '058 patent would have been obvious to a person of ordinary skill in the art at the time of the alleged invention in view of the prior art.

11. I also provide background information on systems using microphones and tissue sensors to detect voiced and unvoiced speech, systems using two microphones and differences in signal strength to detect speech, and systems using cross correlation to determine whether a desired signal is present.

12. Section headings, figure captions, table headings, and table captions in this Declaration are descriptive, not limiting.

**C. Materials Considered**

13. In preparing this Declaration, I have considered the following materials:

| Exhibit No. | Description                                                                                                                                                                                                                                      |
|-------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1001        | U.S. Patent No. 7,246,058 (“the ’058 patent”)                                                                                                                                                                                                    |
| 1003        | U.S. Patent No. 6,006,175 (“Holzrichter”)                                                                                                                                                                                                        |
| 1004        | Excerpts from Kinsler, Lawrence E., et al., FUNDAMENTALS OF ACOUSTICS (Wiley 1982) (“Kinsler”)                                                                                                                                                   |
| 1005        | Certified Translation of Japanese Patent Application Publication No. H-11-305792 (“Takano”)                                                                                                                                                      |
| 1006        | Torick, E.L., et al., <i>An Interphone System for “Hands Free” Operation in High Ambient Noise</i> , 14 I1008 Transactions on Audio and Electroacoustics 4 (Dec. 1966) (“Torick”)                                                                |
| 1007        | U.S. Patent No. 5,208,864 (“Kaneda”)                                                                                                                                                                                                             |
| 1008        | Lyons, Richard G., UNDERSTANDING DIGITAL SIGNAL PROCESSING (Addison-Wesley 1997) (“Lyons”)                                                                                                                                                       |
| 1009        | Excerpts from David Crystal, THE PENGUIN DICTIONARY OF LANGUAGE, SECOND EDITION (Penguin Books 1999)                                                                                                                                             |
| 1010        | Gregory C. Burnett, <i>The Physiological Basis of Glottal Electromagnetic Micropower Sensors (GEMS) and Their Use in Defining an Excitation Function for the Human Vocal Tract</i> , Dissertation, University of California at Davis (Jan. 1999) |
| 1011        | U.S. Patent No. 6,542,857                                                                                                                                                                                                                        |
| 1012        | U.S. Patent No. 7,283,948                                                                                                                                                                                                                        |
| 1013        | Excerpts from Michael J. Chapman <i>et al.</i> SIGNAL PROCESSING IN ELECTRONIC COMMUNICATIONS (Horwood Publishing Limited 1997)                                                                                                                  |

| <b>Exhibit No.</b> | <b>Description</b>                                                                                                              |
|--------------------|---------------------------------------------------------------------------------------------------------------------------------|
| 1014               | The '058 patent's file history                                                                                                  |
| 1015               | Excerpts from Paul A. Lynn, AN INTRODUCTION TO THE ANALYSIS AND PROCESSING OF SIGNALS, SECOND EDITION (Macmillan 1982) ("Lynn") |
| 1016               | U.S. Patent No. 5,809,455                                                                                                       |
| 1018               | Japanese Patent Application Publication No. H-11-305792 ("Takano")                                                              |

14. I have also relied on my education, training, and experience, and my knowledge of pertinent literature in the field of the '058 patent.

## **II. APPLICABLE LEGAL STANDARDS**

15. I have been asked to provide my opinion as to 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 alleged invention, in view of the prior art.

16. I am an electrical engineer by training and profession. The opinions I am expressing in this Declaration involve the application of my training and technical knowledge and experience to the evaluation of certain prior art with respect to the '058 patent.

17. Although I have been involved as a technical expert in patent matters before, I am not an expert in patent law. Therefore, the attorneys from Knobbe, Martens, Olson & Bear, LLP have provided me with guidance as to the applicable

patent law in this matter. The paragraphs below express my understanding of how I must apply current principles related to patent validity to my analysis.

**A. Claim Construction**

18. It is my understanding that in determining whether a patent claim is obvious in view of the prior art, the Patent Office construes the claim by giving the claim terms their plain and ordinary 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 (patent specification and file history). For the purposes of this review, and to the extent necessary, I have interpreted each claim term in accordance with its plain and ordinary meaning as it 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. I have been instructed to assume that the time of the invention is May 30, 2001, which I understand to be the date of the earliest provisional application to which the '058 patent claims priority was filed.

19. I understand that a patent and its prosecution history are considered “intrinsic evidence” and are the most important sources for interpreting claim language in a patent. I also understand that in reading the claim, I must not import limitations from the specification into the claim terms; in other words, I must not narrow the scope of the claim terms by implicitly adding disclosed limitations that

have no express basis in the claims. The prosecution history of related patents and applications can also be relevant.

20. I understand that sources extrinsic to a patent and its prosecution history (such as dictionary definitions and technical publications) may also be used to help interpret the claim language, but that such extrinsic sources cannot be used to contradict the unambiguous meaning of the claim language that is evident from the intrinsic evidence.

21. Unless expressly stated herein, I have applied the plain and ordinary meaning of the claim terms, which I understand is the meaning that a person of ordinary skill in the art would have given to terms in May 2001 based on a review of the intrinsic evidence.

## **B. Obviousness**

22. It is my understanding that a claim is “obvious” if the claimed subject matter as a whole would have been obvious to a person of ordinary skill in the art at the time of the alleged invention. I understand that an obviousness analysis involves a number of considerations. I understand that the following factors must be evaluated to determine whether a claim would have been obvious: (i) the scope and content of the prior art; (ii) the differences, if any, between each claim of the ’058 patent and the prior art; (iii) the level of ordinary skill in the art in May 2001; and (iv) additional considerations, if any, that indicate that the invention was obvious or

not obvious. I understand that these “additional considerations” are often referred to as “secondary considerations” of non-obviousness or obviousness.

23. I also understand that the frame of reference when evaluating obviousness is what a hypothetical person of ordinary skill in the pertinent art would have known in May 2001. I understand that the hypothetical person of ordinary skill is presumed to have knowledge of all pertinent prior art references.

24. I understand that a prior art reference may be a pertinent prior art reference (or “analogous art”) if it is in the same field of endeavor as the patent or if it is pertinent to the problem that the inventors were trying to solve. A reference is reasonably pertinent if it logically would have commended itself to an inventor’s attention in considering the problem at hand. If a reference relates to the same problem as the claimed invention, that supports use of the reference as prior art in an obviousness analysis. Here, all of the references relied on in my obviousness analysis below are in the same field of endeavor as the ’058 patent, e.g., processing audio signals. The references are also pertinent to a particular problem the inventor was focused on, e.g., enhancing the quality of audio signals using signal processing techniques.

25. It is my understanding that something is “inherent in,” and therefore taught by, the prior art, if it necessarily flows from the explicit disclosure of the prior art. I understand that the fact that a certain result or characteristic may be present in

the prior art is not sufficient to establish inherency. However, if the result or characteristic is necessarily present based upon the explicit disclosure in the prior art, it is inherent in the prior art and is therefore disclosed.

26. It is my understanding that the law recognizes several rationales for combining references or modifying a reference to show obviousness of claimed subject matter. Some of these rationales include:

- combining prior art elements according to known methods to yield predictable results;
- simple substitution of one known element for another to obtain predictable results;
- a predictable use of prior art elements according to their established functions;
- using known techniques to improve similar devices (methods, or products) in the same way;
- applying a known technique to a known device (method, or product) ready for improvement to yield predictable results;
- choosing from a finite number of identified, predictable solutions, with a reasonable expectation of success (in which case a claim would have been obvious to try);

- known work in one field of endeavor may prompt variations of it for use in either the same field or a different one based on design incentives or other market forces if the variations would have been predictable to one of ordinary skill in the art; and
- some teaching, suggestion, or motivation in the prior art that would have led one of ordinary skill to modify the prior art reference or to combine prior art reference teachings to arrive at the claimed invention.

27. I understand that “secondary considerations” must be considered as part of the obviousness analysis when present. I further understand that the secondary considerations may include: (1) a long-felt but unmet need in the prior art that was satisfied by the claimed invention; (2) the failure of others; (3) skepticism by experts; (4) commercial success of a product covered by the patent; (5) unexpected results achieved by the claimed invention; (6) industry praise of the claimed invention; (7) deliberate copying of the invention; and (8) teaching away by others. I also understand that evidence of the independent and nearly simultaneous “invention” of the claimed subject matter by others is a secondary consideration supporting an obviousness determination and may support a conclusion that a claimed invention was within the knowledge of a person of ordinary skill as of May 2001. I am not aware of any evidence of secondary considerations that would suggest that the claims of the '058 patent would have been non-obvious in May 2001.

28. I understand that when assessing obviousness, using hindsight is impermissible; that is, what is known today or what was learned from the teachings of the patent should not be considered. The patent should not be used as a road map for selecting and combining items of prior art. Rather, obviousness must be considered from the perspective of a person of ordinary skill at the time the alleged invention was made—May 2001 in this case.

29. I also understand that an obviousness analysis must consider the invention as a whole, as opposed to just a part or element of the invention. I understand this “as a whole” assessment to require showing that one of ordinary skill in the art at the time of invention, confronted by the same problems as the inventor and with no knowledge of the claimed invention, would have selected the elements from the prior art and combined them in the claimed manner.

### **III. PERSON OF ORDINARY SKILL IN THE ART**

30. It is my understanding that when interpreting the claims of the '058 patent and evaluating whether a claim would have been obvious, I must do so based on the perspective of a person of ordinary skill in the art at the relevant priority date. I understand that the earliest claimed priority date of the '058 patent is May 30, 2001. I have been instructed to assume for the purposes of my opinions that the relevant priority date of the '058 patent is May 30, 2001.

31. I understand that in determining the level of ordinary skill in the art, several factors are considered. Those factors may include: (i) the type of problems encountered in the art; (ii) prior art solutions to those problems; (iii) the rapidity with which innovations are made; (iv) the sophistication of the technology; and (v) the educational level of active workers in the field. A person of ordinary skill in the art must have the capability of understanding the scientific and engineering principles applicable to the pertinent art.

32. Based on my review of the specification and claims of the '058 patent, it is my opinion that a person of ordinary skill in the art 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 could substitute for formal education and additional formal education could substitute for work experience.

33. My conclusions below that the claims of the '058 patent were obvious would remain the same even if the priority date, field of endeavor, or level of ordinary skill were slightly different.

34. I meet the above definition of a person of ordinary skill in the art, and did so as of May 2001. Also, I have worked with persons of ordinary skill in the art

through my professional and academic experiences, and I have an understanding of their skill level around May 2001.

#### **IV. TECHNOLOGY BACKGROUND**

35. The '058 patent claims a speech detection system comprised of at least two microphones, a voicing sensor, and a processor that generates cross correlation data. Systems using microphones and tissue sensors to detect speech and two-microphone speech detection systems, as well as the concept of using cross correlation to determine whether a desired signal is present, were well-known in the art long before May 30, 2001.

##### **A. Using Microphones and Tissue Sensors Together to Detect Voiced and Unvoiced Speech Was Well-Known**

36. Voiced speech, which is speech sounds produced using vocal fold vibration, and unvoiced speech, which is speech sounds made without vocal fold vibrations, are well-known categories of human speech. (Ex. 1009.)

37. In 1996, John F. Holzrichter, a professor at the Lawrence Livermore National Laboratory, filed a patent application for a system that uses a voicing sensor and a microphone together to detect voiced and unvoiced speech, which issued as a patent in 1999. (Ex. 1003.)

38. Holzrichter's patent described a system and sensor for detecting voiced and unvoiced speech in the presence of varying background noise. (*Id.*) The system included **electromagnetic (EM) sensors directed towards the speech organs (69)**

and a **traditional acoustic microphone (70)**. (*Id.*, 31:15-17 (disclosing “microphone 70”), 32:31-35 (disclosing “tongue and jaw motion EM sensors”), Fig. 19.) The **EM sensors** use radar waves to detect physiological information associated with human voicing activity, such as jaw, tongue, and vocal fold motions. (*Id.*, 32:31-35, 8:1-24, claim 42.)

39. To identify voiced and unvoiced speech, Holzrichter describes separating a portion of the audio signal received from **microphone 70** into a frame. (*Id.*, 31:61-63.) Then, Holzrichter’s system determines if the strength (also known as the “gain”) of the output signal in that frame exceeds **a preset threshold (73)** and is lower than a **noise threshold (77)**. (*Id.*, 31:67-32:4.) If the gain exceeds the preset threshold (73) and is lower than a noise threshold (77), then the system identifies the frame as speech. (*Id.*, 31:67-32:10.) If the gain is not within these thresholds, then the system can **discard the frame as noise (81)**. (*Id.*, 33:43-46.)

40. For frames identified as speech, Holzrichter’s system determines whether the speech is voiced or unvoiced. (*Id.*, 31:67-32:10.) This determination is made by an algorithm that “looks for vocal fold motion to occur within the same speech time period[.]” (*Id.*, 31:67-32:4, Fig. 19 (box 79).) If vocal movements occurred at the same time as the speech frame, Holzrichter’s system identifies the frame as **voiced speech (78)**. (*Id.*, 32:4-7.) If no vocal movements occurred, the frame is labeled as **nonvoiced speech (79)**. (*Id.*, 32:7-10.) Holzrichter explains that

[illegible]

-15-

**B. Using Cross Correlation to Determine Whether a Desired Signal Is Present Was Well-Known**

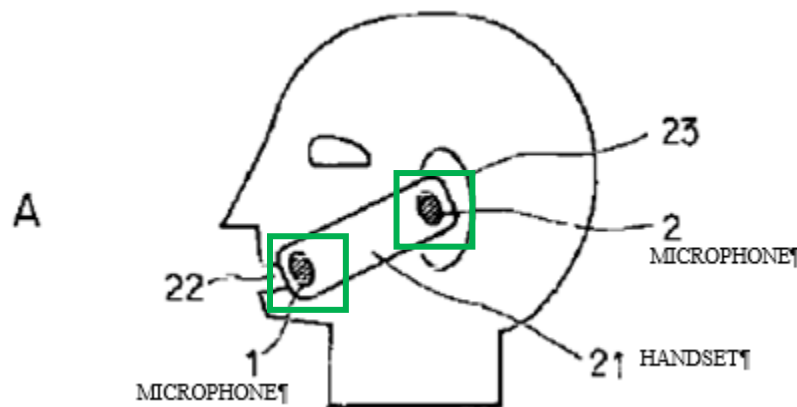
41. Holzrichter does not describe using a specific mathematical function for correlating the vocal fold movements and microphone signals, but various correlation functions were well-known in the art. For decades, the cross-correlation function has been used to determine whether a desired signal (e.g., speech) is present. For example, 40 years ago, Kinsler disclosed methods of detecting a known signal, such as human speech, in an acoustic sample. (Ex. 1004.) Kinsler explained that the “optimum processing system” to search for a “known signal” in a “received signal” is through the operation of “cross correlation.” (*Id.*, 256.) Kinsler taught that the value of the cross-correlation calculation, referred to as  $F(T)$ , could be compared to a pre-determined fixed threshold to determine whether a signal is detected in the presence of background noise. (*Id.*, 256.) If the value of  $F(T)$  meets or exceeds the threshold, then the desired signal (e.g., speech) is deemed to be present in the noise. (*Id.*, 252, 257.)

**C. Using Two Microphones and Differences in Signal Strength to Detect Speech Was Well-Known**

42. The system described in Holzrichter used a single microphone, but similar systems that used two microphones were widely known. For example, Takano, which published in 1999, disclosed a speech recognition system having **two microphones**. Takano described placing these microphones in a headset with

**microphone 1** located closer to a user's mouth and **microphone 2** located farther away from the user's mouth (e.g., near the user's ear). (*Id.*, [0022], [0050]-[0052].)

An example is shown in Takano's Figure 6:



(*Id.*, Fig. 6 (annotated).)

43. Takano discloses using the spatial difference between microphones to detect speech. Specifically, the system receives acoustic signals at each microphone and splits each signal into different frequency components. (*Id.*, [0009], [0023-0025], Figs. 7, 16, 17.) For each frequency component, the system calculates a “level difference” between the two signals. (*Id.*, [0009], [0023-0025], Figs. 7, 16, 17.) Because **microphone 1** is positioned closer to the user's mouth, the “level difference in the outputs of the microphones 1 and 2 for a sound from the sound source of the target signal is large,” while the “the level difference in the outputs of the microphones 1 and 2 due to noise is zero or minimal.” (*Id.*, [0022], [0015].)

44. Takano teaches that the **level difference** can be compared against **a threshold** to identify a signal as speech or noise. (*Id.*, [0009] (“level difference for each frequency component is compared to a preset threshold value by a sound source frequency component selection means”), [0023-0025], Figs. 7, 17.) When the level difference exceeds **a threshold**, Takano identifies the signal as **speech**. (*Id.*, [0009], [0023-0025], Figs. 7, 17.) If the level difference is lower than the threshold, Takano’s system identifies the signal as **noise**. (*Id.*, [0015] (“the output level difference between the two microphones arising due to the target signal is always larger than the level difference of the output level difference of the two microphones arising due to noise”), [0023-0025], Figs. 7, 17.)

45. Takano further teaches that the threshold “can be determined experimentally.” (*Id.*, [0026].) Takano’s method of calculating the “level difference” and comparing it to a threshold to determine whether a signal is speech or noise is shown in Takano’s Figure 9:

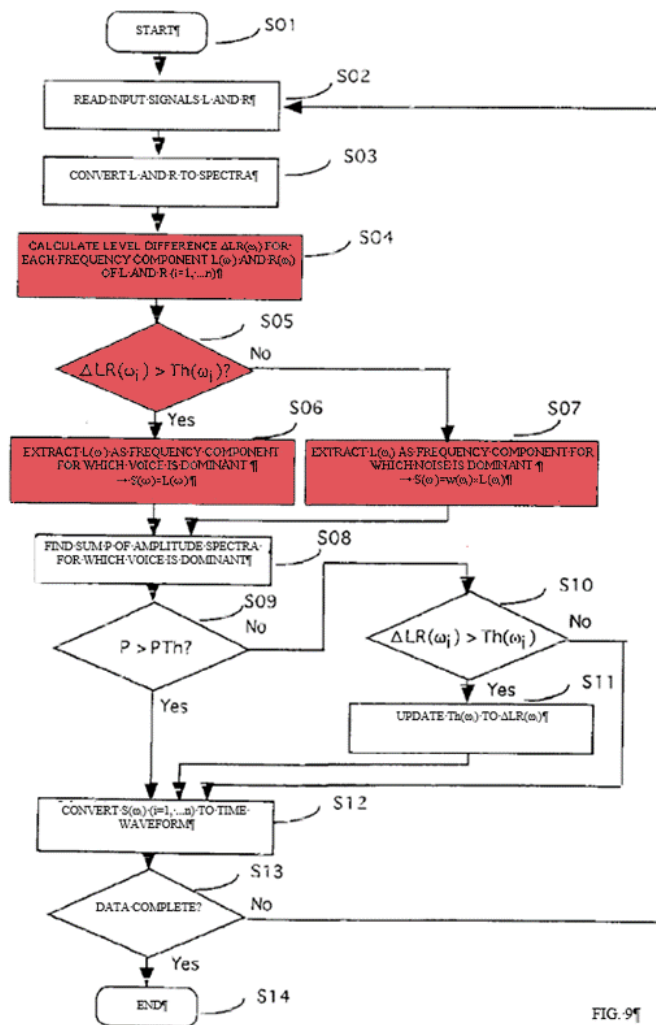


FIG. 9

(*Id.*, Fig. 9 (annotated).)

46. As another example, Torick, which published over 50 years ago, disclosed a system for detecting “speech sounds” using two microphones. (Ex. 1006, 169.) One microphone is used to detect speech, while the other is “used to sense the ambient noise.” (*Id.*)

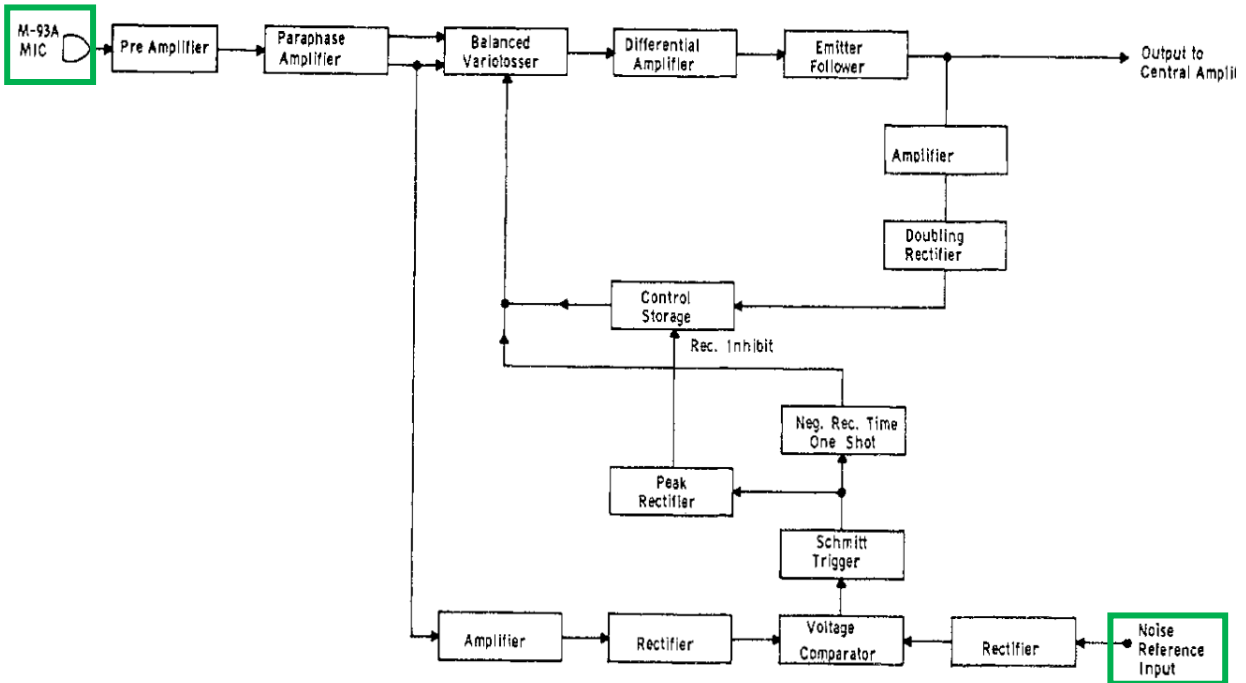


Fig. 4. Crew member's amplifier and control unit.

(*Id.*, Fig. 4 (annotated).)

47. “The noise-sensing microphone output is amplified and equalized to correspond to the far-field response of the noise-cancelling speech microphone. The outputs of both microphones are then continuously compared.” (*Id.*, 169.) When the signals from the two microphones are equal, the system determines that no speech is occurring. (*Id.*) However, when speech sounds occur at the speech microphone, the level between the two microphones is no longer equal and speech is detected. (*Id.*) Torick explains that “this is analogous to the presence of a virtual threshold which rises and falls with corresponding changes in noise levels.” (*Id.*)

48. Kaneda, which published in 1993, is another example. Kaneda disclosed that speech can be distinguished from noise based on the “power P1 of the

output from [a] first microphone” and the “power P2 of the output from [a] second microphone.” (Ex. 1007, 2:1-22, Fig. 3(a)-3(c).) Kaneda explained that “a difference PD ( $=P1-P2$ ) between the short time powers P1 and P2 of the two signals is calculated. When the power difference PD is larger than a given threshold value Pth17, a corresponding time period 18 is detected as a speech period.” (*Id.*, 2:13-18.) If the power difference is lower than the threshold, the period is identified as noise. (*See id.*, 1:51-68, 2:1-22, Fig. 3(a)-3(c).)

49. As these references demonstrate, it was widely known in the art to calculate a difference in gain (power) between acoustic signals received at two microphones and identify the signal as speech when that difference exceeds a threshold and as noise when that difference is less than the threshold.

## **V. THE '058 PATENT**

50. The '058 patent combines elements from the prior art in predictable ways. The patent describes a two microphone system “for detecting voiced and unvoiced speech in acoustic signals having varying levels of background noise.” (Ex. 1001 ('058 patent), Abstract.) The figures below have been annotated with colors for clarity.

51. The '058 patent describes a system for detecting voiced and unvoiced speech using acoustic and non-acoustic (e.g., physiological) sensors. (*Id.*, Abstract,

Figs. 1, 3.) The system includes a **processor (30)** connected to **acoustic microphones 10 (MIC 1 and MIC 2)** and a **voicing sensor 20 (VAD)**:

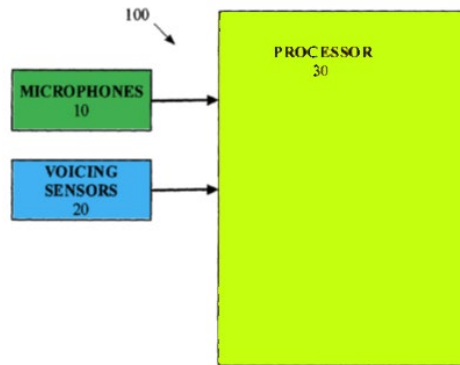


Figure 1

(*Id.*, Fig. 1 (annotated).) The **acoustic microphones 10 (MIC 1 and MIC 2)** receive acoustic signals ( $m_1(n)$  and  $m_2(n)$ ), while the **voicing sensor 20 (VAD)** receives physiological information associated with human voicing activity, such as vocal fold vibrations:

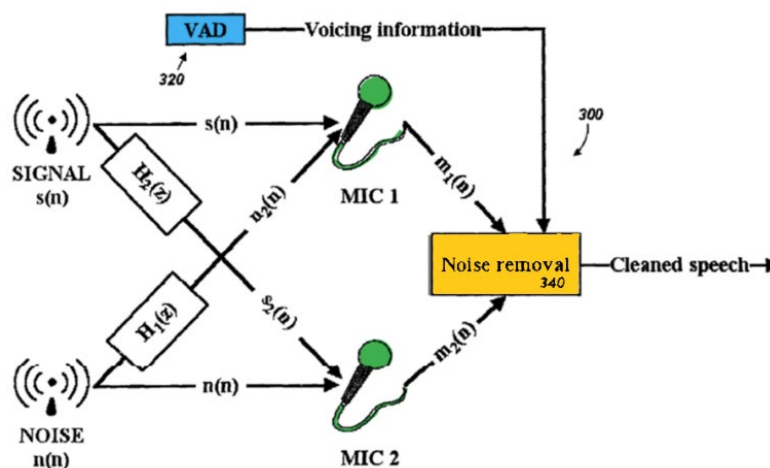


Figure 3

(*Id.*, Fig. 3 (annotated), 2:30-37, 3:64-4:13, 5:26-29.) The **voicing sensor 20** may take several forms, including a radio frequency (radar) sensor. (*Id.*, 3:64-66.) The **acoustic microphones 10** are placed in a headset with MIC 1 placed closer to the user's mouth and MIC 2 placed farther away from the user's mouth, as shown in Figure 7:

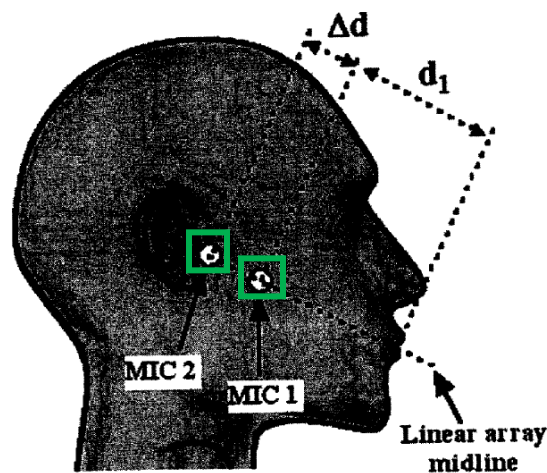


**Figure 7**

(*Id.*, Fig. 7, 7:52-67 (“the difference in signal strength in Mic 1 and Mic 2 ... will be proportional to both  $d_1$  and  $\Delta d$ ”), 8:16-17 (the gain increases when the user is speaking because “Mic 1 is closer to the mouth than Mic 2”).)

52. **Processor 30** processes the signals in two ways. First, it generates cross correlation data correlating the physiological information from the **voicing sensor** and an acoustic signal received from one of the acoustic **microphones**. (*Id.*, 4:64-5:3, 5:11-25.) **Processor 30** compares the cross correlation data to a threshold

and, if the correlation data exceeds the threshold, the processor identifies information in the acoustic signal as voiced speech. (*Id.*, 5:39-45.)

53. Second, the processor generates difference parameters, which represent a relative difference in signal gain between the acoustic signals at the two microphones, by comparing portions of the acoustic signals. (*Id.*, 5:39-45, 7:52-67.) The processor compares the difference parameters to a predetermined gain threshold and, if the difference parameters exceed the threshold, the processor identifies information in the acoustic signals as unvoiced speech. (*Id.*, 8:21-48, Fig. 4.) If the difference parameters are less than the gain threshold, the processor identifies the acoustic signals as noise. (*Id.*, 8:41-43.) The processor may then remove the identified noise (noise removal 340). (*Id.*, 7:18-19.)

## **VI. SUMMARY OF OPINIONS AND LISTING OF GROUNDS**

54. In my opinion, claims 1-5 of the '058 patent would have been obvious to a person of ordinary skill in the art before May 30, 2001 based at least on the below-identified combinations of prior art references:

| Ground | Claims Challenged | Prior Art References                        |
|--------|-------------------|---------------------------------------------|
| 1a     | 1-5               | Holzrichter, Takano, Kinsler                |
| 1b     | 3                 | Holzrichter, Takano, Kinsler, Lyons         |
| 2a     | 1, 2, 4, 5        | Holzrichter, Torick, Kaneda, Kinsler        |
| 2b     | 3                 | Holzrichter, Torick, Kaneda, Kinsler, Lyons |

## **VII. THE REFERENCES ARE PRIOR ART**

55. I understand that the '058 patent issued on July 17, 2007, from U.S. Patent Application 10/159,770 (“the '770 application”), filed May 30, 2002. I also understand that the '770 application claims priority to ten U.S. provisional applications, the earliest of which is U.S. Provisional Application No. 60/294,383 (“the '383 application”), filed May 30, 2001. (Ex. 1001.) The '770 application also claims priority to U.S. Provisional Application Nos. 60/335,100, filed on October 30, 2001; 60/332,202, filed on November 21, 2001; 60/362,162, 60/362,103, 60/362,170, 60/361,981, and 60/362,161, all filed on March 5, 2002; 60/368,208 and 60/368,209, both filed on March 27, 2002; and 60/368,343, filed on March 27, 2002. (*Id.*) For purposes of this Declaration, I will assume the earliest possible priority date of May 30, 2001 without conceding that the '058 patent is entitled to that date.

56. I have been instructed that the references I rely on in the Grounds discussed herein are prior art because they published more than one year before May 30, 2001:

- (1) Holzrichter published on December 21, 1999 (Ex. 1003);
- (2) Kinsler published in 1982 (Ex. 1004);
- (3) Takano published on November 5, 1999 (Ex. 1005);
- (4) Torick published in December 1966 (Ex. 1006);

- (5) Kaneda published on May 4, 1993 (Ex. 1007); and
- (6) Lyons published in 1997 (Ex. 1008).

57. As discussed previously, and as is evident from the discussions of the prior art references that follow in connection with each Ground, all of these references are analogous art because they are from the same field of endeavor as the '058 patent and because they are pertinent to a problem the inventor was focused on.

### **VIII. CLAIMS 1-5 OF THE '058 PATENT WOULD HAVE BEEN OBVIOUS**

58. In view of the background provided above, and further in view of the reasons set forth below, 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 by May 2001.

#### **A. Ground 1a: Holzrichter, Takano, and Kinsler**

59. It is my opinion that Holzrichter, Takano, and Kinsler render claims 1-5 obvious.

#### **1. Claim 1**

##### **a. Preamble**

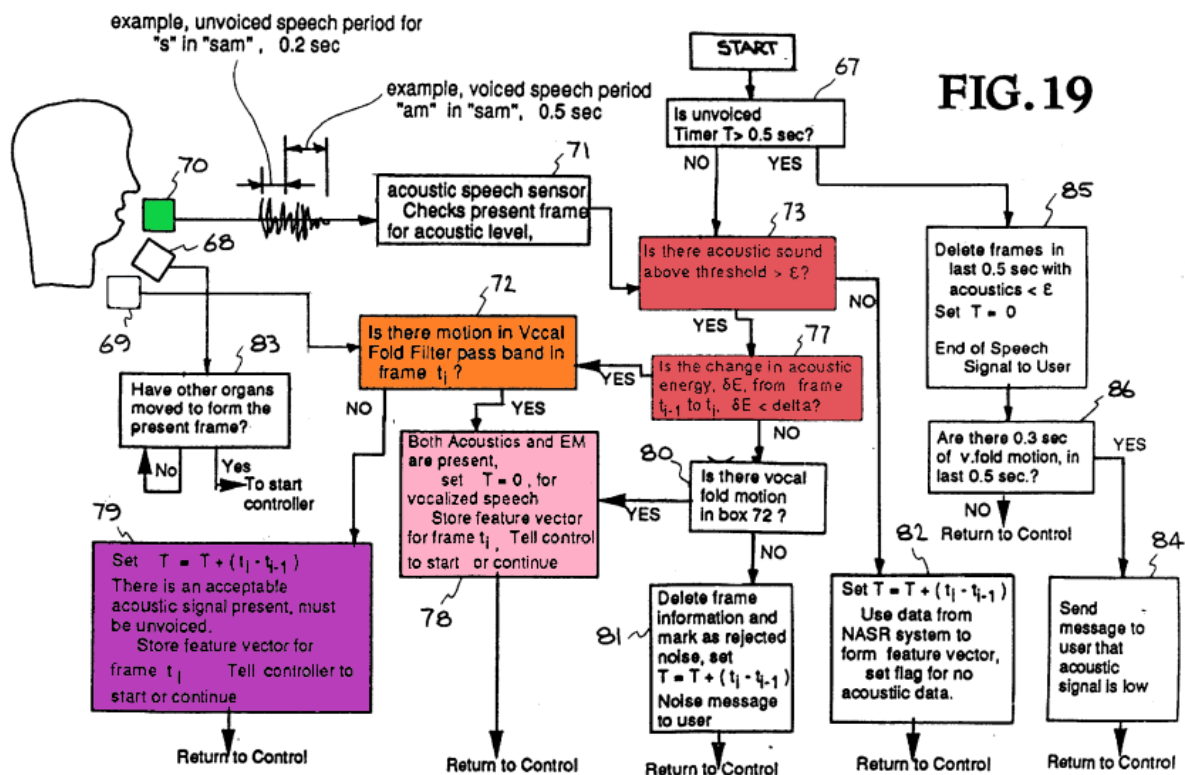
60. The preamble recites “a system for detecting voiced and unvoiced speech in acoustic signals having varying levels of background noise[.]” It is my opinion that Holzrichter discloses such a system.

61. As described further below, Holzrichter discloses a system and method for detecting the “start of speech, end of speech, identification of voiced or unvoiced

phoneme, presence of pause, and extraneous noise presence.” (Ex. 1003, 15:62-65.)

Holzrichter likewise discloses a method for determining whether “voiced speech, or unvoiced speech, or speech or background noise” has occurred. (*Id.*, 52:66-53:2.)

Holzrichter further discloses a “voiced-unvoiced algorithm” (*Id.*, 34:41-67.) Figure 19 “illustrates one method for determining the onset of speech, noise rejection, voiced-unvoiced speech, end of speech, and pauses using acoustic and non-acoustic information together.” (*Id.*, 31:8-11.)



(*Id.*, Fig. 19 (annotated).)

62. Table II provides examples of the voiced and unvoiced speech sounds that Holzrichter’s system can identify.

TABLE II

| Voiced and Non Voiced English sound pairs discernible using EM sensors |                 |                  |                  |
|------------------------------------------------------------------------|-----------------|------------------|------------------|
| voiced                                                                 | unvoiced        | voiced           | unvoiced         |
| b                                                                      | p               | v                | f                |
| d                                                                      | t               | th (as in then)  | th (as in thin)  |
| g as in g ut                                                           | k               | z                | s                |
| j                                                                      | c (as in chore) | g (as in garage) | sh (as in shore) |

(*Id.*, Table II.)

Holzrichter Figure 9A shows the acoustic signal recorded by Holzrichter’s system and Figure 9B shows glottal tissue motion signals recorded by Holzrichter’s system for the word “fox.” (*Id.*, 22:6-28.) “One sees clearly, and it is easily discriminated algorithmically, the unvoiced /f/ and then the voiced /o/, and the transition /x/ in ‘fox’.” (*Id.*)

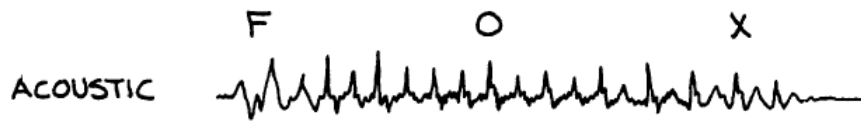
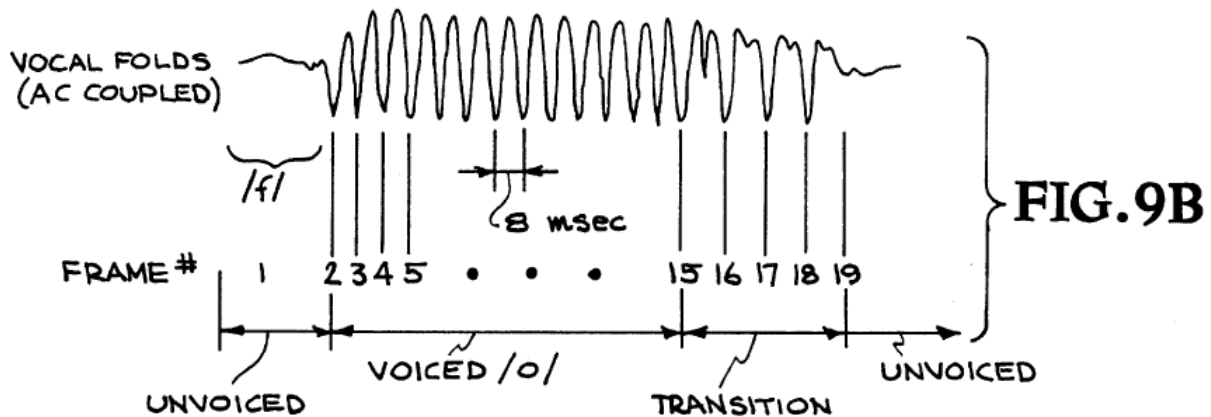


**FIG. 9A**



(*Id.*, Figs. 9A and 9B.)

**b. At Least Two Microphones**

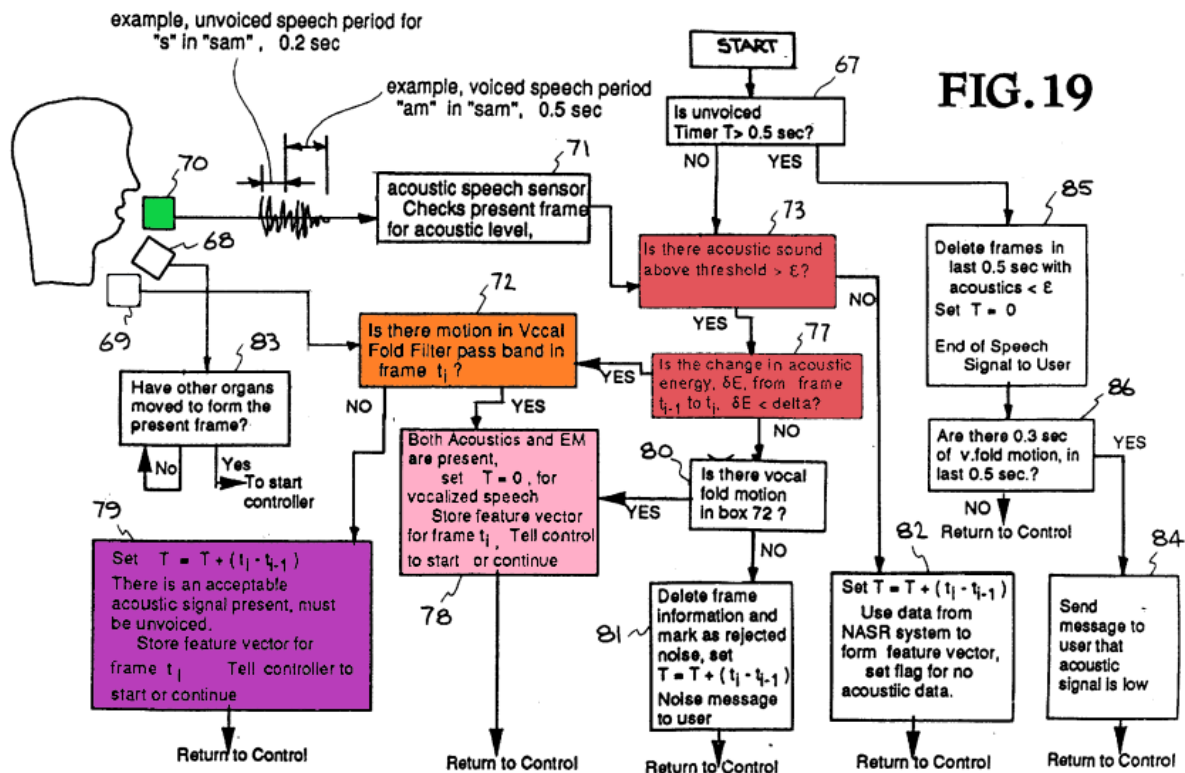
63. Claim 1 recites “at least two microphones that receive the acoustic signals[.]” It is my opinion that Holzrichter combined with Takano discloses this limitation.

64. Holzrichter discloses using a microphone for receiving acoustic signals. (Ex. 1003, 31:15-17.) A person of ordinary skill in the art would have found it obvious to improve Holzrichter’s system by adding a second microphone as disclosed in Takano. There is nothing unique or novel about adding an additional

microphone. A person of ordinary skill in the art would have expected such a modification to improve Holzrichter's methods for classifying input signals as either noise or speech based on signal strength.

65. Holzrichter discloses a microphone (70) for receiving acoustic signals. (Ex. 1003, 31:15-17 (“The acoustic information from microphone 70 is inputted into an acoustic speech sensor, A/D converter and 1 sec. memory 71.”) Holzrichter further discloses that its system can include “microphones against the human face and neck[.]” (*Id.*, 8:50-9:3.) After receiving the acoustic signal, Holzrichter's system uses the strength of the signal to determine whether the signal is noise or speech. (*Id.*, 31:67-32:10, Fig. 19.) If the acoustic “signal exceeds a preset threshold 73, but ... is less than a loud noise threshold 77,” then the system can identify the signal as speech.

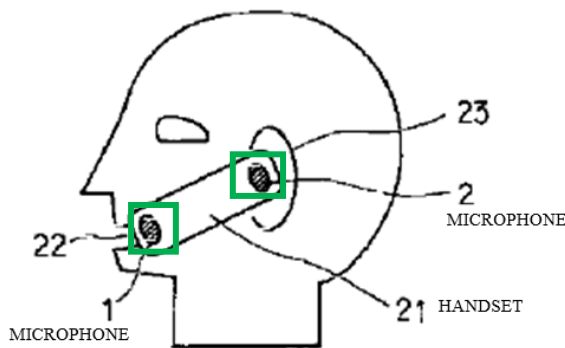
66. Holzrichter's Figure 19 illustrates this process. As shown, sound is received by Holzrichter's **microphone 70**. The system then checks the “acoustic energy” in the signal (**steps 73 and 77**) to determine whether the signal is speech (or noise) and, if speech is present, also evaluates whether the signal is voiced or unvoiced speech (**step 72**, **step 78**, and **step 79**).



(*Id.*, Fig. 19 (annotated).) Thus, Holzrichter discloses using a microphone to receive acoustic signals and then comparing the signal strength against pre-set thresholds to determine whether the signals are speech.

67. Takano discloses a similar process but uses two microphones for receiving acoustic signals. (Ex. 1005, [0009].) Takano discloses “a first microphone that is located at a location that is near to the sound source of the target signal and a second microphone that is located at a location that is further from the sound source of the target signal than the aforementioned location.” (*Id.*, *see also* [0014]-[0015], [0023-0025], [0029], [0049]-[0051], Figs. 1, 6, 7, 16, 17, 18A.)

68. Takano's Figure 6 shows these microphones, which are virtually identical to the two microphones described in the '058 patent:



Ex. 1005 (Takano), Fig. 6

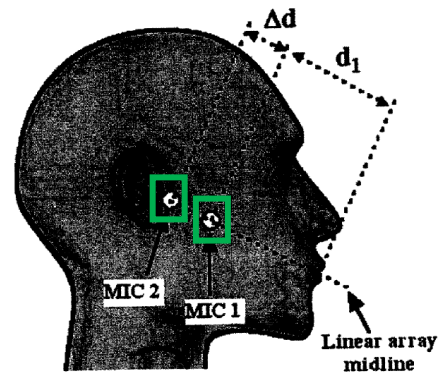


**Figure 7**

Ex. 1001 ('058 Patent), Fig. 7

69. As in Holzrichter, Takano's system receives acoustic signals at these two microphones and then uses the acoustic energy of the signals received by the microphones to determine whether the system should further analyze the signals for speech. (*Id.*, [0009], [0023-0025], [0014]-[0015], [0029], [0040]-[0042], [0049]-[0051], Figs. 1, 6, 7, 8, 9, 16, 17, 18A.). Takano's Figure 9, below, illustrates this process.

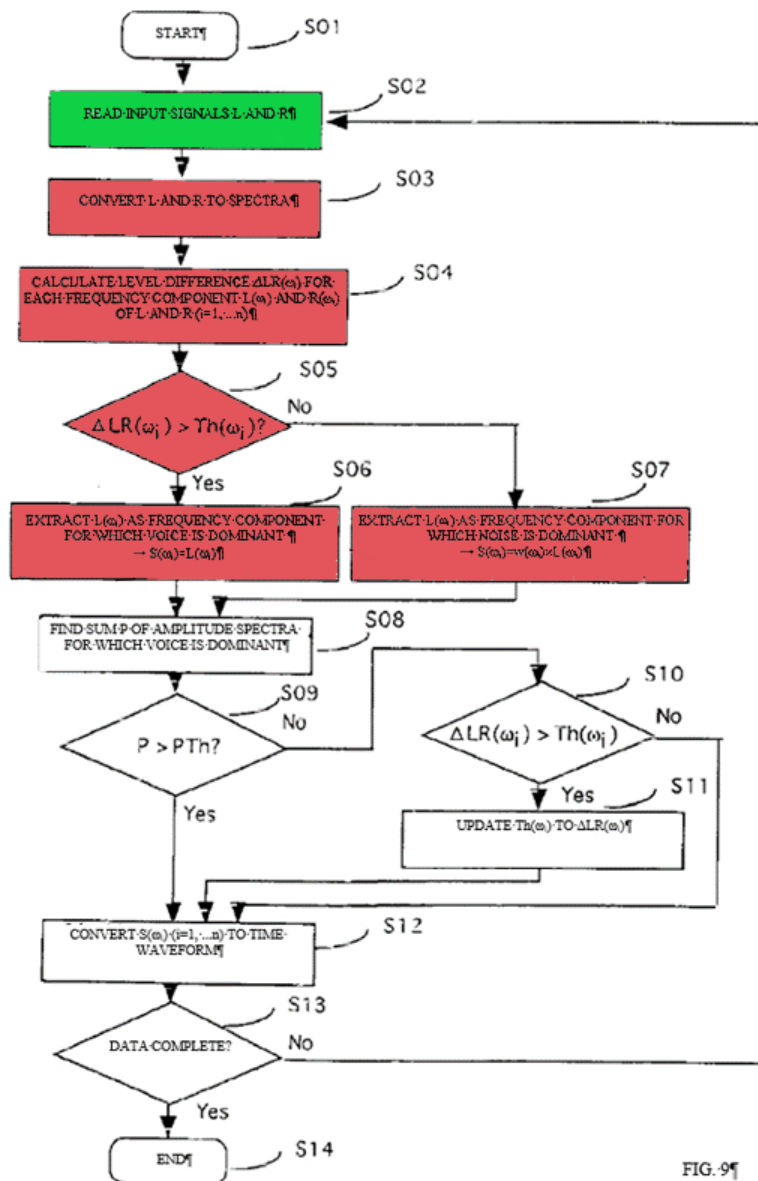


FIG. 9

(*Id.*, Fig. 9 (annotated).)

70. First, at **step S02**, Takano receives acoustic signals at its two microphones. (*Id.*, [0023] (“First the target signal wherein noise is superimposed is picked up by each of the microphones 1 and 2, and read in as digital signals[.]”); *see*

*also id.*, [0027], [0030] (explaining that overlapping steps in Figures 8 and 9 operate in the same manner as in Figure 7).)

71. Then, in **steps S03 and S04**, the system splits the signal into different portions based on frequency and calculates the “level difference” of the two signals for each frequency component. (*Id.* at [0023-0025]; *see also id.*, [0009], [0040]-[0042].) Takano discloses that “respective output signals of the first and second microphones are converted to amplitude spectra and phase spectra by first and second spectrum conversion means, and a level difference is calculated by a level difference calculation means for the amplitude spectrum of each frequency component outputted from the first and second spectrum conversion means.” (*Id.*, [0009].)

72. Next, in steps **S05, S06, and S07**, if the level difference exceeds a threshold, Takano determines if the “target signal,” i.e. speech, is “dominant.” If the level difference is below the threshold “the target signal is not dominant.” (*Id.* [0025]; *see also id.*, [0009]) (“This level difference for each frequency component is compared to a preset threshold value by a sound source frequency component selection means, and a frequency component for which the level difference is greater than the threshold value is determined by the sound source frequency component selection means to be a frequency component for which the target signal is dominant.”), [0040]-[0042].) Takano discloses:

In a voice frequency component selection part 6, a frequency for which the target signal is dominant is selected based on the magnitude correlation of  $\Delta LR(\omega_i)$  and a preset threshold value  $Th(\omega_i)$  for each frequency component. The conditions for determining whether the frequency is one for which the target signal is dominant are determined by the following formulas (S05).

$\Delta LR(\omega_i) > Th(\omega_i) \rightarrow$  the target signal is dominant

$\Delta LR(\omega_i) \leq Th(\omega_i) \rightarrow$  the target signal is not dominant

(*Id.* [0025]-[0026].) Thus, like the '058 patent, Takano discloses using two microphones to receive acoustic signals and then comparing the difference in the gain of the two microphones against a pre-set threshold to determine whether the signals are speech.

73. A person of ordinary skill in the art would have been motivated to combine the teachings of Holzrichter and Takano, and to add a second microphone to Holzrichter's system, because doing so would allow a person of ordinary skill in the art to replace Holzrichter's method of distinguishing speech from noise using a single microphone (e.g., steps 73 and 77) with Takano's improved method of distinguishing speech from noise using the level difference between two microphones (e.g., Takano steps S02-S07). Because Holzrichter already recognizes that the strength of acoustic signals (i.e. acoustic energy) could be used to determine whether a signal is speech or noise (Ex. 1003, 31:67-32:10, Fig. 19), Holzrichter provides the motivation to look for improved methods for classifying signals as noise

or speech based on signal strength, such as the system of Takano. A person of ordinary skill in the art would have found it obvious to add a second microphone from Takano to the system of Holzrichter because Takano's system for classifying noise or speech relies on the use of two microphones. Indeed, the methods described in Takano have many advantages.

74. First, Takano explains that its two-microphone system assists in overcoming background noise and is more capable of distinguishing between a user's speech and noise. (Ex. 1005, [0001] ("The present invention relates to a sound collection device for suppressing ambient noise components in a sound source signal containing ambient noise and extracting a target signal, a voice recognition device which has the sound collection function of the sound collection device and achieves a high recognition rate..."), [0008] (object if the invention is to have "a noise suppression processing function"), [0022]-[0029], [0059], Figs. 6A-6C.) This would be particularly advantageous to Holzrichter's system because it would allow Holzrichter's system to perform additional processing only on audio samples that are more likely to be speech, thereby saving computing resources.

75. Second, Takano discloses that using a second microphone provides better recovery of high-frequency components in voiced speech, thereby leading to better sound quality in the processed voice signal:

In the invention of the seventh mode, the microphone installed at a position near the sound source of the target signal and the microphone installed at a position farther from the sound source of the target signal in the sound collection device or the voice recognition device of the first through sixth modes are incorporated into a handset, a headset, or an ear microphone set so that the former microphone is situated at a position closer to the mouth than the latter microphone during use. Therefore, the noise resistance performance of transmission signals can be improved in a conventional handset, headset, or ear microphone set, and the voice recognition precision can be improved when using voice commands or the like in a handset. Conventionally, a bone conduction microphone and a receiver are formed integrally as an ear microphone set for obtaining a transmission signal with excellent sound resistance. However, the frequency components of a voice captured by a bone conduction microphone are biased toward low-frequency components, and there are few high-frequency components, so the sound quality is poor. There is also a problem with acoustic coupling between the bone conduction microphone and the receiver. However, in the invention of the seventh mode, sound is collected based on air conduction, and the distance between the receiver and the microphone can be secured, so an ear microphone set free of the problems described above can be provided.

(Ex. 1005, [0059].) A person of ordinary skill in the art would have been motivated to include this improvement in Holzrichter because it would allow Holzrichter's system to more accurately distinguish between speech and noise at high frequencies.

76. Third, the proposed combination represents nothing more than a predictable use of prior art elements according to established functions (i.e. using two microphones to detect speech).

77. Fourth, the combination represents nothing more than the simple addition of one known element (Takano's second microphone) to another known element (Holzrichter's system) to obtain predictable results. Indeed, Takano's second microphone would provide an improved method for something Holzrichter already contemplates, namely using the output strength of microphones to distinguish between noise and speech. Thus, the combination also reflects nothing more than a combination of known elements (microphones) to achieve an expected result (improved speech detection).

78. A person of ordinary skill in the art would have had reasonable expectation of success in combining Holzrichter and Takano because they relate to similar systems (systems for identifying acoustic signals such as speech or noise) and because the use of microphones to receive acoustic signals was well-known in the art.

**c. Voicing Sensor That Receives  
Physiological Information Associated  
With Human Voicing Activity**

79. Claim 1 further recites "at least one voicing sensor that receives physiological information associated with human voicing activity[.]" Holzrichter discloses this limitation.

80. Holzrichter discloses a radar sensor (EM sensor 68) that is used to detect voicing activity. (Ex. 1003, 32:31-35, Fig. 19, 9:23-54 (system uses "range

gated EM Sensor (i.e., radar)"); *see also id.*, 8:25-49, , 10:50-65, 11:66-12:11, 31:17-18, Figs. 4, 5, 8A-C, 12, 13, 24A-B.) For example, “tongue and jaw motion EM sensors as shown in FIG. 4 (modules 21 and/or 22) and in FIG. 19 (modules 68, 83) can measure precursor motions of the tongue and/or jaw motion as they move to a position in preparation for the first sound.” (*Id.*, 32:31-35.) Thus, Holzrichter discloses a voicing sensor that receives physiological information associated with human voicing activity as claimed.

**d. Processor Coupled Among the  
Microphones and the Voicing Sensor**

81. Claim 1 next recites “at least one processor coupled among the microphones and the voicing sensor[.]” It is my opinion that Holzrichter discloses this limitation.

82. Holzrichter discloses a processor coupled to microphone 70 and EM sensors 68. (Ex. 1003, 31:15-20, 18:49-67, 9:55-10:14, 18:24-48, Figs. 5, 13.) “The acoustic information from microphone 70 is inputted into an acoustic speech sensor, A/D converter and 1 sec. memory 71. Vocal fold motion information from EM sensor 69 is input into integrator and band pass filter 72. The processors make decisions as described below and as listed in Table I.” (*Id.*, 31:15-20.) It would have been obvious to a person of ordinary skill in the art that, when adding the second microphone of Takano to Holzrichter’s system, the microphone would be coupled to the processor to allow Holzrichter’s system to use the second microphone

to identify noise and speech. A person of ordinary skill in the art would be motivated to do this for the reasons discussed above. *Supra* § VIII.A.1.b.

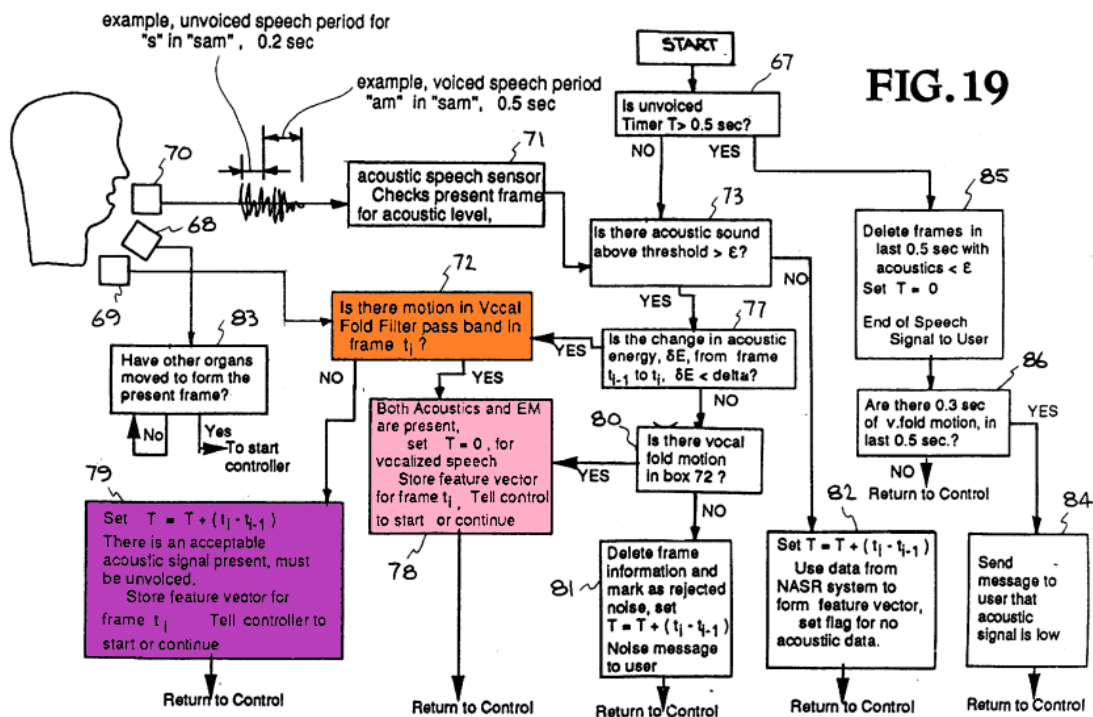
83. Claim 1 further recites that the processor performs five functions, each of which is addressed below.

**e. Generating Cross Correlation Data  
Between the Physiological Information  
and an Acoustic Signal**

84. Claim 1 recites that the processor “generates cross correlation data between the physiological information and an acoustic signal received at one of the two microphones[.]” It is my opinion that Holzrichter, alone or in combination with Kinsler, renders this limitation obvious.

85. Holzrichter discloses that its system determines whether both speech and vocal fold motion are present within a certain speech frame based on the physiological information received at EM sensor 69 and an acoustic signal received at microphone 70. (Ex. 1003, 31:61-32:27, Figs. 19, 15A-15D.) “[T]he algorithm looks for vocal fold motion to occur within the same speech time period in box 79 .... If acoustics and vocal fold motion are present, then a voiced speech unit has occurred .... If satisfactory acoustics, but no vocal fold motion occurs in box 72, the sound in frame  $t_i$  is labeled as nonvoiced speech in box 79.” (*Id.*, 31:61-32:27.) Figure 19 of Holzrichter illustrates this process.

86. In **step 72**, Holzrichter determines if vocal motion is occurring at the same time as the audio signal received by the microphone. (Ex. 1003, 31:61-32:27, Fig. 19.) If vocal motion is present, Holzrichter identifies the acoustic signal as voiced speech (**step 78**). (*Id.*) If vocal motion is not present and Holzrichter previously identified the acoustic signal as speech, then Holzrichter identifies the acoustic signal as unvoiced speech (**step 79**). (*Id.*)



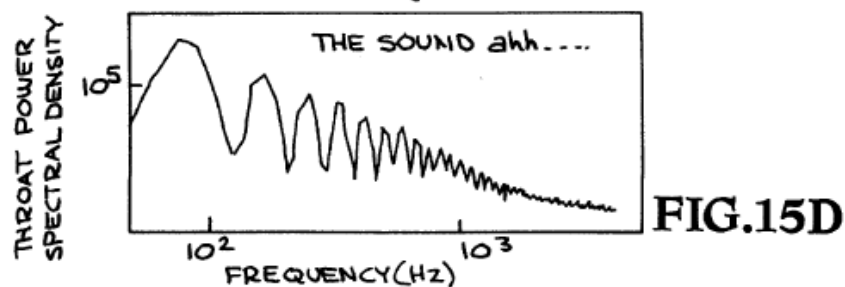
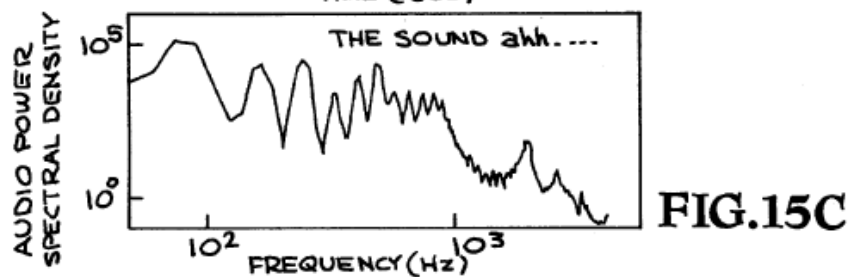
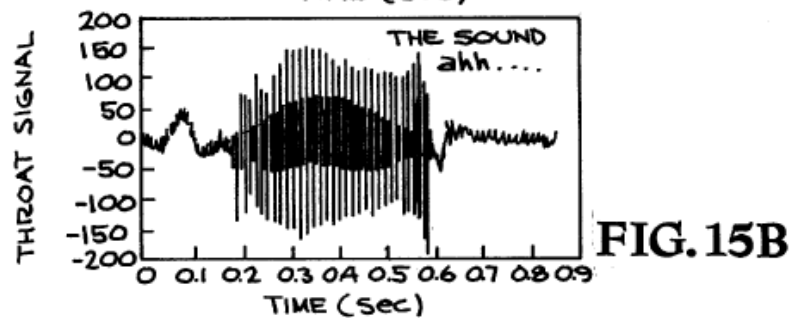
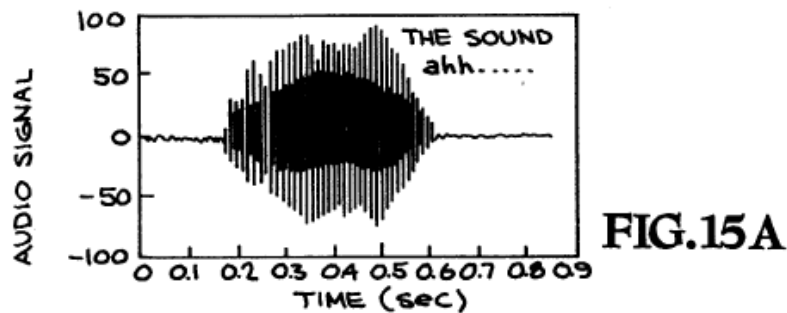
(*Id.*, Fig. 19 (annotated).)

87. Holzrichter further discloses that the signal it receives from the EM throat sensor and the audio signal received by the microphone are closely correlated. (Ex. 1003, 26:44-27:4 (explaining "how easily" speech organ location information

“can be correlated with simultaneous acoustic word signals from a microphone.”);

*see also id.*, 17:4-23, 20:11-34, 26:44-27:4, 31:61-32:27.) Holzrichter Figures 15A-

D show the correlation between the two signals:



(*Id.*, Fig. 15A-D.)

88. Although Holzrichter does not expressly state that the signals can be correlated using a cross correlation function, using such a function would have been obvious, particularly in view of Kinsler.

89. Kinsler discloses that “the only way to decide if a signal is present” is to calculate an output  $A_I$  from a “detection system.” This output is then compared against a “*threshold criterion*  $A_T$ ” such that “if  $A_I > A_T$  there is a signal present and if  $A_I < A_T$  there is no signal.” (Ex. 1004, 252.) Kinsler then explains various types of detection systems and their application. (*Id.*, 255-257.)

90. Kinsler discloses that the “optimum processing system” to search for a “known signal” (e.g., the audio signal from a microphone) in a “received signal” (e.g., the vocal movement signal from the EM sensor) is “cross correlation.” (Ex. 1004, 256.) To use the cross-correlation method, first, a value from the cross correlation function  $F(T)$  would be calculated. (*Id.*) Specifically, the maximum value of  $F(T)$  (cross correlation data) could be calculated and then can then be compared against the detection threshold (“DT”). (*Id.*, 252, 256-57; *see also id.*, 256 ( $F(T)$  “plays the role of the detection system output  $A$ ” for evaluating whether a signal is present).) If the converted value meets or exceeds the threshold, then the two signals are correlated (i.e., one signal is present in the other). (*Id.*, 252, 257.)

91. A person of ordinary skill in the art would have been motivated to combine the teachings of Holzrichter and Kinsler, and to use Kinsler’s “optimum

processing system” to calculate cross correlation data using a cross correlation function to determine if the vocal movements detected by Holzrichter’s EM sensor correlate with the acoustic signal received by Holzrichter’s microphone.

92. First, a person of ordinary skill in the art would have been motivated by Holzrichter’s disclosure that the vocal movements detected by Holzrichter’s EM sensor are highly correlated with speech. (Ex. 1003, 26:44-27:4, 17:4-23, 20:11-34, 26:44-27:4, 31:61-32:27).) For example, Holzrichter discloses:

The advantage of this sensor configuration is since there are so many waves, one wave is always reflecting off an organ interface at a range gate time that allows detection, thus simplifying the finding of the speech organ interfaces. This simple experiment also indicates how easily this speech organ location information can be correlated with simultaneous acoustic word signals from a microphone. The acoustic signals from a microphone sensing the same words as the EM Sensor(s) were displayed on a separate trace of the oscilloscope and were sampled in an A/D converter and were stored in a memory with identical start times and with identical time bin numbers as the memory bins for the EM sensor data.

(*Id.*, 26:53-65.) Based on this disclosure, a person of ordinary skill in the art would have been motivated to look for other references, such as Kinsler, for methods of determining when two signals are correlated with each other so that the system of Holzrichter can detect when the signal received by Holzrichter’s microphone is voiced speech.

93. Second, a person of ordinary skill in the art would have been motivated by Kinsler's disclosure that its method is the "optimum processing system" for determining if two signals are correlated to consider improving Holzrichter's system by correlating the Holzrichter's EM sensor and microphone signals using cross correlation. Indeed, the cross correlation function disclosed by Kinsler is a well-known method of determining whether signals are correlated. (Ex. 1015, 113-116.) A person of ordinary skill in the art would have understood that Kinsler's methods would have been a more accurate way of determining that two signals were correlated than Holzrichter's determination that the audio and vocal movement signals occurred at the same time. For example, using Kinsler's methods would prevent Holzrichter's system from identifying sound from Holzrichter's microphone that occurred at the same time as some non-voicing vocal movements (such as swallowing) as voice speech. Thus, a person of ordinary skill in the art would have used Kinsler's teachings to improve Holzrichter's method of correlating signals.

94. Third, the proposed combination represents nothing more than a predictable use of prior art elements according to established functions (e.g., determining if an audio signal and a vocal movement signal are correlated using a cross-correlation function). (*Id.*)

95. Fourth, the combination represents nothing more than the simple addition of one known element (Kinsler's cross-correlation methods) to another

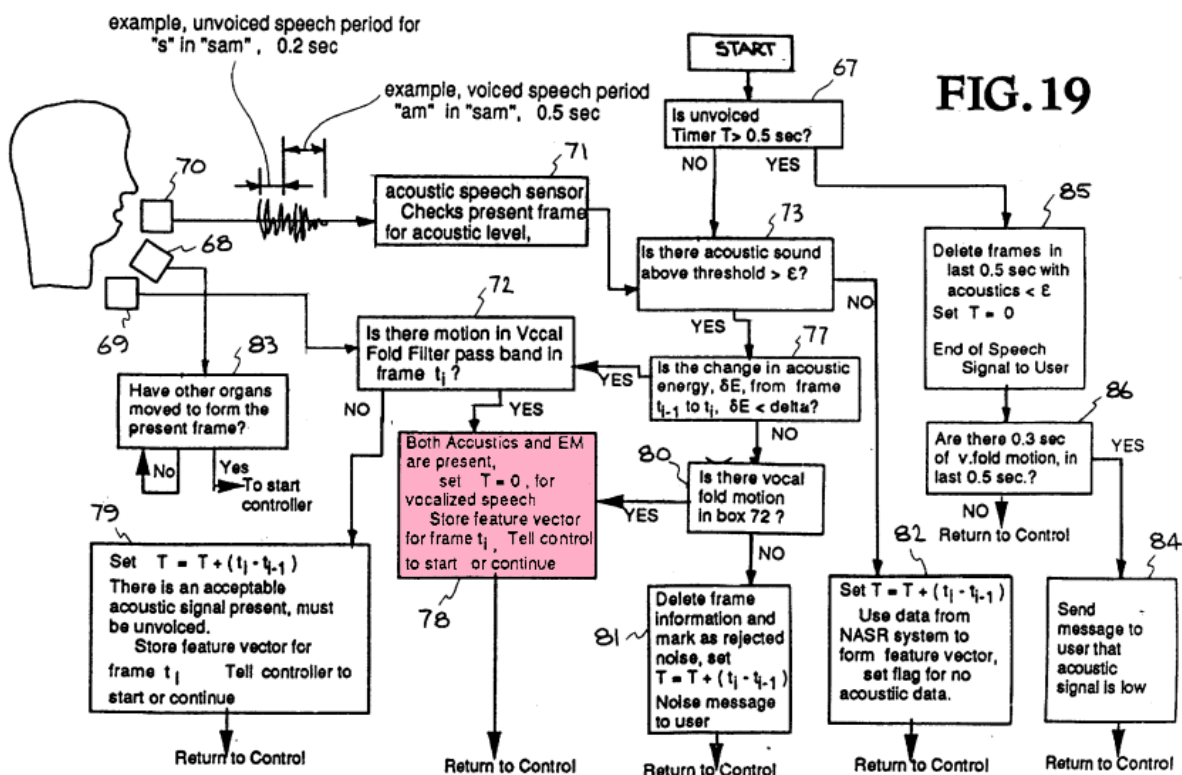
known element (Holzrichter's system) to obtain predictable results (identifying when an acoustic signal and a vocal movement signal are correlated). Indeed, the addition of Kinsler's methods to Holzrichter would simply improve what Holzrichter already contemplates—correlating an acoustic signal and a vocal movement signal. Thus, the combination also reflects nothing more than a combination of known elements to achieve an expected result.

96. A person of ordinary skill in the art would have had a reasonable expectation of success in combining Holzrichter and Kinsler because Holzrichter expressly recognizes that the signals could be “easily” correlated (Ex. 1003, 26:44-27:4) and because the cross-correlation function was widely known and used by those of skill in the art (Ex. 1013, 231-237). Additionally, processors, like those in Holzrichter, commonly perform mathematical functions like cross correlation on audio signals. Thus, a person of ordinary skill in the art would reasonable expect that Holzrichter's system could perform cross correlation.

**f. Identifying Information as Voiced Speech**

97. Claim 1 recites that the processor “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[.]” It is my opinion that Holzrichter combined with Kinsler discloses this limitation.

98. Holzrichter discloses that its system identifies information in the acoustic signal as voiced speech when it determines that the vocal motion signal is correlated to the acoustic signal. (Ex. 1003, 31:61-32:27, Figs. 19, 15A-15D.)



(*Id.*, Fig. 19 (annotated).) Kinsler discloses that the optimum way to determine whether two signals are correlated is to calculate the value of the cross correlation function  $F(T)$  (cross correlation data), specifically the maximum value, and then compare it against a threshold DT. (Ex. 1004, 252, 256-257.) If the value exceeds the threshold, then the two signals are correlated (i.e., one signal is present in the other). (*Id.*, 252, 257.)

99. Thus, Holzrichter combined with Kinsler discloses that identifying information in the acoustic signal as voiced speech when the cross correlation data (the value of  $F(t)$ ) exceeds a threshold (DT). A person of ordinary skill in the art would be motivated to combine Kinsler and Holzrichter, and to use Kinsler's cross correlation analysis, for the reasons discussed above.

**g. Generating Difference Parameters  
Between the Acoustic Signals Received at  
Each of the Two Receivers**

100. Claim 1 recites that the processor “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[.]” It is my opinion that Holzrichter combined with Takano discloses this limitation.

101. Takano discloses calculating a “level difference” (difference parameters) for each frequency component of the signals it receives at its two microphones (portions of the received acoustic signals). (Ex. 1005, [0009], [0023]-[0024], [0014]-[0015], [0029], [0040]-[0042] [0049]-[0051], Figs. 1, 6, 7, 16, 17, 18A.) This “level difference” is calculated based on the difference in “amplitude spectrum components” (i.e., gain) of the two signals. (*Id.*)

A first mode of the present invention is provided with a first microphone installed at a position close to a sound source of a target signal and a second microphone installed at a position farther from the sound source of the target signal

than the first position, wherein respective output signals of the first and second microphones are converted to amplitude spectra and phase spectra by first and second spectrum conversion means, and a level difference is calculated by a level difference calculation means for the amplitude spectrum of each frequency component outputted from the first and second spectrum conversion means. This level difference for each frequency component is compared to a preset threshold value by a sound source frequency component selection means, and a frequency component for which the level difference is greater than the threshold value is determined by the sound source frequency component selection means to be a frequency component for which the target signal is dominant. A frequency component for which it is assessed that the target signal is dominant is extracted by a sound source frequency component extraction means from the output amplitude spectrum of the first spectrum conversion means.

(*Id.*, [0009].) Thus, the level differences calculated by Takano are difference parameters that are representative of the relative difference in signal gain between portions of the received acoustic signals.

102. As explained above, *supra* § VIII.A.1.b., a person of ordinary skill in the art would have been motivated to incorporate the second microphone of Takano and Takano's associated methods into Holzrichter to allow Holzrichter to more accurately identify an acoustic signal as speech for further processing.

**h. Identifying Information as Unvoiced  
Speech When the Difference Parameters  
Exceed a Gain Threshold**

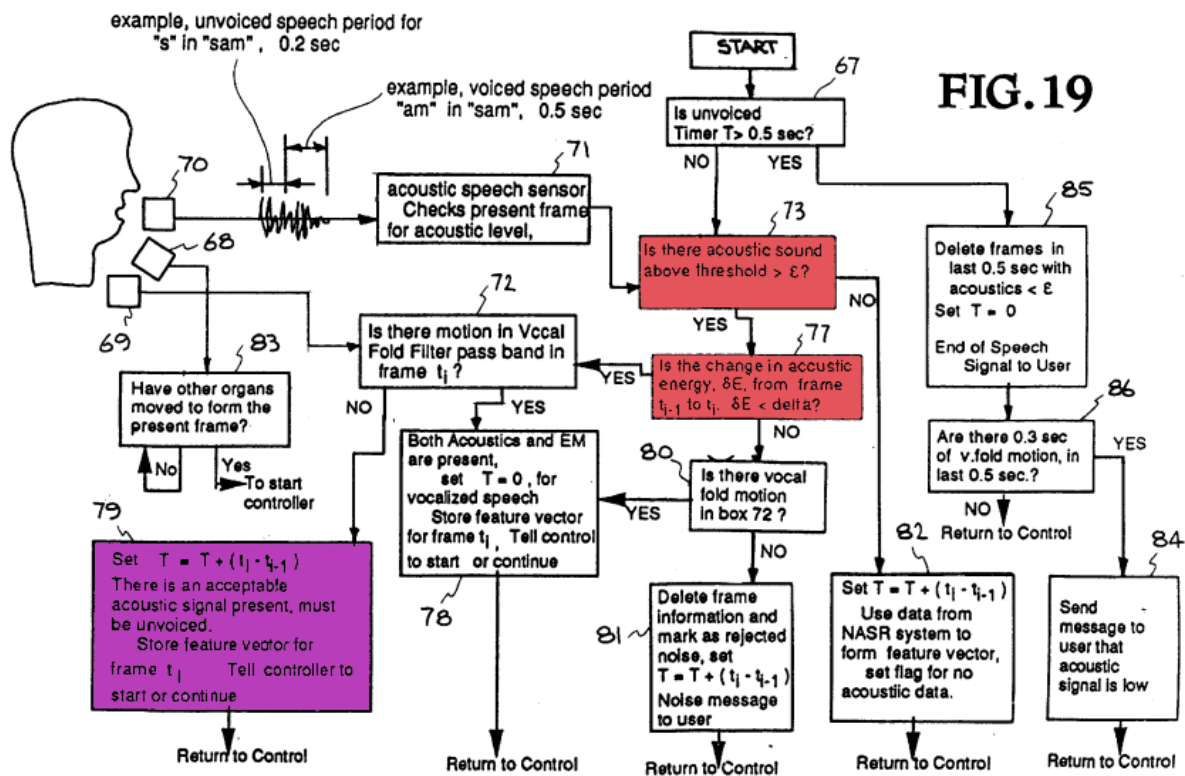
103. Claim 1 recites that the processor “identifies information of the acoustic signals as unvoiced speech when the difference parameters exceed a gain

threshold[.]” It is my opinion that Holzrichter combined with Takano discloses this limitation.

104. Holzrichter discloses that the strength of the output signal can be used to determine whether the signal is noise or speech. (Ex. 1003, 31:67-32, Fig. 19.) Specifically, Holzrichter identifies information of the acoustic signals as unvoiced speech when the strength of the received acoustic signal “exceeds a preset threshold 73” and no vocal motion is present. (*Id.*, 31:61-32:27.)

Speech onset testing is begun after acoustic and EM speech time frames and feature vectors are formed. For this example, each speech frame “i” is defined by its end time  $t_i$ . The algorithm FIG. 19 is entered at box 67 where a test on timer T for the cumulative duration of unvoiced speech over several speech time frames is conducted (e.g., is  $T > 0.5$  sec.?). If the duration of testing is shorter than 0.5 sec, then if the acoustic microphone 70 output signal exceeds a preset threshold 73, but when it is less than a loud noise threshold 77, then the algorithm looks for vocal fold motion to occur within the same speech time period in box 79. If acoustics and vocal fold motion are present, then a voiced speech unit has occurred and the unvoiced timer, T, is set to zero in box 78, and processing on frame  $t$  continues. If satisfactory acoustics, but no vocal fold motion occurs in box 72, the sound in frame  $t_i$  is labeled as nonvoiced speech in box 79 and processing continues. If the acoustic signal is too loud in box 77, and vocal fold motion is detected at box 80, the signal is processed as loud voiced speech box 78. If acoustic speech is below a threshold box 73 (a test not used for silent or whispered speech applications), the EM sensor feature vector is processed in box 82, the acoustic signal is labeled as too weak in box 82, the non-voiced or silence counter T is increased at box 82 by the time  $t_i - t_{i-1}$  of the most recent speech time frame, and the

(*Id.*) Holzrichter illustrates this process in Figure 19, where Holzrichter’s system identifies speech as unvoiced speech (**step 79**) if it determines the strength of the output signal from the microphone is adequate (**steps 73 and 77**).



-51-

105. As discussed above, Holzrichter does not disclose the use of a second microphone or the claimed difference parameters. But Takano does. Takano discloses that the “level difference  $\Delta LR(\omega_i)$ ” (difference parameter) is compared against a threshold “ $Th(\omega_i)$ .” (Ex. 1005, [0009], [0023]-[0025]; *id.*, [0014]-[0015], [0029], [0040]-[0042] [0049]-[0051], Figs. 1, 6, 7, 16, 17, 18A.) If the level difference exceeds the threshold, “the target signal is dominant,” meaning that the signal is identified as speech. (*Id.* [0025]; *id.*, [0014]-[0015], [0029], [0040]-[0042] [0049]-[0051], Figs. 1, 6, 7, 16, 17, 18A.) Takano’s Figure 9 shows this process:

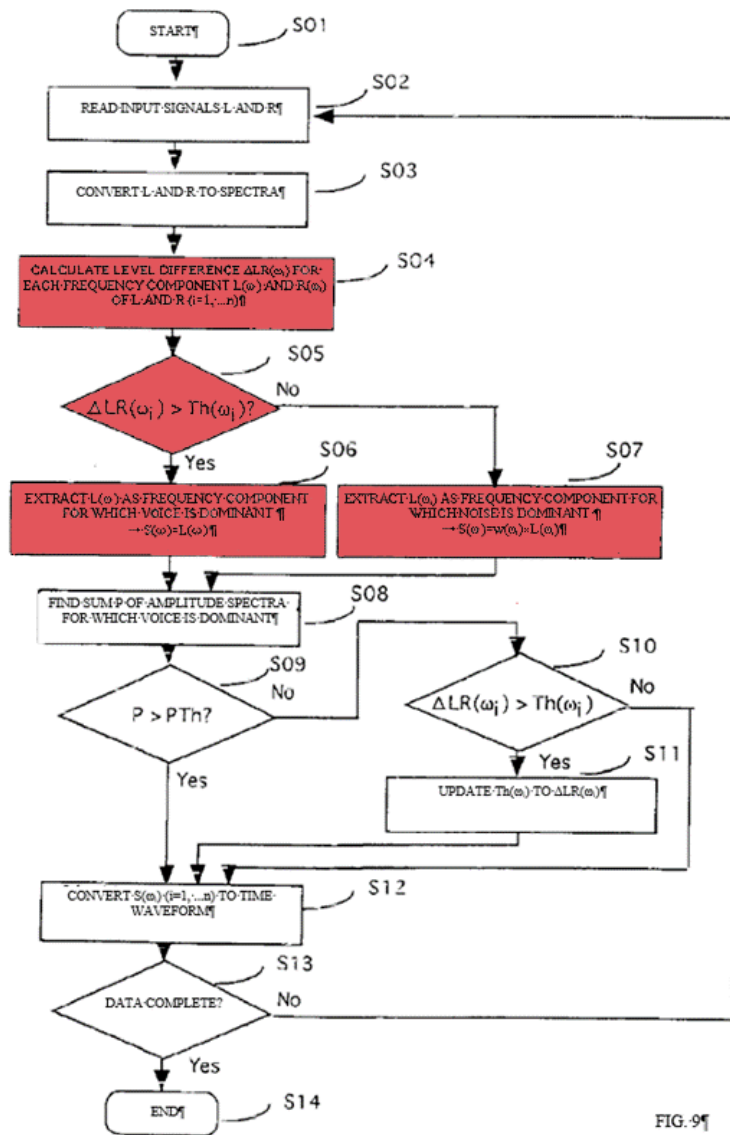


FIG. 9

(*Id.*, Fig. 9 (annotated).)

106. As explained above, *supra* § VIII.A.1.b., a person of ordinary skill in the art would have been motivated to use the two microphone system of Takano in place of the single microphone system of Holzrichter to determine if an acoustic signal is noise or speech. In such a combination, a person of ordinary skill in the art

would have replaced Holzrichter's evaluation of the output strength of a single microphone (e.g., Holzrichter steps 73 and 77) with Takano's improved evaluation of the level difference between two microphones (e.g., Takano steps S04-S07). The combined system would identify a signal as unvoiced speech only when the level difference of the two microphones exceeds a threshold. Thus, a person of ordinary skill in the art would have understood that the combined system would identify information of the acoustic signals as unvoiced speech when the difference parameters exceed a gain threshold.

**i. Identifying Information as Noise When  
the Difference Parameters Are Less Than  
the Gain Threshold**

107. Claim 1 recites that the processor "identifies information of the acoustic signals as noise when the difference parameters are less than the gain threshold." Holzrichter combined with Takano discloses this limitation.

108. Takano discloses that, when the level difference is lower than the threshold, "the target signal is not dominant," meaning that the signal is identified as noise instead of speech. (Ex. 1005, [0025], [0022] ("the level difference in the outputs from the microphones 1 and 2 due to noise is zero or minimal"), [0014]-[0015], [0029], [0040]-[0042] [0049]-[0051], Figs. 1, 6, 7, 16, 17, 18A.) As discussed above, *supra* § VIII.A.1.b., it would have been obvious to a person of

ordinary skill in the art to incorporate this feature of Takano into Holzrichter's system.

## **2. Claim 2**

109. Independent claim 2 is similar to claim 1, except it is a method claim. Thus, It is my opinion that claim 2 is obvious for the same reasons as claim 1.

## **3. Claim 3**

110. Claim 3 depends from claim 2 but adds the step of “generating the gain threshold using standard deviations corresponding to the generation of the difference parameters.” It is my opinion that Holzrichter combined with Takano discloses or renders obvious this limitation.

111. Takano teaches that its threshold “can be determined experimentally.” (Ex. 1005, [0026].) A person of ordinary skill in the art would have known that one way of determining the threshold empirically is to measure the “level difference” generated by Takano in various circumstances and then calculate the threshold based on standard deviations corresponding to those level differences. Standard deviation is one of a finite number of statistical methods (mean, variance, and standard deviation) for analyzing signals, like audio signals, that contain “random noise.” (Ex. 1008, 477.) Furthermore, standard deviation (or variance, which is the standard deviation squared) is no doubt the most common statistical measurement after the mean value. It is quite common to use standard deviation as a way of normalizing

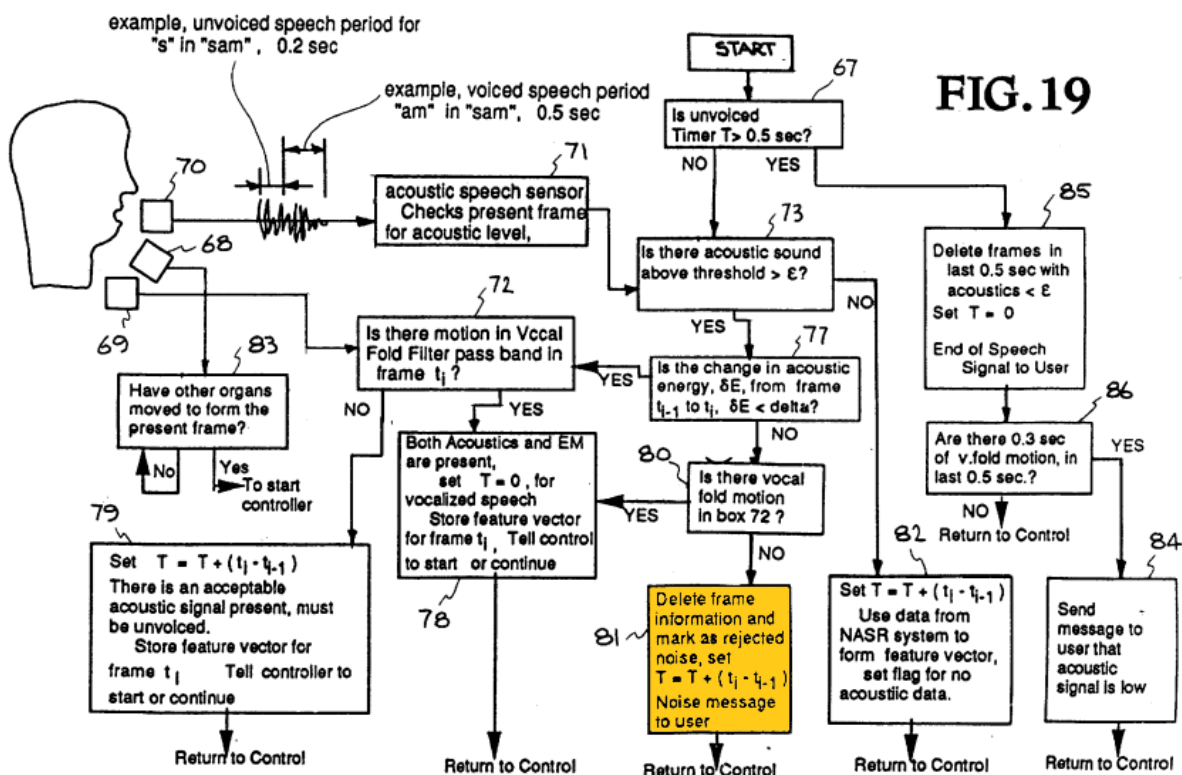
differences in mean values for a statistic like the difference in levels from the two microphones to normalize for differences in the operating environment. Indeed, the '058 patent recognizes that determining the threshold based on standard deviation is one of the “standard signal processing techniques” that could be used to determine the threshold. (Ex. 1001, 8:41-48; *see also* Ex. 1016, 2:22-34 (explaining that the standard deviation of the peak values of a signal is a statistical characteristic that can be used to determine if a signal is voiced or unvoiced).) Accordingly, this limitation would have been obvious to a person of ordinary skill in the art in view of Holzrichter and Takano.

#### 4. Claim 4

112. Claim 4 depends from claim 2 and adds the step of “performing denoising on the identified noise.” It is my opinion that Holzrichter discloses this limitation.

113. Holzrichter discloses a “method for determining the onset of speech, ***noise rejection***, voiced-unvoiced speech, end of speech, and pauses using acoustic and non-acoustic information together.” (Ex. 1003, 31:9-12 (emphasis added)). Holzrichter explains that, when its system identifies a portion of the acoustics signal as noise, that portion is removed and is not subsequently processed. (*Id.*, 33:39-34:7, Fig. 19; *id.*, 31:8-14, 33:20-34:7 (“The algorithm in FIG. 19 illustrated by boxes 67, 73, 77, 80, 81 detects the unusual noise event, removes the frame

information, and marks the frame as ‘damaged’ for subsequent processing.”), 50:59-51:9, Table 1.)



(Id., Fig. 19 (annotated).)

## 5. Claim 5

114. Claim 5 depends from claim 2 and recites that “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.” It is my opinion that Holzrichter discloses this limitation.

115. Holzrichter discloses that its EM sensor is a “radar” (radio frequency) sensor. (Ex. 1003, 8:1-24, claim 42 (“The apparatus of claim 41 wherein each EM wave generating, transmitting and receiving unit is a RF, microwave, millimeter wave, infrared, or visible wave radar.”), 9:23-54 (system uses “range gated EM sensor (i.e., radar)”)).)

**B. Ground 1b: Holzrichter, Takano, Kinsler, and Lyons**

116. It is my opinion that Holzrichter, Takano, Kinsler, and Lyons render claim 3 obvious.

**1. Claim 3**

117. Claim 3 recites the additional step of “generating the gain threshold using standard deviations corresponding to the generation of the difference parameters.”

118. As discussed above, Takano teaches that its threshold “can be determined empirically.” (Ex. 1005, [0026].) It is my opinion that, although it would have been obvious to a person of ordinary skill in the art to generate the gain threshold using standard deviations as discussed above, it also would have been obvious in further view of Lyons.

119. 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. 1008,

319, *see also id.*, 476.) Lyons teaches that the standard deviation of digital signals is defined as the positive square root of the variance:

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

(*Id.*, 320.) “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  $\sigma$  reflects the amount of the fluctuating, or AC, component of the signal[.]”

(*Id.*, 320, *see also id.*, 479.)

120. Lyons discloses that its digital signal processing has many applications, including “voice recognition [and] data communication” (*id.*, 1), and cites to more than 20 publications directed to digital signal processing used in speech recognition (*id.*, 21, 47, 127, 128, 155, 156, 217, 218, 295, 317, 318, 441, 442).

121. It would have been obvious to a person of ordinary skill in the art in view of Lyons that one way to determine the threshold empirically (as taught by Takano) would be to measure Takano’s “level difference” in various circumstances where Takano’s microphones are receiving noise. A person of ordinary skill in the art could then calculate the standard deviation (as taught by Lyons) of the different level differences, which would represent an upper bound for how high Takano’s level difference could reach when receiving noise. A person of ordinary skill in the

art would then use this upper bound as the threshold for Takano to determine when it is receiving noise instead of speech.

122. A person of ordinary skill in the art would have been motivated to generate the gain threshold in Takano using standard deviations as taught by Lyons for several reasons.

123. First, a person of ordinary skill in the art would have understood, based on Lyons, that calculating the gain threshold using standard deviations corresponding to generation of the difference parameters would ensure that the gain threshold is meaningful.

124. Second, 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 to minimize the effects of noise in measuring the true value of a parameter. (Ex. 1008, 319, *see also id.* 476.) A person of ordinary skill in the art would have understood that the standard deviation is applicable to the processing of speech signals disclosed in Takano because they are a type of digital signal. Indeed, it is quite common to use standard deviation as a way of normalizing differences in mean values for a statistic like the difference in levels from the two microphones to normalize for differences in the operating environment. In fact, the '058 patent recognizes that determining the threshold based on standard deviation is one of the “standard signal processing techniques” that could be used to determine the

threshold. (Ex. 1001, 8:41-48; *see also* Ex. 1016, 2:22-34 (explaining that the standard deviation of the peak values of a signal is a statistical characteristic that can be used to determine if a signal is voiced or unvoiced).)

125. Third, Lyons cites to over 20 publications related to speech signal processing, demonstrating that Lyons’ digital processing techniques, including the standard deviation calculation, are applicable to speech processing methods such as Takano’s. (Ex. 1008, 21, 47, 127, 128, 155, 156, 217, 218, 295, 317, 318, 441, 442.)

126. Fourth, a person of ordinary skill in the art would have been motivated to use, and would have had a reasonable expectation of success in using, standard deviation to calculate Takano’s threshold because doing so would have been combining prior art elements (difference parameters and a standard deviation) according to known methods (calculating standard deviations) to yield predictable results (standard deviation of the speech signals to generate a gain threshold).

127. Fifth, it would have been obvious to try using standard deviations to generate Takano’s gain threshold because standard deviation is one of a finite number of statistical methods (mean, variance, and standard deviation) for analyzing signals, like audio signals, that contain “random noise.” Indeed, the ’058 patent recognizes that determining the threshold based on standard deviation is one of the “standard signal processing techniques” that could be used to determine the threshold. (Ex. 1001, 8:41-48; *see also* Ex. 1016, 2:22-34 (explaining that the

standard deviation of the peak values of a signal is a statistical characteristic that can be used to determine if a signal is voiced or unvoiced).) Therefore, generating a gain threshold using standard deviations corresponding to the generation of the difference parameters would reflect merely a choice from a finite number of identified, predictable solutions, with a reasonable expectation of success.

**C. Ground 2a: Holzrichter, Torick, Kaneda, and Kinsler**

128. It is my opinion that Holzrichter, Torick, Kaneda, and Kinsler render claims 1, 2, 4, and 5 obvious.

**1. Claim 1**

129. It is my opinion that the preamble, the voicing sensor, processor, and first two processor functionality limitations (generating cross correlation data and identifying information as voiced speech when the cross correlation data exceeds a correlation threshold) of claim 1 are disclosed by the combination of Holzrichter and Kinsler as discussed above. The remaining limitations of claim 1 are discussed below.

**a. At Least Two Microphones**

130. Claim 1 recites “at least two microphones that receive the acoustic signals[.]” It is my opinion that Holzrichter combined with Torick discloses this limitation.

131. Holzrichter discloses using the strength of the output signal of a microphone to determine if a signal is speech or noise. (*Supra* VIII.A.1.h.) Torick discloses a system for detecting “speech sounds” using two microphones. (Ex. 1006, 169.) One microphone is used to detect speech, while the other is “used to sense the ambient noise.” (*Id.*)

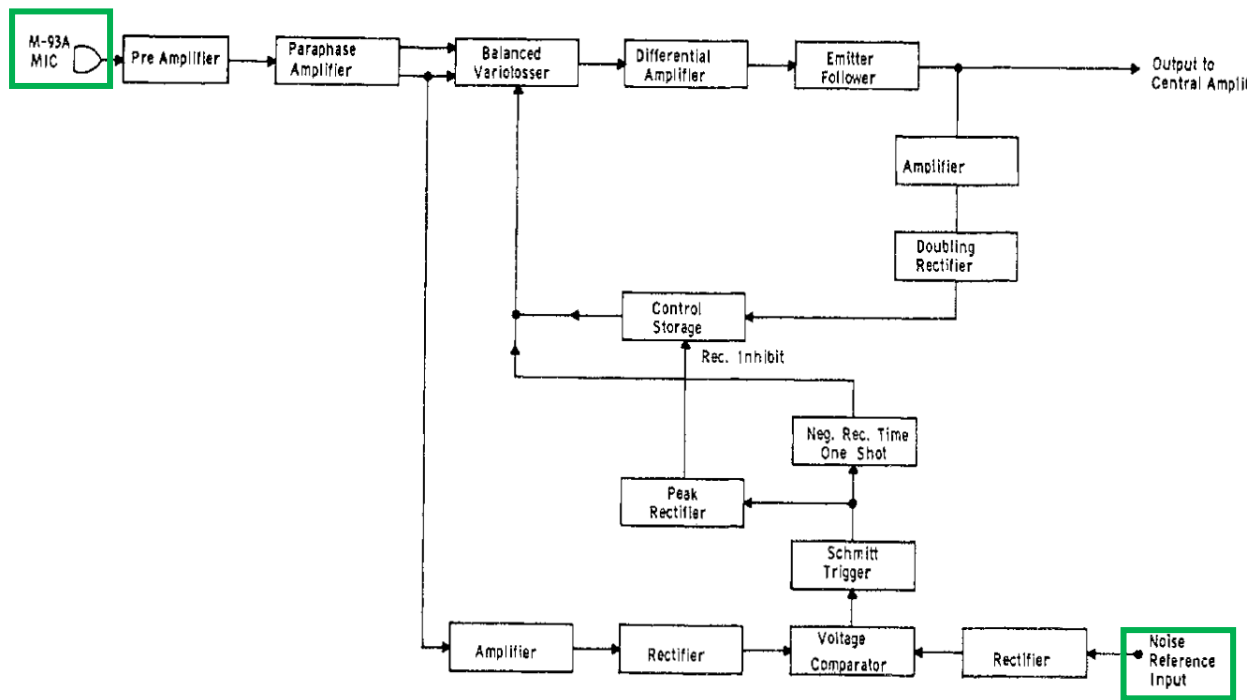


Fig. 4. Crew member's amplifier and control unit.

(*Id.*, Fig. 4 (annotated).)

132. Torick uses these two microphones to implement “voice actuated switching” and “eliminate the press-to-talk switch” for microphone-based systems. (*Id.*, 168). To accomplish this, Torick explains that “[t]he noise-sensing microphone output is amplified and equalized to correspond to the far-field response of the noise-cancelling speech microphone. The outputs of both microphones are then

continuously compared.” (*Id.*, 169.) When the signals from the two microphones are equal, the system determines that no speech is occurring. (*Id.*) However, when speech sounds occur at the speech microphone, the level between the two microphones is no longer equal and speech is detected. (*Id.*) Torick explains that “the method is analogous to the presence of a virtual threshold which rises and falls with corresponding changes in noise levels.” (*Id.*) Torick’s system allows for voice-actuated switching even in “high noise environments.” (*Id.*)

133. A person of ordinary skill in the art would have understood that the second microphone of Torick could be added to Holzrichter’s system so that a person of ordinary skill in the art could replace Holzrichter’s method of distinguishing speech from noise using a single microphone (*e.g.* steps 73 and 77) with Torick’s improved method of distinguishing speech from noise by comparing the outputs two microphones.

134. A person of ordinary skill in the art would have been motivated to combine the teachings of Holzrichter and Torick, and to add a second microphone to Holzrichter’s system, for several reasons.

135. First, because Holzrichter recognizes the strength of acoustic signals could be used to determine noise or speech (Ex. 1003, 31:67-32:4), a person of ordinary skill in the art reading Holzrichter would have been motivated to look for

other methods for classifying signals as noise or speech based on signal strength.

Torick discloses such as a system.

136. Second, adding a second microphone as taught by Torick would have improved the system's performance because Torick's two-microphone system allows for the accurate identification of speech even in "high noise environments." (Ex. 1006, 169.) This would be particularly advantageous to the Holzrichter system because it would allow Holzrichter's system to perform additional processing only on audio samples that are more likely to be speech, thereby saving computing resources.

137. Third, the proposed combination represents nothing more than a predictable use of prior art elements according to established functions (i.e., using two microphones to detect speech).

138. Fourth, the combination represents nothing more than the simple addition of a known element (Torick's second microphone) to another known element (Holzrichter's system) to obtain predictable results. Indeed, adding Torick's second microphone would enable Holzrichter's system to improve a function that it already contemplates (using the output strength of microphones to distinguish between noise and speech). Thus, the combination also reflects nothing more than a combination of known elements (microphone) to achieve an expected result (improved speech detection).

139. A person of ordinary skill in the art would have had a reasonable expectation of success in combining Holzrichter and Torick because they relate to similar systems (systems for identifying acoustic signals as speech or noise) and because the use of microphones to receive acoustic signals was well known in the art.

**b. Generating Difference Parameters  
Between the Acoustic Signals Received at  
Each of the Two Receivers**

140. It is my opinion that Holzrichter combined with Torick in light of Kaneda discloses this limitation. As explained above, a person of ordinary skill in the art would have been motivated to incorporate the second microphone of Torick and Torick's associated methods into Holzrichter to allow Holzrichter's system to more accurately identify an acoustic signal as speech for further processing.

141. Torick discloses that its "noise-sensing microphone output is amplified and equalized to correspond to the far-field response of the noise-cancelling speech microphone." (Ex. 1006, 169.) "The outputs of both microphones are then continuously compared." (*Id.*) When the signals from the two microphones are equal, the system determines that no speech is occurring. (*Id.*) However, when speech sounds occur at the speech microphone, the level between the two microphones is no longer equal and speech is detected. (*Id.*)

142. It would have been obvious to a person of ordinary skill in the art reading Torick that, in order for the system to function properly, there should be some tolerance or range within which Torick would consider the outputs of two microphones to be equal and outside of which Torick would consider the outputs to not be equal, thereby detecting speech. Thus, a person of ordinary skill in the art would look for methods to compare the outputs of the two signals and determine whether the two signals are within a certain range of each other.

143. Kaneda teaches that one way to compare the outputs of two microphones is to calculate a “power difference PD” (difference parameters) by subtracting the power, or gain, of one microphone from the other. (Ex. 1007, 2:1-22, Fig. 3(a)-3(c).) Thus, Torick in light of Kaneda discloses 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.

144. It would have been obvious to compare the outputs of Torick’s microphones in the manner taught by Kaneda because subtracting one signal from another is a standard technique for comparing two signals. Further, a person of ordinary skill in the art would have been motivated to combine Kaneda’s method with Torick because it would allow a person of ordinary skill in the art to add sufficient tolerance to Torick’s system to allow the system to operate. Indeed, the

proposed modification represents nothing more than the simple addition of one known element (Kaneda's subtraction) to another known element (Torick's microphones) to obtain predictable the predictable result of comparing the two signals from Torick's microphones.

**c. Identifying Information as Unvoiced  
Speech When the Difference Parameters  
Exceed a Gain Threshold**

145. It is my opinion that Holzrichter combined with Torick in light of Kaneda discloses this limitation.

146. Torick further discloses that it “detects speech sounds” when the output of the speech microphone is greater than the noise sensing microphone, which is “analogous to the presence of a virtual threshold which rises and falls with corresponding changes in noise levels.” (Ex. 1006, 169.) Kaneda discloses identifying a portion of a signal as speech when “[w]hen the power difference PD is larger than a given threshold value  $P_{th}$ .” (Ex. 1007, 2:13-18.) As discussed, it would be obvious to use the mathematical methods disclosed in Kaneda to detect speech using the microphones of Torick.

147. Further, as explained previously, Holzrichter discloses that, when its system identifies a signal as speech, it further analyzes the signal to determine if the acoustic signal is voiced or unvoiced speech. Holzrichter discloses that it will only identify the signal as unvoiced speech when no vocal motion is present **and** the

output of signal is sufficient to identify the signal as speech. (Ex. 1003, 31:61-32:27.)

148. As explained above, *supra* VIII.C.1.a., a person of ordinary skill in the art would have been motivated to use Torick's two-microphone system in place of Holzrichter's single-microphone system to determine if an acoustic signal is noise or speech. In such a combination, a person of ordinary skill in the art would replace Holzrichter's use of an output signal from a single microphone to determine if a signal is noise (e.g., Holzrichter steps 73 and 77) with Torick's improved evaluation of the difference in the outputs between two microphones to determine if a signal is noise. Accordingly, the combined system would identify a signal as unvoiced speech only when there is no vocal motion present **and** when the difference in outputs between Torick's two microphones exceeds a threshold. Thus, a person of ordinary skill in the art would understand that the combined system identifies information of the acoustic signals as unvoiced speech when the difference parameters exceed a gain threshold.

**d. Identifying Information as Noise When  
the Difference Parameters Are Less Than  
the Gain Threshold**

149. It is my opinion that Holzrichter combined with Torick in light of Kaneda discloses this limitation.

150. Torick further discloses that it does not “detect[] speech sounds” (i.e., it detects noise) when the output of the speech microphone is the same as the noise sensing microphone. (Ex. 1006, 169.) Kaneda similarly discloses identify a portion of a signal as noise when the power difference PD is lower than a threshold. (See Ex. 1007, 1:51-68, 2:1-22, Fig. 3(a)-3(c).) As discussed, it would be obvious to use the mathematical methods disclosed in Kaneda to detect speech using the microphones of Torick.

151. Further, as discussed above, *supra* VIII.C.1.a., it would have been obvious to a person of ordinary skill in the art to incorporate this feature of Torick into the system of Holzrichter. Indeed, Holzrichter similarly discloses identifying a signal as noise when it “exceeds a preset threshold 73, but when it is less than a loud noise threshold 77.” (Ex. 1003, 31:61-32:27.) Based on these disclosures, a person of ordinary skill in the art would have found it obvious that that the system could identify acoustic signals as noise when the difference parameters are less than the gain threshold.

## **2. Claim 2**

152. Claim 2 is a method claim that is substantially similar to the system claimed in claim 1. It is my opinion that the combination of Holzrichter, Torick, and Kinsler would have rendered this claim obvious for the same reasons as claim 1.

**3. Claim 4**

153. As discussed above, *supra* § VIII.A.4., it is my opinion that Holzrichter discloses the additional limitations of claim 4.

**4. Claim 5**

154. As discussed above, *supra* § VIII.A.5., it is my opinion that Holzrichter discloses the additional limitations of claim 5.

**D. Ground 2b: Holzrichter, Torick, Kaneda, Kinsler, and Lyons**

155. It is my opinion that Holzrichter, Torick, Kaneda, Kinsler, and Lyons render claim 3 obvious.

**1. Claim 3**

156. Claim 3 recites the additional step of “generating the gain threshold using standard deviations corresponding to the generation of the difference parameters.”

157. As explained, *supra* § VIII.B.1, Lyons teaches that the standard deviation of a signal “is an indication of the magnitude of the fluctuating component of the signal.” (Ex. 1008, 479). Thus, it would have been obvious to a person of ordinary skill in the art in view of Lyons that one way to determine the threshold taught by Torick and Kaneda would be to measure the power differences received by Torick’s microphones in various circumstances when Torick is receiving noise. A person of ordinary skill in the art could then calculate the standard deviation of

the different power differences, which would represent an upper bound for how high the power difference could reach when receiving noise. A person of ordinary skill in the art would then use this upper bound as the threshold for Takano to determine when it is receiving noise instead of speech. A person of ordinary skill in the art would have been motivated to do this for the same reasons as discussed.

## **IX. SECONDARY CONSIDERATIONS**

158. I am not aware of any secondary considerations of non-obviousness. If Patent Owner identifies any alleged evidence of secondary considerations in the future, I reserve the right to respond to that information.

## **X. CONCLUSION**

159. For the foregoing reasons, 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 at the time of the alleged invention in view of the prior art discussed above.

160. I reserve the right to supplement my opinions in the future to address or respond to any issues that the Patent Owner may raise, as well as new information including, but not limited to, any claim constructions advanced by the Patent Owner or adopted by the Board in the Institution Decision, and respond to any alleged secondary considerations as they become available to me.

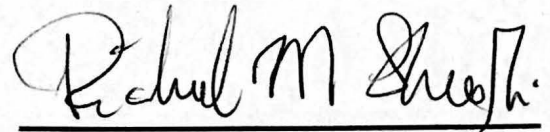
161. I declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and

Amazon.com v. Jawbone LLC

Declaration of Dr. Richard M. Stern – U.S. Patent No. 7,246,058

further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code.

Executed on November 13, 2022 at Pittsburgh, Pennsylvania.

A handwritten signature in black ink, reading "Richard M. Stern, Ph.D.", written over a horizontal line.

Richard M. Stern, Ph.D.

56605603