

Filed: June 23, 2023

IPR2023-01119
U.S. Patent No. 7,246,058
Petition for *Inter Partes* Review
Attorney Docket SONY 7.1R-019

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

SONY ELECTRONICS INC.,
Petitioner(s),

v.

JAWBONE INNOVATIONS, LLC,
Patent Owner.

IPR2023-01119
U.S. Patent No. 7,246,058

**PETITION FOR *INTER PARTES* REVIEW OF
UNITED STATES PATENT NO. 7,246,058
PURSUANT TO 35 U.S.C. §§ 311-319, 37 C.F.R. § 42**

TABLE OF CONTENTS

	Page
TABLE OF AUTHORITIES.....	v
EXHIBIT LIST.....	vii
I. INTRODUCTION.....	1
II. BACKGROUND.....	2
A. Using Microphones and Tissue Sensors Together to Detect Voiced and Unvoiced Speech Was Known.....	2
B. Using Cross Correlation to Determine Whether a Desired Signal Is Present Was Known.....	5
C. Using Two Microphones and Differences in Signal Strength to Detect Speech Was Known.....	6
III. THE '058 PATENT	11
A. Overview	11
B. Prosecution History.....	15
IV. STATEMENT OF RELIEF REQUESTED	16
A. Grounds.....	16
B. The References Are Prior Art.....	16
C. The Asserted References Are Analogous Art.....	17
V. LEVEL OF ORDINARY SKILL	17
VI. CLAIM CONSTRUCTION	18
VII. GROUNDS FOR UNPATENTABILITY	19
A. Ground 1a: Holzrichter, Takano, and Kinsler	19
1.Claim 1	19

a.	Preamble	19
b.	At Least Two Microphones	19
c.	Voicing Sensor That Receives Physiological Information Associated With Human Voicing Activity	28
d.	Processor Coupled Among the Microphones and the Voicing Sensor	29
i.	Generating Cross Correlation Data Between the Physiological Information and an Acoustic Signal	30
ii.	Identifying Information as Voiced Speech	36
iii.	Generating Difference Parameters Between the Acoustic Signals Received at Each of the Two Receivers	38
iv.	Identifying Information as Unvoiced Speech When the Difference Parameters Exceed a Gain Threshold	39
v.	Identifying Information as Noise When the Difference Parameters Are Less Than the Gain Threshold	42
2.	CLAIM 2	43
3.	CLAIM 3	43
4.	CLAIM 4	44
5.	CLAIM 5	45
B.	Ground 1b: Holzrichter, Takano, Kinsler, and Lyons	46
1.	Claim 3	46
C.	Ground 2a: Holzrichter, Torick, Kaneda, and Kinsler	50

1. Claim 1	50
a. At Least Two Microphones.....	50
b. Generating Difference Parameters Between the Acoustic Signals Received at Each of the Two Receivers	54
c. Identifying Information as Unvoiced Speech When the Difference Parameters Exceed a Gain Threshold	56
d. Identifying Information as Noise When the Difference Parameters Are Less Than the Gain Threshold	58
2. Claim 2.....	59
3. Claim 4.....	59
4. Claim 5	59
D. Ground 2b: Holzrichter, Torick, Kaneda, Kinsler, and Lyons	59
1. Claim 3	59
VIII. SECONDARY CONSIDERATIONS OF NONOBVIOUSNESS.....	60
IX. DISCRETIONARY DENIAL UNDER § 314(A) IS NOT APPROPRIATE	60
A. Factor 1: Potential Stay	61
B. Factor 2: Proximity of Trial to FWD	62
C. Factor 3: Investment in Parallel Proceeding.....	63
D. Factor 4: Overlapping Issues	64
E. Factor 5: The Parties	64

F.	Factor 6: Other Circumstances	65
X.	DISCRETIONARY DENIAL UNDER § 325(D) IS NOT APPROPRIATE	65
XI.	MANDATORY NOTICES	67
A.	Real Party-In-Interest (37 C.F.R. § 42.8(b)(1)).....	67
B.	Related Matters (37 C.F.R. § 42.8(b)(2))	68
C.	Lead and Backup Counsel (37 C.F.R. § 42.8(b)(3)).....	69
D.	Service Information (37 C.F.R. § 42.8(b)(4)).....	69
E.	Payment of Fees Pursuant to 37 C.F.R. § 42.103.....	69
F.	Grounds for Standing Under 37 C.F.R. § 42.104(a).....	70
XII.	CONCLUSION	70

TABLE OF AUTHORITIES

	Page(s)
CASES	
<i>Apple v. Fintiv</i> , IPR2020-00019, Paper 11 (P.T.A.B. Mar. 20, 2020)	60
<i>Cal. Inst. of Tech. v. Broadcom Ltd.</i> , 25 F.4th 976 (Fed. Cir. 2022).....	64
<i>GAF Materials LLC v. Kirsch Rsch. & Dev., LLC</i> , IPR2021-00192, Paper 14 (P.T.A.B. May 25, 2021).....	62
<i>Global Tel-Link Corp. v. HLFIP Holding, Inc.</i> , IPR2021-00444, Paper 14 (P.T.A.B. July 22, 2021)	62
<i>Google LLC v. Jawbone Innovations LLC</i> , IPR2022-00630, Paper 10 (P.T.A.B. Sept. 13, 2022).....	64
<i>In re GPAC Inc.</i> , 57 F.3d 1573 (Fed. Cir. 1995).....	17
<i>Haliburton Energy Servs., Inc. v. U.S. Well Servs., LLC</i> , IPR2021-01036, Paper 12 (Jan. 19, 2022)	66
<i>Huawei Techs. Co. v. WSOU Invs., LLC</i> , IPR2021-00226, Paper 10 (P.T.A.B. June 10, 2021).....	64
<i>Jawbone Innovations, LLC v. Sony Elecs. Inc.</i> , No. 2:23-cv-01161-CCC-LDW (D.N.J.).....	61
<i>Kirsch Rsch. & Dev., LLC v. GAF Materials, LLC</i> , Civil Action No. 20-13683 (JMV), 2021 U.S. Dist. LEXIS 111667 (D.N.J. June 15, 2021)	62
<i>Leapfrog Enters. v. Fisher-Price, Inc.</i> , 485 F.3d 1157 (Fed. Cir. 2007).....	60

IPR2023-01119

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

<i>Newell Cos. v. Kenney Mfg. Co.</i> , 864 F.2d 757 (Fed. Cir. 1988).....	60
<i>Nidec Motor Corp. v. Zhongshan Broad Ocean Motor Co.</i> , 868 F.3d 1013 (Fed. Cir. 2017).....	18
<i>PNC Bank, N.A., v. United Servs. Auto. Ass’n</i> , IPR2022-00075, Paper 21 (June 13, 2022).....	66
<i>Qualcomm Inc. v. UNM Rainforest Innovations</i> , IPR2021-00582, Paper 11 (P.T.A.B. Aug. 27, 2021)	63
<i>Samsung Elecs. Am. Inc. v. Snik LLC</i> , IPR2020-01428, Paper 10 (P.T.A.B. Mar. 9, 2021)	63
<i>Sand Revolution II, LLC v. Cont’l Intermodal Grp.-Trucking LLC</i> , IPR2019-01393, Paper 24 (P.T.A.B. June 16, 2020).....	64
<i>Skechers U.S.A., Inc. v. Nike, Inc.</i> , IPR2021-00160, Paper 10 (P.T.A.B. May 19, 2021).....	62
<i>Unwired Planet, LLC v. Google Inc.</i> , 841 F.3d 995 (Fed. Cir. 2016).....	17
<i>Vivid Techs., Inc. v. Am. Sci. & Eng’g, Inc.</i> , 200 F.3d 795 (Fed. Cir. 1999).....	18

STATUTES, RULES & OTHER AUTHORITIES

35 U.S.C. § 102	16
35 U.S.C. § 103	16
35 U.S.C. § 112	18
35 U.S.C. § 314(a).....	60, 65
35 U.S.C. § 325(d).....	65, 66

TABLE OF EXHIBITS

Exhibit No.	Description
1001	U.S. Patent No. 7,246,058 (“the ’058 patent”)
1002	Declaration of Richard Stern, Ph.D.
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 <i>IEEE Transactions on Audio and Electroacoustics</i> 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

Exhibit No.	Description
1013	Excerpts from Michael J. Chapman <i>et al.</i> SIGNAL PROCESSING IN ELECTRONIC COMMUNICATIONS (Horwood Publishing Limited 1997)
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
1017	Curriculum Vitae of Richard Stern, Ph.D.
1018	Japanese Patent Application Publication No. H-11-305792
1019	Declaration of Carol Peterson Regarding Kinsler, Lyons, and Torick
1020	Declaration of Brian Toth Regarding English Translation of Japanese Patent App. Pub. No. H-11-305792

Petitioner Sony Electronics Inc. (“Sony” or “Petitioner”) request *inter partes* review (“IPR”) of Claims 1-5 of U.S. Patent No. 7,246,058 (“the ’058 patent”), which Jawbone Innovations, LLC (“Patent Owner” or “PO”) purportedly owns.

I. INTRODUCTION

Voiced speech—speech sounds produced using vocal fold vibration—and unvoiced speech—speech sounds made without vocal fold vibrations—are well known categories of human speech. (Ex. 1009.) The challenged claims recite a system for identifying voiced and unvoiced speech using conventional components that were widely known in the art: a “voicing sensor” and two microphones.

In 1996, Professor John F. Holzrichter 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. (Ex. 1003.) That patent issued in 1999 and is assigned to the Regents of the University of California. (*Id.*)

Gregory C. Burnett (“Burnett”) worked under Dr. Holzrichter at the Lawrence Livermore National Laboratory for years on research related to the voicing sensor invented by Dr. Holzrichter. (Ex. 1010; Ex. 1011 (listing Holzrichter and Burnett as co-inventors); Ex. 1012 (same).) Burnett even thanked “the efforts and enthusiasm of John Holzrichter” for “tirelessly [] promot[ing] this new sensor and its possible applications” and suggesting “several experiments” in his Ph.D. dissertation.

(Ex. 1010, v.)

In 2001, after leaving Dr. Holzrichter's lab, Burnett filed the '058 patent, which describes nothing more than known uses of the very sensor that Dr. Holzrichter developed. (Ex. 1001.) The '058 patent, which identifies Burnett as the sole inventor, describes trivial changes to the system described in Dr. Holzrichter's patent. For example, Burnett's '058 patent describes using two microphones instead of one microphone and processing data using the conventional mathematical technique of cross-correlation. But two-microphone systems for detecting speech were disclosed as early as 1966 and cross-correlation for audio processing was described as early as 1982. (Ex. 1006, 169, Ex. 1004, 252, 257; Ex. 1013, 231-237.) Thus, Burnett's '058 patent merely combines Dr. Holzrichter's known sensor with well-known signal processing features in a conventional manner. The challenged claims were obvious to a person skilled in the art in view of Holzrichter's teachings, and the Board should cancel the claims.

II. BACKGROUND

A. Using Microphones and Tissue Sensors Together to Detect Voiced and Unvoiced Speech Was Known.

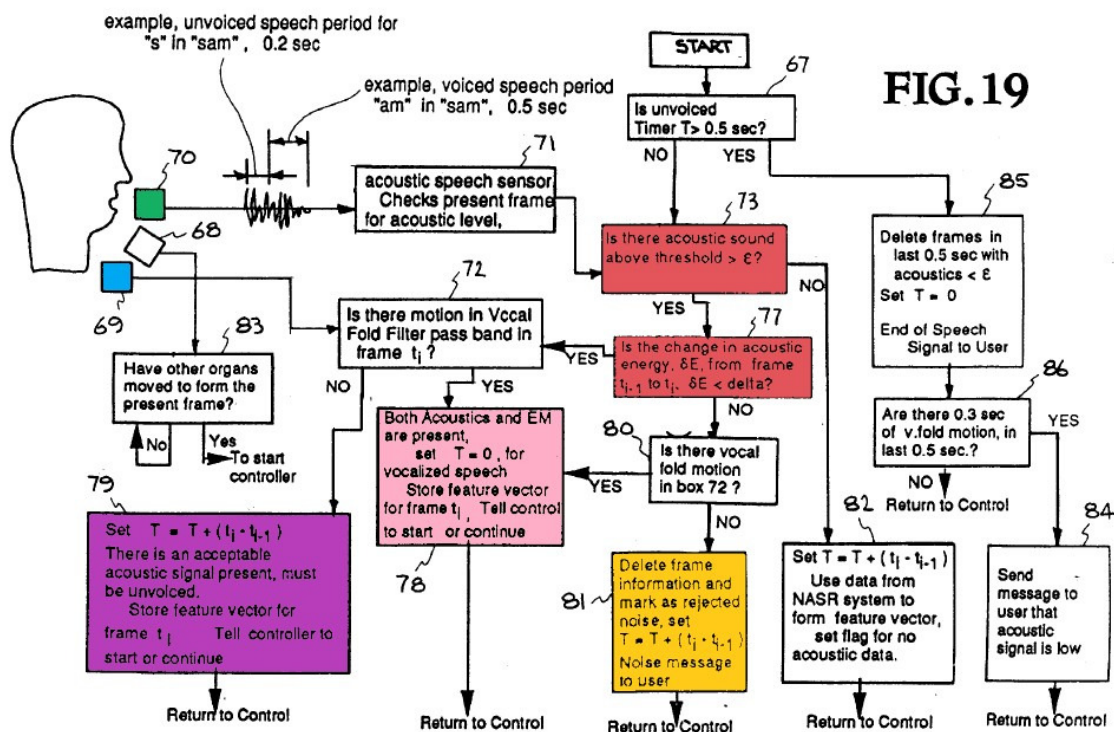
Dr. Holzrichter's patent issued in 1999. That patent described a system and sensor for detecting voiced and unvoiced speech in the presence of varying

background noise. (Ex. 1003; Ex. 1002 ¶ 38.)) The system included **electromagnetic (EM) sensors directed towards the speech organs (69)** and a **traditional acoustic microphone (70)**. (Ex. 1003, 31:15-17 (disclosing “microphone 70”), 32:31-35 (disclosing “tongue and jaw motion EM sensors”), Fig. 19; Ex. 1002 ¶ 38.) The **EM sensors** use radar waves to detect physiological information associated with human voicing activity, such as jaw, tongue, and vocal fold motions. (Ex. 1003, 32:31-35, 8:1-24, cl.42; Ex. 1002 ¶ 38.)

To identify voiced and unvoiced speech, Holzrichter describes separating a portion of the audio signal received from **microphone 70** into a frame. (Ex. 1003, 31:61-63; Ex. 1002 ¶ 39.) 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)**. (Ex. 1003, 31:67-32:4; Ex. 1002 ¶ 39.) If the gain exceeds the preset threshold (73) and is lower than a noise threshold (77), then the system identifies the frame as speech. (Ex. 1003, 31:67- 32:10; Ex. 1002 ¶ 39.) If the gain is not within these thresholds, then the system can **discard the frame as noise (81)**. (Ex. 1003, 33:43-46; Ex. 1002 ¶ 39.)

For frames identified as speech, Holzrichter’s system determines whether the speech is voiced or unvoiced. (Ex. 1003, 31:67-32:10; Ex. 1002 ¶ 40.) This determination is made by an algorithm that “looks for vocal fold motion to occur

within the same speech time period[.]” (Ex. 1003, 31:67-32:4, Fig. 19 (box 79); Ex. 1002 ¶ 40.) If vocal movements occurred at the same time as the speech frame, Holzrichter’s system identifies the frame as **voiced speech (78)**. (Ex. 1003, 32:4-7; Ex. 1002 ¶ 40.) If no vocal movements occurred, the frame is labeled as **nonvoiced speech (79)**. (Ex. 1003, 32:7-10; Ex. 1002 ¶ 40.) Holzrichter explains that speech organ location information “can be *correlated with* simultaneous acoustic word signals from a microphone.” (Ex. 1003, 26:44-27:4 (emphasis added); Ex. 1002 ¶ 40.) Thus, Holzrichter discloses a method of detecting voiced and unvoiced speech based on the correlation between physiological information (vocal movements as detected by a voicing sensor) and acoustic speech signals (as detected by a microphone). (Ex. 1002 ¶ 40.) This process is shown in Holzrichter’s Figure 19:



(Ex. 1003, Fig. 19 (annotated); Ex. 1002 ¶ 40.) Holzrichter shows that it was known in the art to: (a) compare the strength of a microphone signal to a threshold and identify the signal as speech when a threshold is exceeded; and (b) correlate physiological vocal movements with an acoustic signal to determine whether the speech is voiced or unvoiced. (Ex. 1002 ¶ 40.)

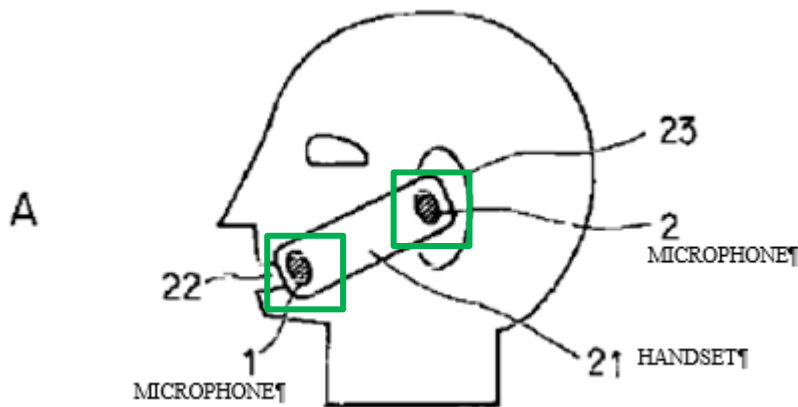
B. Using Cross Correlation to Determine Whether a Desired Signal Is Present Was Known.

Although Holzrichter does not describe using a specific mathematical function for correlating the vocal fold movements and microphone signals, various correlation functions were well known in the art. (Ex. 1002 ¶ 41.) For decades, the

cross-correlation function has been used to determine whether a desired signal (e.g., speech) is present. (*Id.*) For example, 40 years ago, Kinsler disclosed methods of detecting a known signal, such as human speech, in an acoustic sample. (Ex. 1004; Ex. 1002 ¶ 41.) Kinsler explained that the “optimum processing system” to search for a “known signal” in a “received signal” is through “cross correlation.” (Ex. 1004, 256; Ex. 1002 ¶ 41.) Kinsler taught that the cross-correlation method is used to calculate a value using a “cross correlation” function, $F(T)$. (Ex. 1004, 256; Ex. 1002 ¶ 41.) The value is then compared against a threshold. (Ex. 1004, 252, 256- 57; Ex. 1002 ¶ 41.) If the converted value meets or exceeds the threshold, then the desired signal (e.g., speech) is present in the noise. (Ex. 1004, 252, 257; Ex. 1002 ¶ 41.)

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

Although the system described in Holzrichter used a single microphone, similar systems that used two microphones were widely known. (Ex. 1002 ¶ 42.) For example, Takano, which published in 1999, disclosed a speech recognition system having **two microphones**. (*Id.*) 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). (Ex. 1005, [0022], [0050]-[0052]; Ex. 1002 ¶ 42.) An example is shown in Takano’s Figure 6:



(Ex. 1005, Fig. 6 (annotated); Ex. 1002 ¶ 42.)

Takano discloses using the spatial difference between microphones to detect speech. (Ex. 1002 ¶ 43.) Specifically, the system receives acoustic signals at each microphone and splits each signal into different frequency components. (Ex. 1005, [0009], [0023-0025], Figs. 7, 16, 17; Ex. 1002 ¶ 43.) For each frequency component, the system calculates a “level difference” between the two signals. (Ex. 1005, [0009], [0023-0025], Figs. 7, 16, 17; Ex. 1002 ¶ 43.) 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.” (Ex. 1005, [0022], [0015]; Ex. 1002 ¶ 43.)

Takano teaches that the **level difference** can be compared against **a threshold** to identify a signal as speech or noise. (Ex. 1005, [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; Ex. 1002 ¶ 44.) When the level difference exceeds **a threshold**, Takano identifies the signal as **speech**. (Ex. 1005, [0009], [0023-0025], Figs. 7, 17; Ex. 1002 ¶ 44.) If the level difference is lower than the threshold, Takano’s system identifies the signal as **noise**. (Ex. 1005, [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; Ex. 1002 ¶ 44.)

Takano further teaches that the threshold “can be determined experimentally.” (Ex. 1005, [0026]; Ex. 1002 ¶ 45.) 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:

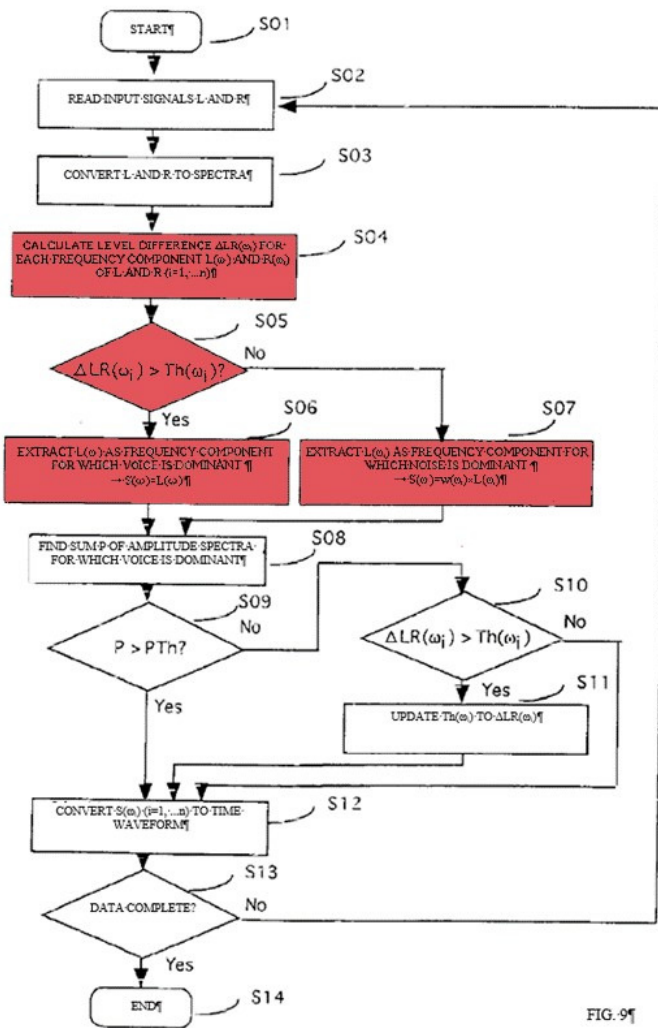


FIG. 9

(Ex. 1005, Fig. 9 (annotated); Ex. 1002 ¶ 45.)

As another example, Torick—which published over 50 years ago—disclosed a system for detecting “speech sounds” using two microphones. (Ex. 1006, 169; Ex. 1002 ¶ 46.) One microphone is used to detect speech, while the other is “used to sense the ambient noise.” (Ex. 1006, 169; Ex. 1002 ¶ 46.)

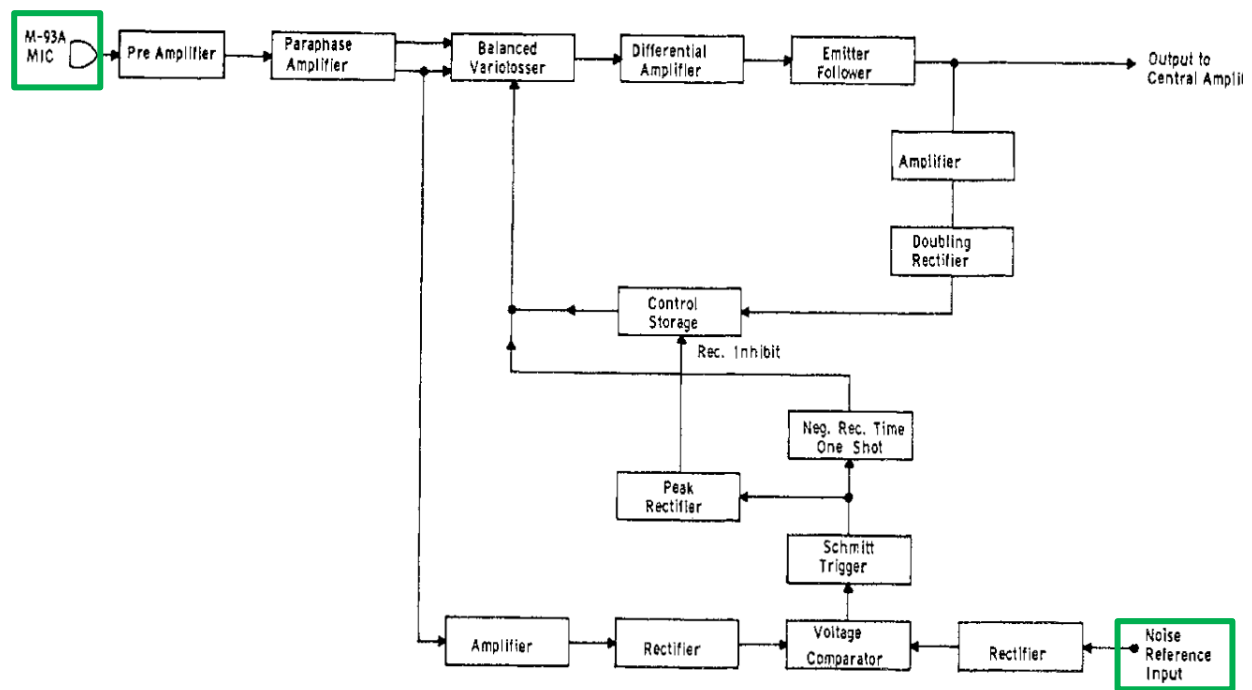


Fig. 4. Crew member's amplifier and control unit.

(Ex. 1006, Fig. 4 (annotated); Ex. 1002 ¶ 46.)

“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.” (Ex. 1006, 169; Ex. 1002 ¶ 47.) When the signals from the two microphones are equal, the system determines that no speech is occurring. (Ex. 1006, 169; Ex. 1002 ¶ 47.) However, when speech sounds occur at the speech microphone, the level between the two microphones is no longer equal and speech is detected. (Ex. 1006, 169; Ex. 1002 ¶ 47.) Torick explains that “this is analogous to the presence of a virtual threshold which rises and falls with corresponding changes in noise levels.” (Ex. 1006, 169;

Ex. 1002 ¶ 47.)

Kaneda, which published in 1993, is another example. (Ex. 1002 ¶ 48.) 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); Ex. 1002 ¶ 48.) 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.” (Ex. 1007, 2:13-18; Ex. 1002 ¶ 48.) If the power difference is lower than the threshold, the period is identified as noise. (Ex. 1007, 1:51-68, 2:1-22, Fig. 3(a)- 3(c); Ex. 1002 ¶ 48.)

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. (Ex. 1002 ¶ 49.)

III. THE '058 PATENT

A. Overview

Like Holzrichter, the '058 patent describes a system for detecting voiced and unvoiced speech using acoustic and non-acoustic (e.g., physiological) sensors.

(Ex. 1001, Abstract, Figs. 1, 3; Ex. 1002 ¶ 51.) The system includes a **processor (30)** connected to **acoustic microphones 10 (MIC 1 and MIC 2)** and a **voicing sensor 20 (VAD)**:

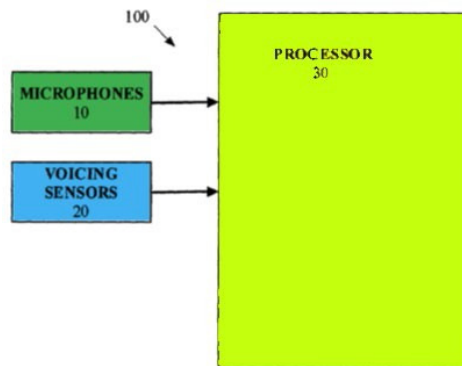


Figure 1

(Ex. 1001, Fig. 1 (annotated); Ex. 1002 ¶ 51.) The **acoustic microphones 10 (MIC 1 and MIC 2)** receive acoustic signals ($m1(n)$ and $m2(n)$), while the **voicing sensor 20 (VAD)** receives physiological information associated with human voicing activity, such as vocal fold vibrations:

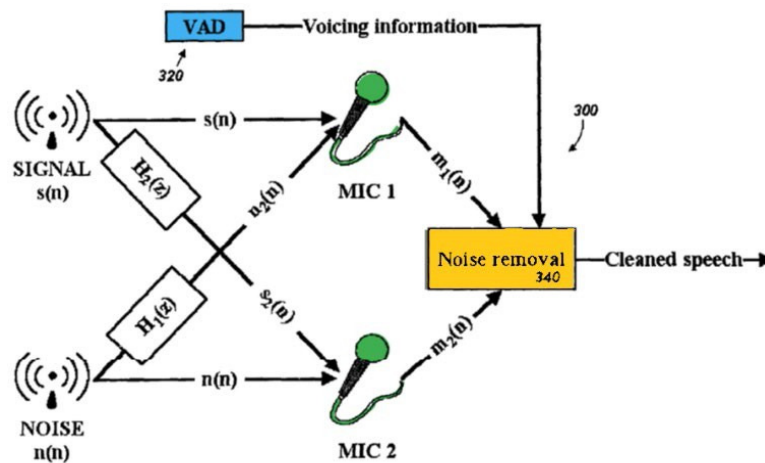


Figure 3

(Ex. 1001, Fig. 3 (annotated), 2:30-37, 3:64-4:13, 5:26-29; Ex. 1002 ¶ 51.) The **voicing sensor 20** may take several forms, including a radio frequency (radar) sensor. (Ex. 1001, 3:64-66; Ex. 1002 ¶ 51.) 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:

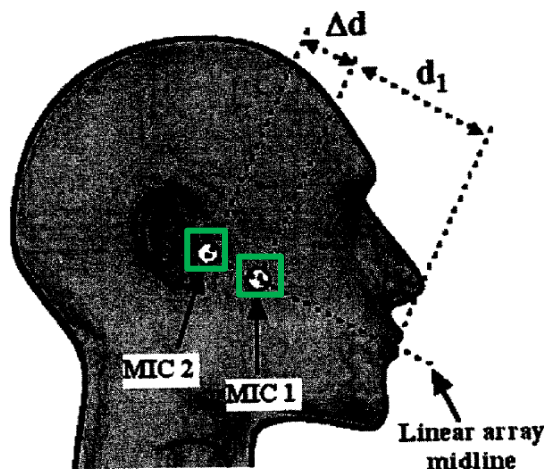


Figure 7

(Ex. 1001, Fig. 7, 7:52-67 (“the difference in signal strength in Mic 1 and Mic 2 ... will be proportional to both d_1 and Δd ”), 8:16-17 (the gain increases when the user is speaking because “Mic 1 is closer to the mouth than Mic 2”); Ex. 1002 ¶ 51.)

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**. (Ex. 1001, 4:64-5:3, 5:11-25; Ex. 1002 ¶ 52.) **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. (Ex. 1001, 5:39-45; Ex. 1002 ¶ 52.)

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. (Ex. 1001, 5:39-45, 7:52-67; Ex. 1002 ¶ 53.) 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. (Ex. 1001, 8:21-48, Fig. 4; Ex. 1002 ¶ 53.) If the difference parameters are less than the gain threshold, the processor identifies the acoustic signals as noise. (Ex. 1001,

8:41-43; Ex. 1002 ¶ 53.) The processor may then remove the identified noise (noise removal 340). (Ex. 1001, 7:18-19; Ex. 1002 ¶ 53.)

B. Prosecution History

The application that issued as the '058 patent was filed on May 30, 2002. (Ex. 1014, 296-333.) After the PTO rejected all claims as obvious over U.S. Patent No. 5,590,241 ("Park") and U.S. Patent No. 6,233,551 ("Cho"), the applicant argued that Park fails to disclose using two microphones and voicing activity detection. (*Id.*, 131.) The applicant further argued that Cho fails to disclose using two microphones and detecting voiced and unvoiced speech, as claimed. (*Id.*, 132.) The PTO maintained its rejection of claims 1-5, but withdrew the rejection of original claims 6 and 7 without any explanation. (*Id.*, 114-119.) However, the only feature that claim 6 added to claim 5 was the use of a voicing sensor to generate cross correlation data, which was well known in the art. Despite this, the '058 patent was issued on July 17, 2007 with the allowed claims and a few additional new dependent claims. (Ex. 1001.)

IV. STATEMENT OF RELIEF REQUESTED

A. Grounds

The Board should cancel the claims under § 103 as follows:

Ground	Claims Challenged	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

B. The References Are Prior Art.

The '058 patent was filed May 30, 2002, claiming priority to ten U.S. provisional patent applications, the earliest of which is U.S. Provisional No. 60/294,383, filed May 30, 2001. (Ex. 1001; Ex. 1002 ¶¶ 55.) Thus, the earliest possible priority date for the '058 patent is May 30, 2001.¹ All of the references in the Grounds are prior art under pre-AIA § 102(b) because they published more than one year earlier:

- (1) Holzrichter published on December 21, 1999 (Ex. 1003);
- (2) Kinsler published in 1982 (Ex. 1004; Ex. 1019 ¶¶ 16-27);
- (3) Takano published on November 5, 1999 (Ex. 1005);

¹ Petitioner does not concede the claims are entitled to the claimed priority date.

- (4) Torick published in December 1966 (Ex. 1006; Ex. 1019 ¶¶ 28-39);
- (5) Kaneda published on May 4, 1993 (Ex. 1007); and
- (6) Lyons published in 1997 (Ex. 1008; Ex. 1019 ¶¶ 40-55).

C. The Asserted References Are Analogous Art.

The references are analogous art because they are from the same field as the '058 patent, e.g., processing audio signals. (Ex. 1002 ¶¶ 24, 57; Ex. 1001, 1:18-19); *Unwired Planet, LLC v. Google Inc.*, 841 F.3d 995, 1000 (Fed. Cir. 2016). They are also pertinent to a problem the inventor was focused on, e.g., enhancing the quality of audio signals using signal processing techniques. (Ex. 1002 ¶¶ 24, 57; Ex. 1001, Abstract, 1:34-47.)

V. LEVEL OF ORDINARY SKILL

Based on the relevant factors, *In re GPAC Inc.*, 57 F.3d 1573, 1579 (Fed. Cir. 1995), a person of ordinary skill in the art (“POSITA”) would have had a minimum of a bachelor’s degree in computer engineering, computer science, electrical engineering, mechanical engineering, or a similar field, and approximately three years of industry or academic experience in a field related to acoustics, speech recognition, speech detection, or signal processing. (Ex. 1002 ¶ 32.) Work experience can substitute for formal education and additional formal education can substitute for work experience. (*Id.*)

VI. CLAIM CONSTRUCTION

Although the parties may propose constructions for certain terms in the litigation for purposes of determining infringement, no claim terms require construction to resolve the obviousness challenges here.² *Nidec Motor Corp. v. Zhongshan Broad Ocean Motor Co.*, 868 F.3d 1013, 1017 (Fed. Cir. 2017); *Vivid Techs., Inc. v. Am. Sci. & Eng'g, Inc.*, 200 F.3d 795, 803 (Fed. Cir. 1999).

²In the litigation, the parties may dispute whether some challenged claims are invalid under 35 U.S.C. § 112. For purposes of this proceeding only, Petitioner assumes the claims are not invalid under § 112.

VII. GROUND FOR UNPATENTABILITY

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

Holzrichter, Takano, and Kinsler render claims 1-5 obvious. (Ex. 1002 ¶¶ 59-114.)

1. Claim 1

a. Preamble

The preamble recites “a system for detecting voiced and unvoiced speech in acoustic signals having varying levels of background noise.” Holzrichter discloses such a system. Ex. 1002 ¶¶ 60-62.)

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, Fig. 19, 22:6-28, 31:8-14 (“Fig. 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.”), 34:41-67 (discussing a “voiced-unvoiced algorithm”), 52:66-53:2 (discussing a method for determining whether “voiced speech, or unvoiced speech, or speech or background noise” has occurred), Table II, Figs. 9A- 9B; Ex. 1002 ¶¶ 61-62.)

b. At Least Two Microphones

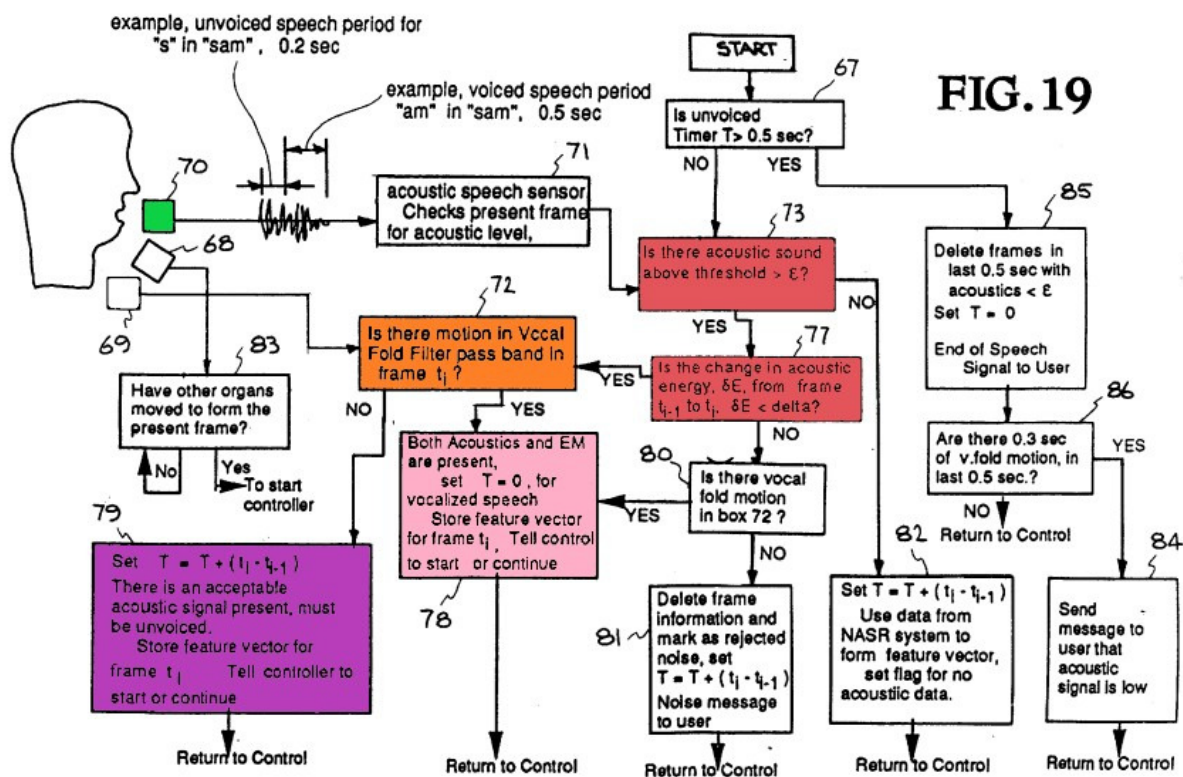
Claim 1 recites “at least two microphones that receive the acoustic signals[.]”

Holzrichter discloses using a microphone for receiving acoustic signals. (Ex. 1003, 31:15-17; Ex. 1002 ¶ 64.) However, it would have been obvious to improve Holzrichter's system by adding a second microphone as disclosed in Takano. (Ex. 1002 ¶¶ 63-78.) As described below, such a modification would improve Holzrichter's methods for classifying input signals as either noise or speech based on signal strength. (Ex. 1002 ¶ 64.)

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.”), 8:50-9:3 (Holzrichter's systems can include “microphones against the human face and neck”), 31:67-32:4, Fig. 19; Ex. 1002 ¶ 65.) 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; Ex. 1002 ¶ 65.) Specifically, if the received “signal exceeds a preset threshold 73, but ... is less than a loud noise threshold 77,” then the system identifies the signal as speech. (Ex. 1002 ¶ 65.)

Holzrichter's Figure 19 illustrates this process. (Ex. 1002 ¶ 66.) As shown, sound is received by Holzrichter's **microphone 70**. (*Id.* ¶ 66.) 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.* ¶ 66.)



(Ex. 1003, Fig. 19 (annotated); Ex. 1002 ¶ 66.) 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. (Ex. 1002 ¶ 66.)

Takano discloses a similar process but uses two microphones for receiving acoustic signals. (Ex. 1005, [0009] (“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”),

[0014]-[0015], [0023-0025], [0029], [0049]-[0051], Figs. 1, 6, 7, 16, 17, 18A; Ex. 1002 ¶ 67.) Takano's Figure 6 shows these microphones, which are virtually identical to the two microphones described in the '058 patent (Ex. 1002 ¶ 68):

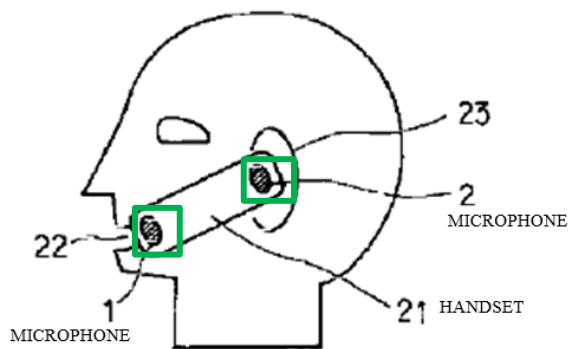
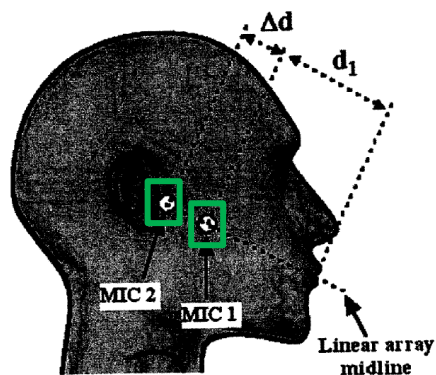


Figure 7

Ex. 1005 (Takano), Fig. 6

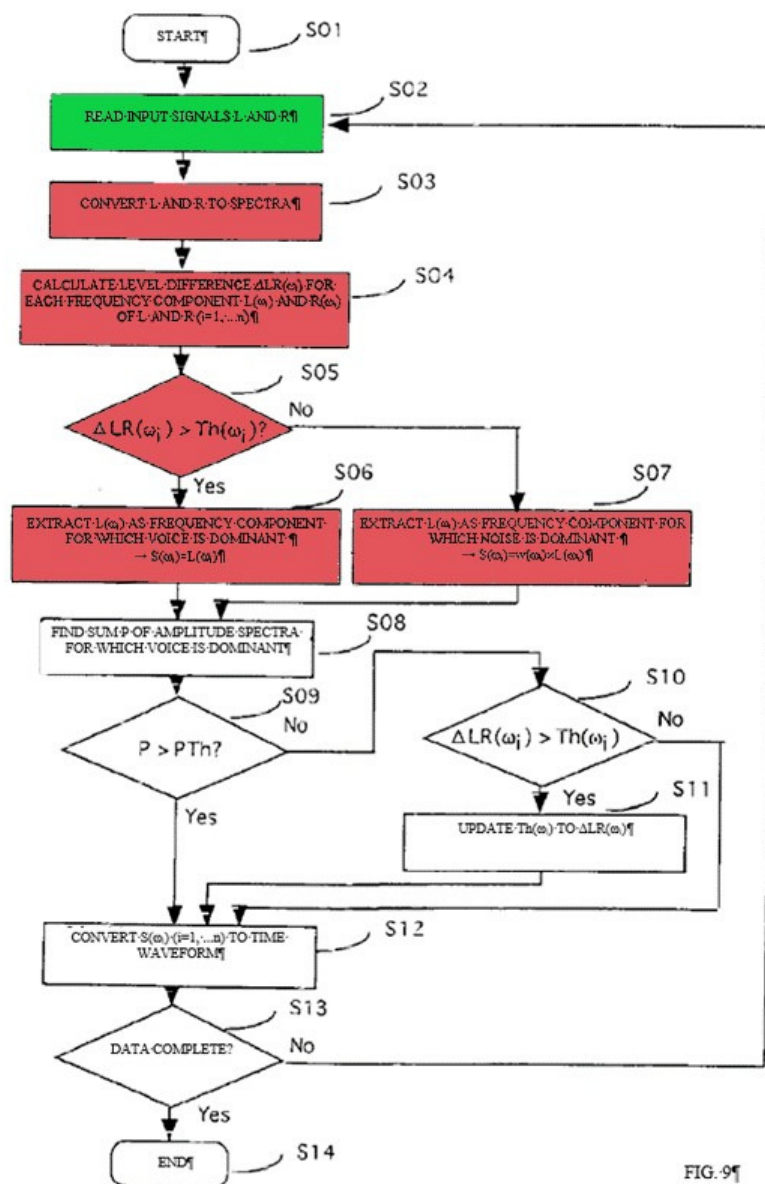
Ex. 1001 ('058 Patent), Fig. 7

Takano's system receives acoustic signals at these two microphones and then, as in Holzrichter, uses the acoustic energy of the signals received by the microphones to determine whether the system should further analyze the signals for speech. (Ex. 1005, [0009], [0023-0025], [0014]-[0015], [0029], [0040]-[0042], [0049]-[0051], Figs. 1, 6, 7, 8, 9, 16, 17, 18A; Ex. 1002 ¶ 69.). Takano's Figure 9 illustrates this process.

First, at **step S02**, Takano receives acoustic signals at its two microphones. (Ex. 1005, [0023] (explaining step S02 in Figure 7); *see also id.*, [0027], [0030] (explaining that overlapping steps in Figures 8 and 9 operate in the same manner as in Figure 7); Ex. 1002 ¶ 70.)

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. (Ex. 1005 at [0023-0025]; *see also id.*, [0009] ("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"), [0040]-[0042]; Ex. 1002 ¶ 71.)

Next, in steps **S05, S06, and S07**, if the level difference exceeds a threshold, Takano determines the “target signal” (speech) is “dominant.” (Ex. 1002 ¶ 72.) If the level difference is below the threshold “the target signal is not dominant.” (Ex. 1005, [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]; Ex. 1002 ¶ 72.)



(Ex. 1005, Fig. 9 (annotated).) 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. (Ex. 1002 ¶ 72.)

A POSITA 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 POSITA 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). (Ex. 1002 ¶¶ 73-78.) Because Holzrichter already recognizes that the strength of acoustic signals (acoustic energy) could be used to determine whether a signal is speech or noise (Ex. 1003, 31:67-32:10, Fig. 19), a POSITA reading Holzrichter would have been motivated to look for improved methods for classifying signals as noise or speech based on signal strength, such as the system of Takano. (Ex. 1002 ¶ 73.) Because Takano's system for classifying noise or speech relies on the use of two microphones, a POSITA would be motivated to add a second microphone to the system of Holzrichter to use the methods described in Takano. (*Id.*) Indeed, the methods described in Takano have many advantages.

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 of the invention is to have “a noise suppression processing function”), [0022]-[0029], [0059], Figs. 6A-6C; Ex. 1002 ¶ 74.) 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. (Ex. 1002 ¶ 74.)

Second, Takano discloses that using a second microphone provides better recovery of high-frequency components in voiced speech. (Ex. 1005, [0059]; Ex. 1002 ¶ 75.) A POSITA 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.

Third, the proposed combination represents nothing more than a predictable use of prior art elements according to established functions (e.g., using two microphones to detect speech). (Ex. 1002 ¶ 76.)

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. (Ex. 1002 ¶ 77); *KSR Int’l Co. v. Teleflex Inc.*, 550 U.S. 398, 417 (2007). 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. (Ex. 1002 ¶ 77.) Thus, the combination also reflects nothing more than a combination of known elements (microphones) to achieve an expected result (improved speech detection). (*Id.*)

A POSITA 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. (*Id.* ¶ 78.)

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

Claim 1 further recites “at least one voicing sensor that receives physiological information associated with human voicing activity[.]” Holzrichter discloses this limitation. (Ex. 1002 ¶¶ 79-80.)

Holzrichter discloses a radar sensor (EM sensor 68) that is used to detect voicing activity. (Ex. 1003, 32:31-35 (“An important example is that 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”), 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; Ex. 1002 ¶ 80.)

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

Claim 1 next recites “at least one processor coupled among the microphones and the voicing sensor[.]” Holzrichter discloses this limitation. (Ex. 1002 ¶ 81.)

Holzrichter discloses a processor coupled to microphone 70 and EM sensors 68. (Ex. 1003, 31:15-20 (“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.”), 18:49-67, 9:55-10:14, 18:24-48, Figs. 5, 13; Ex. 1002 ¶ 82.) It would have been obvious to a POSITA 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. (Ex. 1002 ¶ 82.) A POSITA would be motivated to do this for the reasons discussed above. (*Id.*; *supra* §VI.A.1.b.)

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

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

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[.]” Holzrichter, alone or in combination with Kinsler, renders this limitation obvious. (Ex. 1002 ¶¶ 84-96.)

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 (“[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”), Figs. 19, 15A-15D; Ex. 1002 ¶ 85.) Figure 19 of Holzrichter illustrates this process.

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; Ex. 1002 ¶ 86.) If vocal motion is present, Holzrichter identifies the acoustic signal as voiced speech (**step 78**). (Ex. 1003, 31:61-32:27, Fig. 19;

FIG. 19

example, unvoiced speech period for "s" in "sam", 0.2 sec

example, voiced speech period "am" in "sam", 0.5 sec

START 67

Is unvoiced Timer T > 0.5 sec? 67

NO YES

Is there acoustic sound above threshold > E? 73

NO YES

Is the change in acoustic energy, δE , from frame t_{i-1} to t_i , $\delta E < \delta$? 77

YES NO

Is there vocal fold motion in box 72? 80

YES NO

Delete frame information and mark as rejected noise, set $T = T + (t_i - t_{i-1})$ Noise message to user 81

Return to Control

Both Acoustics and EM are present, set $T = 0$, for vocalized speech Store feature vector for frame t_i , Tell controller to start or continue 78

Return to Control

Is there motion in Vocal Fold Filter pass band in frame t_i ? 72

YES NO

Have other organs moved to form the present frame? 83

No Yes

To start controller 79

Set $T = T + (t_i - t_{i-1})$ There is an acceptable acoustic signal present, must be unvoiced. Store feature vector for frame t_i , Tell controller to start or continue 79

Return to Control

Delete frames in last 0.5 sec with acoustics < E Set $T = 0$ End of Speech Signal to User 85

Are there 0.3 sec of v-fold motion, in last 0.5 sec.? 86

YES NO

Send message to user that acoustic signal is low 84

Return to Control

Set $T = T + (t_i - t_{i-1})$ Use data from NASR system to form feature vector, set flag for no acoustic data. 82

Return to Control

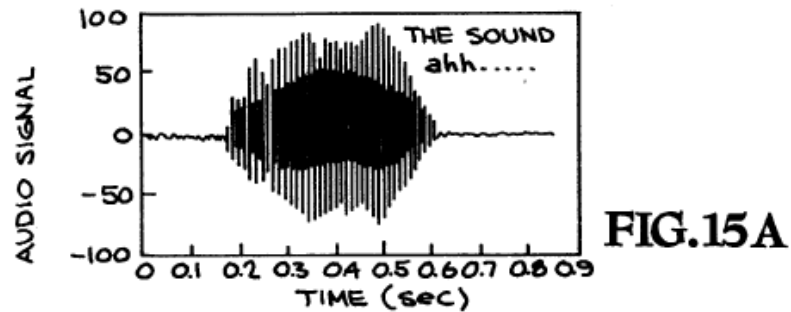
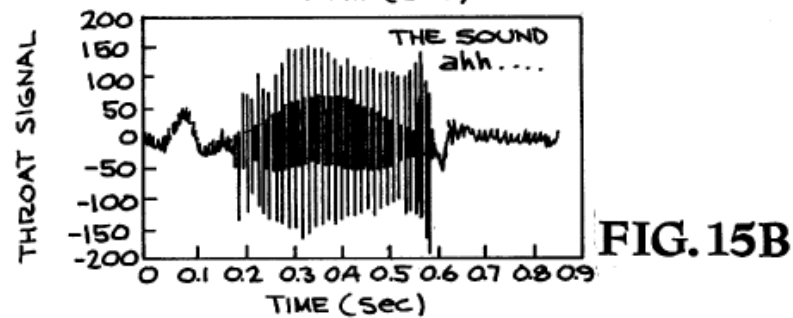
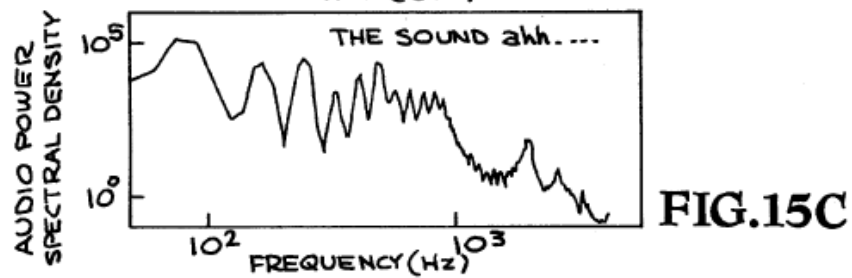
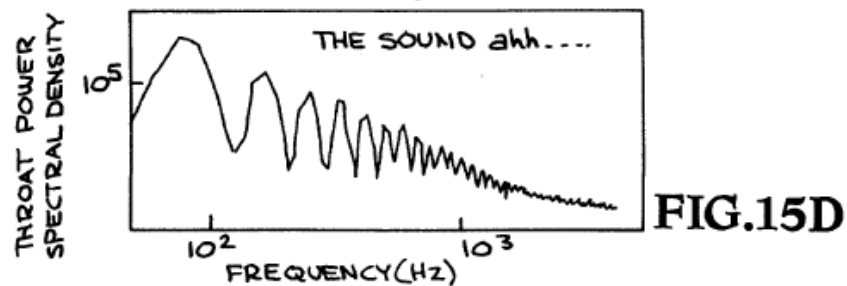
Return to Control

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; Ex. 1002 ¶ 87.)

Holzrichter

Figures 15A-D show the correlation between the two signals:



(Ex. 1003, Fig. 15A-D; Ex. 1002 ¶ 87.)

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. (Ex. 1002 ¶ 88.)

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, Ex. 1002 ¶ 89.) Kinsler then explains various types of detection systems and their application. (Ex. 1004, 255-257; Ex. 1002 ¶ 89.)

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; Ex. 1002 ¶ 90.) To use the cross-correlation method, first, a value (cross correlation data) is calculated using a cross correlation function, $F(T)$. (Ex. 1004, 256; Ex. 1002 ¶ 90.) The value can then be compared against the detection threshold (“DT”). (Ex. 1004, 252, 256-57; Ex. 1002 ¶ 90.) If the converted value meets or exceeds the threshold, then the two signals are correlated (i.e., one signal is present in the other). (Ex. 1004, 252, 257, *see also id.*, 256 ($F(T)$ “plays the role of the detection system output A ” for evaluating whether a

signal is present); Ex. 1002 ¶ 90.)

A POSITA 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. (Ex. 1002 ¶ 91.)

First, a POSITA would have been motivated by Holzrichter's disclosure that the vocal movements detected by Holzrichter's EM sensor are highly correlated with speech. (Ex. 1002 ¶¶ 91-96; Ex. 1003, 26:44-27:4 (explaining "how easily" speech organ location information "can be correlated with simultaneous acoustic word signals from a microphone."), 17:4-23, 20:11-34, 26:44-27:4, 31:61-32:27.) Based on this disclosure, a POSITA would have been motivated to look to 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. (Ex. 1002 ¶ 92.)

Second, a POSITA would have been motivated by Kinsler's disclosure that its method is the "optimum processing system" for determining if two signals are correlated. (Ex. 1002 ¶ 93.) Indeed, the cross-correlation function disclosed by Kinsler is a well-known method of determining whether signals are correlated. (*Id.*;

Ex. 1015, 113-116; Ex. 1002 ¶ 93.) A POSITA 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. (Ex. 1002 ¶ 93.) 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. (Ex. 1002 ¶ 93.) Thus, a POSITA would have used Kinsler's teachings to improve Holzrichter's method of correlating signals. (*Id.*)

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). (Ex. 1002 ¶ 94.)

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). (Ex. 1002 ¶ 95); *KSR*, 550 U.S. at 417. Indeed, the addition of Kinsler's methods to Holzrichter would simply improve what Holzrichter already contemplates—correlating an acoustic

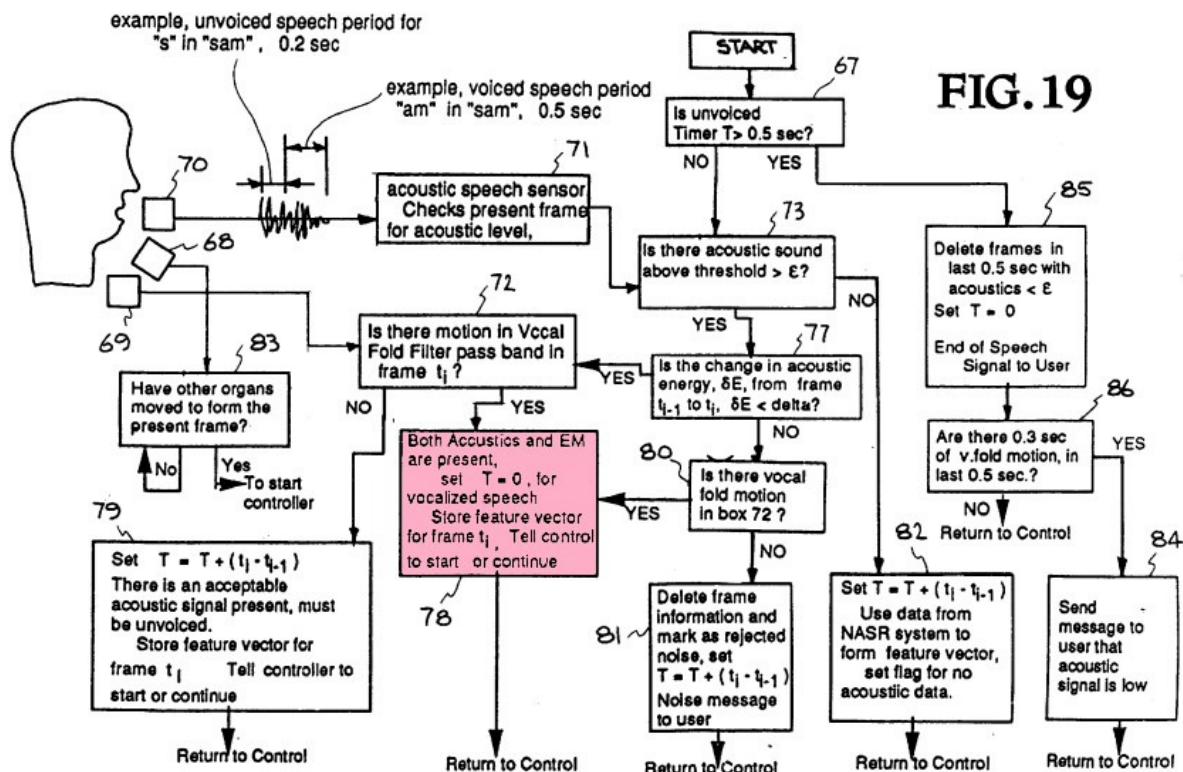
signal and a vocal movement signal. (Ex. 1002 ¶ 95.) Thus, the combination also reflects nothing more than a combination of known elements to achieve an expected result. (*Id.*)

A POSITA 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; Ex. 1002 ¶ 96) and because the cross-correlation function was widely known and used by those of skill in the art (Ex. 1002 ¶ 96; *see also* Ex. 1013, 231-237). Additionally, processors, like those in Holzrichter, commonly perform mathematical functions like cross correlation on audio signals. (Ex. 1002 ¶ 96.) Thus, a POSITA would reasonable expect that Holzrichter’s system could perform cross correlation. (*Id.*)

ii. Identifying Information as Voiced Speech

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[.]” Holzrichter combined with Kinsler discloses this limitation. (Ex. 1002 ¶¶ 97-99.)

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; Ex. 1002 ¶ 98.)



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). (Ex. 1002 ¶ 99.) A POSITA would

be motivated to combine Kinsler and Holzrichter, and to use Kinsler's cross correlation analysis, for the reasons discussed above. (Ex. 1002 ¶ 99; *supra* §VI.A.1.d.1.)

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

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[.]” Holzrichter combined with Takano discloses this limitation. (Ex. 1002 ¶¶ 100-102.)

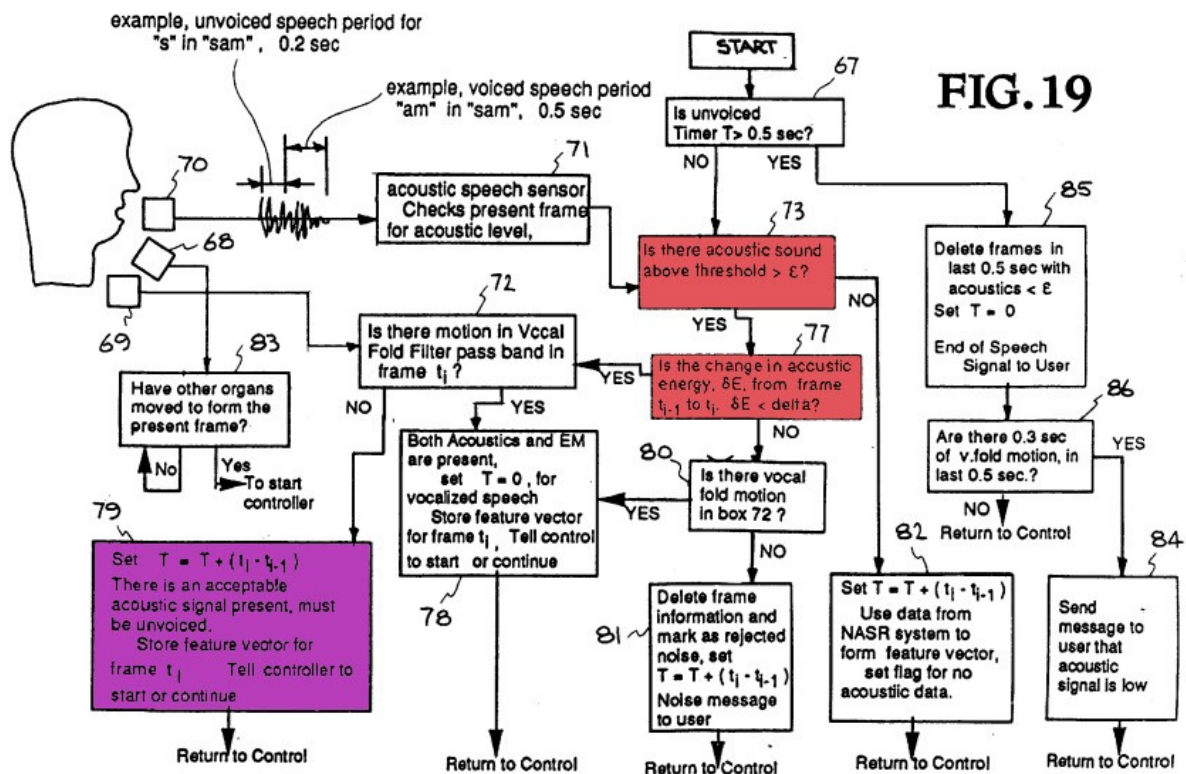
Takano discloses calculating a “level difference” (difference parameters) for each frequency component of the signals it receives at its two microphones (portions of the received acoustic signals). (Ex. 1005, [0009], [0023]-[0024], [0014]-[0015],[0029], [0040]-[0042] [0049]-[0051], Figs. 1, 6, 7, 16, 17, 18A; Ex. 1002 ¶ 101.) This “level difference” is calculated based on the difference in “amplitude spectrum components” (i.e., gain) of the two signals. (*Id.*) 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. (Ex. 1002 ¶ 101.)

As explained above, *supra* §VI.A.1.b., a POSITA 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. (Ex. 1002 ¶ 102.)

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

Claim 1 recites that the processor “identifies information of the acoustic signals as unvoiced speech when the difference parameters exceed a gain threshold[.]” Holzrichter combined with Takano discloses this limitation. (Ex. 1002 ¶¶ 103-106.)

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; Ex. 1002 ¶ 104.) 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. (Ex. 1003, 31:61-32:27; Ex. 1002 ¶ 104.) 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**). (Ex. 1002 ¶ 104.)

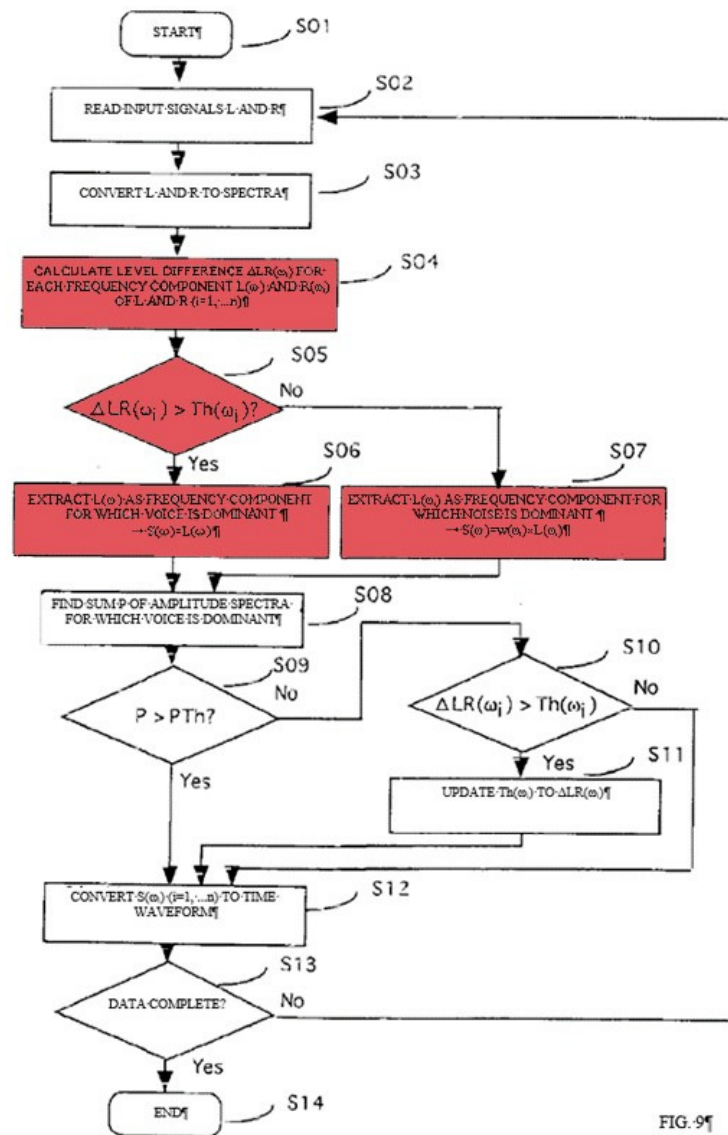


(Ex. 1003, Fig. 19 (annotated); Ex. 1002 ¶ 104.)

As discussed above, Holzrichter does not disclose the use of a second microphone or the claimed difference parameters. (Ex. 1002 ¶ 105.) But Takano does. (*Id.*) 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]; *see also id.*, [0014]-[0015], [0029], [0040]-[0042] [0049]-[0051], Figs. 1, 6, 7, 16, 17, 18A; Ex. 1002 ¶ 105.) If the level difference exceeds the threshold, “the target signal is dominant,” meaning that the signal is identified as speech. (Ex. 1005, [0025]; *see also id.*, [0014]-[0015], [0029], [0040]-[0042]

[0049]-[0051], Figs. 1, 6, 7, 16, 17, 18A; Ex. 1002 ¶ 105.) Takano's Figure 9

shows this process:



(Ex. 1005, Fig. 9 (annotated); Ex. 1002 ¶ 105.)

As explained above, *supra* § VI.A.1.b., a POSITA 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. (Ex. 1002 ¶ 106.) In such a combination, a POSITA 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). (*Id.*) The combined system would identify a signal as unvoiced speech only when the level difference of the two microphones exceeds a threshold. (*Id.*) Thus, a POSITA 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. (*Id.*)

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

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. (Ex. 1002 ¶¶ 107-108.) 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; Ex. 1002 ¶ 108.) As discussed above, *supra* §VI.A.1.b., it would have been obvious to a POSITA to incorporate this feature of Takano into Holzrichter's system. (Ex. 1002 ¶ 108.)

2. Claim 2

Independent claim 2 is similar to claim 1, except it is a method claim. Thus, claim 2 is obvious for the same reasons as claim 1. (Ex. 1002 ¶ 109; *supra* §VI.A.1.)

3. Claim 3

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.” Holzrichter combined with Takano discloses or renders obvious this limitation. (Ex. 1002 ¶¶ 110-111.)

Takano teaches that its threshold “can be determined experimentally.” (Ex. 1005, [0026]; Ex. 1002 ¶ 111.) A POSITA 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. (Ex. 1002 ¶ 111.) 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; Ex. 1002 ¶ 111.) Standard deviation is

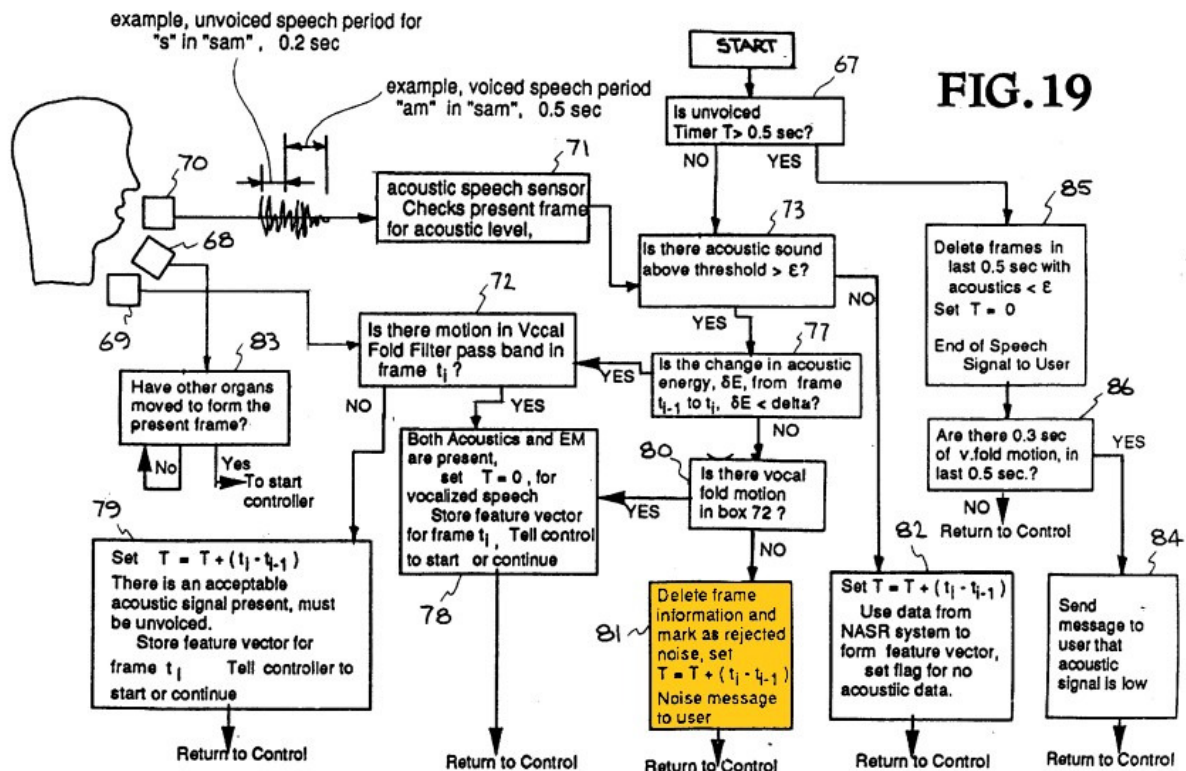
also commonly used to compare and evaluate statistics like the difference in levels between two microphones. (Ex. 1002 ¶ 111.) Indeed, the 058 patent recognizes that determining the threshold based on standard deviation is one of the “standard signal processing techniques” that could be used to determine 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); Ex. 1002 ¶ 111.) Accordingly, this limitation would have been obvious to a POSITA in view of Holzrichter and Takano. (Ex. 1002 ¶ 111.)

4. Claim 4

Claim 4 depends from claim 2 and adds the step of “performing denoising on the identified noise.”

Holzrichter discloses this limitation. (Ex. 1002 ¶¶ 112-113.) Holzrichter teaches 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); Ex. 1002 ¶ 113.) Holzrichter explains that, when its system identifies a portion of the acoustics signal as noise, that portion is removed and is not subsequently processed. (Ex. 1003, 33:39-34:7, Fig. 19; *see also 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; Ex. 1002 ¶ 113.)



(Ex. 1003, Fig. 19 (annotated); Ex. 1002 ¶ 113.)

5. Claim 5

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.”

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)”); Ex. 1002 ¶¶ 114-115.)

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

1. Claim 3

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

As discussed above, Takano teaches that its threshold “can be determined empirically.” (Ex. 1005, [0026]; Ex. 1002 ¶ 118.) Although it would have been obvious to a POSITA to generate the gain threshold using standard deviations as discussed above (*supra* §VI.A.3), it also would have been obvious in further view of Lyons. (Ex. 1002 ¶¶ 117-127.)

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; Ex. 1002 ¶ 119.) Lyons teaches that the standard deviation of digital signals is defined as the positive square root of the variance:

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

(Ex. 1008, 320; Ex. 1002 ¶ 119.) “If the sequence $x(n)$ represents a time series of signal samples, we can say that x_{ave} specifies the constant, or DC, value of the signal, [and] the standard deviation σ reflects the amount of the fluctuating, or AC, component of the signal[.]” (Ex. 1008, 320, *see also id.*, 479; Ex. 1002 ¶ 119.)

Lyons discloses that its digital signal processing has many applications, including “voice recognition [and] data communication” (Ex. 1008, 1; Ex. 1002 ¶ 120), and cites to more than 20 publications directed to digital signal processing used in speech recognition (Ex. 1008, 21, 47, 127, 128, 155, 156, 217, 218, 295, 317, 318, 441, 442; Ex. 1002 ¶ 120).

It would have been obvious to a POSITA 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. (Ex. 1002 ¶ 121.) A POSITA 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. (*Id.*) A POSITA would then use this

upper bound as the threshold for Takano to determine when it is receiving noise instead of speech. (*Id.*)

A POSITA would have been motivated to generate the gain threshold in Takano using standard deviations as taught by Lyons for several reasons. (Ex. 1002 ¶ 122.)

First, POSITA 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. (Ex. 1002 ¶ 123.)

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. (*Id.* ¶ 124; Ex. 1008, 319, 476.) A POSITA 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. (Ex. 1002 ¶ 124.) Indeed, standard deviation is also commonly used to compare and evaluate statistics like the difference in levels between two microphones. (*Id.*) 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); Ex. 1002 ¶ 124.)

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; Ex. 1002 ¶ 125.)

Fourth, a POSITA 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). (Ex. 1002 ¶ 126.)

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." (Ex. 1002 ¶ 127.) 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); Ex. 1002 ¶ 127.) 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. (Ex. 1002 ¶ 127.)

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

1. Claim 1

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. (Ex. 1002 ¶ 129; *supra* §§VI.A.1.a, VI.A.1.c; VI.A.1.d; VI.A.1.d.i.-ii.) The remaining limitations of claim 1 are discussed below.

a. At Least Two Microphones

Holzrichter discloses using the strength of the output signal of a microphone to determine if a signal is speech or noise. (Ex. 1002 ¶¶ 130-139; *supra* VI.A.1.b.)

Torick discloses a system for detecting “speech sounds” using two microphones.

(Ex. 1006, 169; Ex. 1002 ¶ 131.) One microphone is used to detect speech, while

the other is “used to sense the ambient noise.” (Ex. 1006, 169; Ex. 1002 ¶ 131.)

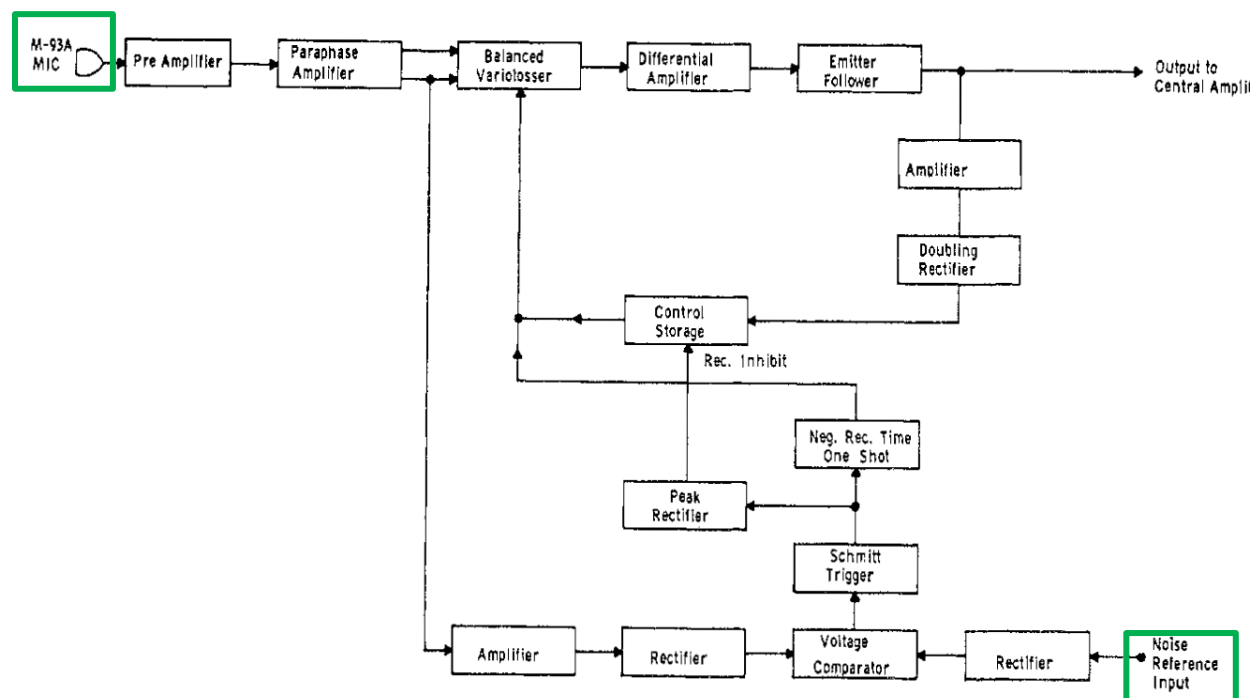


Fig. 4. Crew member's amplifier and control unit.

(Ex. 1006, Fig. 4 (annotated); Ex. 1002 ¶ 131.)

Torick uses these two microphones to implement “voice actuated switching” and “eliminate the press-to-talk switch” for microphone-based systems. (Ex. 1006, 168; Ex. 1002 ¶ 132.). 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.” (Ex. 1006, 169; Ex. 1002 ¶ 132.) When the signals

from the two microphones are equal, the system determines that no speech is occurring. (Ex. 1006, 169; Ex. 1002 ¶ 132.) However, when speech sounds occur at the speech microphone, the level between the two microphones is no longer equal and speech is detected. (Ex. 1006, 169; Ex. 1002 ¶ 132.) Torick explains that “the method is analogous to the presence of a virtual threshold which rises and falls with corresponding changes in noise levels.” (Ex. 1006, 169; Ex. 1002 ¶ 132.) Torick’s system allows for voice-actuated switching even in “high noise” environments. (Ex. 1006, 169; Ex. 1002 ¶ 132.)

A POSITA would have understood that the second microphone of Torick could be added to Holzrichter’s system so that a POSITA 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. (Ex. 1002 ¶ 133.)

A POSITA 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. (*Id.* ¶¶ 134-139.)

First, because Holzrichter recognizes the strength of acoustic signals could be used to determine noise or speech (Ex. 1003, 31:67-32:4; Ex. 1002 ¶ 135), a POSITA reading Holzrichter would have been motivated to look for other

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

(Ex. 1002 ¶ 135.) Torick discloses such as a system. (*Id.*)

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; Ex. 1002 ¶ 136.) 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. (Ex. 1002 ¶ 136.)

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). (Ex. 1002 ¶ 137.)

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. (Ex. 1002 ¶ 138); *KSR*, 550 U.S. at 417. 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). (Ex. 1002 ¶ 138.) Thus, the combination also reflects nothing more than a

combination of known elements (microphone) to achieve an expected result (improved speech detection). (*Id.*)

A POSITA 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. (Ex. 1002 ¶ 139.)

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

Holzrichter combined with Torick in light of Kaneda discloses this limitation. (Ex. 1002 ¶¶ 140-144.) As explained above, *supra* VI.C.1.a, a POSITA 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. (*Id.*)

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; Ex. 1002 ¶ 141.) “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.*)

It would have been obvious to a POSITA 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. (Ex. 1002 ¶ 142.) Thus, a POSITA 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. (*Id.*)

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 of one microphone from the other. (Ex. 1007, 2:1-22, Fig. 3(a)-3(c); Ex. 1002 ¶ 143.) 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. (Ex. 1002 ¶ 143.)

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. (*Id.* ¶ 144.) Further, a POSITA would have been motivated to combine Kaneda’s method with Torick because

it would allow a POSITA to add sufficient tolerance to Torick's system to allow the system to operate. (*Id.*) 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. (*Id.*); *KSR*, 550 U.S.at 417.

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

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; Ex. 1002 ¶ 146.) 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_{th17} .” (Ex. 1007, 2:13-18; Ex. 1002 ¶ 146.) As discussed, *supra* IV.C.1.b, it would be obvious to use the mathematical methods disclosed in Kaneda to detect speech using the microphones of Torick. (Ex. 1002 ¶ 146.)

Further, as explained previously, *supra* VI.A.1.d.iv., 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. (Ex. 1007,

2:13-18; Ex. 1002 ¶ 147.) 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; Ex. 1002 ¶ 147.)

As explained above, *supra* § IV.C.1.a, a POSITA 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. (Ex. 1002 ¶ 148.) In such a combination, a POSITA 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. (*Id.*) 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. (*Id.*) Thus, a POSITA would understand that the combined system identifies information of the acoustic signals as unvoiced speech when the difference parameters exceed a gain threshold. (*Id.* ¶¶ 145-148)

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

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; Ex. 1002 ¶ 150.) 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); Ex. 1002 ¶ 150.) As discussed, *supra* IV.C.1.b, it would be obvious to use the mathematical methods disclosed in Kaneda to detect speech using the microphones of Torick. (Ex. 1002 ¶ 150.)

Further, as discussed above, *supra* VI.C.1.a, it would have been obvious to a POSITA to incorporate this feature of Torick into the system of Holzrichter. (Ex. 1002 ¶ 151.) 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; Ex. 1002 ¶ 151.) Based on these disclosures, a POSITA 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. (Ex. 1002 ¶¶ 149-151.)

2. Claim 2

Claim 2 is a method claim that is substantially similar to the system claimed in claim 1. (Ex. 1002 ¶ 152.) The combination of Holzrichter, Torick, and Kinsler would have rendered this claim obvious for the same reasons as claim 1. (*Id.*)

3. Claim 4

Holzrichter discloses the additional limitations of claim 4. (Ex. 1002 ¶ 153; *supra* §VI.A.4.)

4. Claim 5

Holzrichter discloses the additional limitations of claim 5. (Ex. 1002 ¶ 154; *supra* §VI.A.5.)

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

1. Claim 3

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

As explained *supra* Section VI.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; Ex. 1002 ¶ 157.) Thus, it would have been obvious to a POSITA 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. (Ex. 1002 ¶ 157.) A POSITA 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. (*Id.*) A POSITA would then use this upper bound as the threshold for Takano to determine when it is receiving noise instead of speech. (*Id.*) A POSITA would have been motivated to do this for the same reasons as discussed in Section IV.B.1. (*Id.*)

VIII. SECONDARY CONSIDERATIONS OF NONOBVIOUSNESS

Secondary considerations should be considered but do not control the obviousness conclusion. *Newell Cos. v. Kenney Mfg. Co.*, 864 F.2d 757, 768 (Fed. Cir. 1988). Where a strong *prima facie* obviousness showing exists, even relevant secondary considerations supported by substantial evidence may not dislodge obviousness conclusion. *Leapfrog Enters. v. Fisher-Price, Inc.*, 485 F.3d 1157, 1162 (Fed. Cir. 2007). Petitioner is aware of no evidence supporting a claim for secondary considerations. (Ex. 1002 ¶ 158.)

IX. DISCRETIONARY DENIAL UNDER § 314(A) IS NOT APPROPRIATE

The Board should decline to exercise its discretion under § 314(a) because efficiency, fairness, and the merits support institution. *Apple v. Fintiv*, IPR2020-00019, Paper 11 (P.T.A.B. Mar. 20, 2020) (“*Fintiv*”). Each *Fintiv* factor is

addressed below.

A. Factor 1: Potential Stay

On February 28, 2023, Patent Owner (“PO”) sued Petitioner for infringement of the ’058 patent in *Jawbone Innovations, LLC v. Sony Elecs. Inc.*, No. 2:23-cv-01161-CCC-LDW (D.N.J.) (“the Litigation”). There are currently 16 instituted IPRs covering every patent claim asserted against Petitioner in the February 28, 2023 complaint, including the claims of the ’058 Patent, as follows:

- US10,779,080 (IPR2022-01059, IPR2023-00284);
- US8,019,091 (IPR2022-00649, IPR2023-00253);
- US7,246,058 (IPR2022-00623 (instituting IPR involving ’058 patent on behalf of Petitioner Google); IPR2023-00222);
- US8,321,213 (IPR2022-00797, IPR2023-00276, IPR 2023-00279);
- US8,467,543 (IPR2022-01027);
- US8,503,691 (IPR2022-01060, IPR2022-01061);
- US11,122,357 (IPR2022-01124, IPR2023-00251); and
- US8,326,611 (IPR2023-00285, IPR2023-286).

On June 12, 2023, Jawbone amended its complaint to add one additional patent.

The District of New Jersey (“DNJ”) has a “liberal” policy toward staying cases when IPRs have been instituted, to avoid wasting “a court’s time and

resources” and because “a USPTO decision could drastically alter the nature of the litigation.” *Kirsch Rsch. & Dev., LLC v. GAF Materials, LLC*, Civil Action No. 20-13683 (JMV), 2021 U.S. Dist. LEXIS 111667, at *5 (D.N.J. June 15, 2021). This is particularly true when, as here, the litigation is at its early stages. *Id.* at *13. Given that the Litigation is at its inception and all asserted claims (other than the claims in a single patent added months after the original complaint) are being challenged in instituted IPRs, the likelihood of the grant of a stay weighs against discretionary denial.

B. Factor 2: Proximity of Trial to FWD

The DNJ has not set a trial date. Indeed, the only date that has been set by the district court is an initial case management conference for July 18, 2023. (The Litigation, ECF No. 18.) Thus, this factor weighs against discretionary denial. *Global Tel-Link Corp. v. HLFIP Holding, Inc.*, IPR2021-00444, Paper 14, at 16-19 (P.T.A.B. July 22, 2021) (“The lack of a scheduled trial date . . . typically weighs against exercising discretion to deny the Petition.”); *GAF Materials LLC v. Kirsch Rsch. & Dev., LLC*, IPR2021-00192, Paper 14, at 12-13 (P.T.A.B. May 25, 2021) (finding that the second *Fintiv* factor “weighs against exercise of [the Board’s] discretion to deny institution” when no trial date has been set); *Skechers U.S.A., Inc. v. Nike, Inc.*, IPR2021-00160, Paper 10, at 9-11 (P.T.A.B. May 19, 2021)

(granting institution and finding that this factor weighs against the Board exercising discretion because “there is no trial date set in the California action” and “without a trial date we cannot objectively determine that the District Court will address validity issues before our final written decision”); *Qualcomm Inc. v. UNM Rainforest Innovations*, IPR2021-00582, Paper 11, at 11 (P.T.A.B. Aug. 27, 2021) (finding “the considerations of the second *Fintiv* factor weigh against discretionary denial” where the trial date was “removed from the calendar in the parallel District Court proceeding”).

C. Factor 3: Investment in Parallel Proceeding

PO filed its complaint on February 28, 2023, and served it on March 16, 2023. (The Litigation, ECF Nos. 1 and 5.) The parties have not conducted any discovery, and an initial scheduling conference will not occur until July 18, 2023. (*Id.*, ECF No. 18.) Thus, the parties have invested very little in the Litigation, and little is expected to be invested in the Litigation over the next several months (if the case is not stayed). Even if a scheduling order subsequently imposes some discovery deadlines over the following 6 months (*e.g.*, fact discovery, claim construction), much work will still remain, including expert reports, expert discovery, dispositive motions, pretrial motions, and trial. This “remaining investment” significantly outweighs any investment made thus far. *Samsung Elecs.*

Am. Inc. v. Snik LLC, IPR2020-01428, Paper 10, at 11 (P.T.A.B. Mar. 9, 2021);
Huawei Techs. Co. v. WSOU Invs., LLC, IPR2021-00226, Paper 10, at 13-14
(P.T.A.B. June 10, 2021).

D. Factor 4: Overlapping Issues

If this IPR is instituted and the Litigation is stayed, Petitioner could not pursue in the Litigation any invalidity ground raised or that could have been reasonably raised in this IPR. *Cal. Inst. of Tech. v. Broadcom Ltd.*, 25 F.4th 976, 991 (Fed. Cir. 2022). Thus, duplicative efforts or conflicting decisions are unlikely.

Petitioner hereby stipulates that, if this IPR is instituted but the Litigation is not stayed, Petitioner will not argue in the Litigation invalidity of the '213 patent based on the grounds in this petition. This factor weighs against discretionary denial. *Sand Revolution II, LLC v. Cont'l Intermodal Grp.-Trucking LLC*, IPR2019-01393, Paper 24, at 11-12 (P.T.A.B. June 16, 2020) (informative).

E. Factor 5: The Parties

The parties are the same, but trial in the Litigation is highly unlikely to start before this IPR reaches a final written decision, and most likely will occur long after a final written decision. Thus, this factor is neutral. *See Google LLC v.*

Jawbone Innovations LLC, IPR2022-00630, Paper 10, at 17 (P.T.A.B. Sept. 13, 2022).

F. Factor 6: Other Circumstances

The prior art here discloses a system using a voicing sensor and microphones to detect voiced and unvoiced speech. To Petitioner's knowledge, other than Amazon's IPR petition that Petitioner is seeking to join in its accompanying motion for joinder (IPR2023-00222), the grounds presented here have never been considered by the PTO, any court, or any jury. The petition's merits are particularly strong, which favors institution. *Fintiv*, IPR2020-00019, Paper 11, at 18.

Additionally, the P.T.A.B. has already instituted IPR of the '058 patent. *Google*, IPR2022-00623, Paper 12. Thus, the P.T.A.B. will be intimately familiar with the patent and claims as well as the relevant technology. Judicial efficiency therefore favors having the same panel consider the validity of the claims in view of the additional prior art references that Petitioner has identified.

For the above reasons, the *Fintiv* analysis favors institution and the Board should not deny institution under §314(a).

**X. DISCRETIONARY DENIAL
UNDER § 325(D) IS NOT APPROPRIATE**

The Board may deny institution if the Office previously considered “the same or substantially the same prior art or arguments[.]” 35 U.S.C. § 325(d). Here, all

grounds rely on one or more of Kinsler, Takano, Kaneda, Lyons, and Torick, which the Examiner never considered. Nor did the Examiner consider “substantially the same prior art or arguments.” 35 U.S.C. § 325(d). The Examiner allowed the claims during examination because the art he relied on did not disclose generating cross-correlation data. The prior art references relied on herein disclose that feature. Further, the Board has not considered any IPR petition (other than in Amazon’s IPR that Sony seeks to join) based on the Grounds presented herein. Accordingly, they are not — and could not be — cumulative of previously-considered references. *See PNC Bank, N.A., v. United Servs. Auto. Ass’n*, IPR2022-00075, Paper 21, at 9-12 (June 13, 2022) (instituting IPR when one out of eight references was not previously considered by PTO); *Haliburton Energy Servs., Inc., v. U.S. Well Servs., LLC*, IPR2021-01036, Paper 12, at 15-16 (Jan. 19, 2022) (instituting IPR when primary references were previously considered by the Patent Office but secondary references were not).

Although this petition presents the same grounds as the petition filed by Amazon (IPR2023-00222), Petitioner is filing this petition to preserve its ability to maintain a petition on the merits in the event that Amazon terminates its involvement. If Petitioner is joined to the Amazon IPR, Petitioner will take an

understudy role. Thus, the proposed grounds will not be cumulative of references previously considered by the Board.

XI. MANDATORY NOTICES

Pursuant to 37 C.F.R. § 42.8(a)(1), the mandatory notices identified in 37 § 42.8(b) are provided below as part of this Petition.

A. Real Party-In-Interest (37 C.F.R. § 42.8(b)(1))

Sony Electronics Inc., Sony Group Corporation, Sony Corporation, and Sony Corporation of America are the real parties in interest.

B. Related Matters (37 C.F.R. § 42.8(b)(2))

To the best knowledge of Petitioner, the '058 patent is or has been involved in the following district court litigations and petitions for *inter partes* review:

Name	Number	Court	Filed
<i>Jawbone Innovations, LLC v. Amazon.com, Inc.</i>	5:22-cv-06727	N.D. Cal.	Nov. 29, 2021 (transferred from E.D. Tex. on Nov. 1, 2022)
<i>Jawbone Innovations, LLC v. Amazon.com, Inc.</i>	2:21-cv-00435	E.D. Tex.	Nov. 29, 2021
<i>Jawbone Innovations, LLC v. Google LLC</i>	3:23-cv-00466	N.D. Cal.	Sept. 23, 2021 (transferred from W.D. Tex. on Feb. 1, 2023)
<i>Jawbone Innovations, LLC v. Google LLC</i>	6:21-cv-00985	W.D. Tex.	Sept. 23, 2021
Petition for <i>Inter Partes</i> Review	IPR2022-00623	P.T.A.B.	March 8, 2022
Petition for Inter Partes Review	IPR2023-00222	P.T.A.B.	Nov. 17, 2022
<i>Jawbone Innovations, LLC v. Meta Platforms, Inc.</i>	6:23-cv-00158	W.D. Tex.	Feb. 28, 2023
<i>Jawbone Innovations, LLC v. ZTE Corporation</i>	2:23-cv-00082	E.D. Tex.	Feb. 28, 2023
<i>Jawbone Innovations, LLC v. Panasonic Holdings</i>	2:23-cv-00081	E.D. Tex.	Feb. 28, 2023
<i>Jawbone Innovations, LLC v. Guangdong Oppo Mobile Telecommunications Corp., Ltd.</i>	2:23-cv-00079	E.D. Tex.	Feb. 28, 2023
<i>Jawbone Innovations, LLC v. LG Electronics, Inc.</i>	2:23-cv-00078	E.D. Tex.	Feb. 28, 2023

<i>Jawbone Innovations, LLC v. HTC Corporation</i>	2:23-cv-00077	E.D. Tex.	Feb. 28, 2023
<i>Jawbone Innovations, LLC v. Sony Electronics Inc.</i>	2:21-cv-00186	D.N.J.	Feb. 28, 2023

C. Lead and Backup Counsel (37 C.F.R. § 42.8(b)(3))

Petitioner provides the following designation of counsel:

LEAD COUNSEL	BACKUP COUNSEL
Gregory S. Gewirtz Registration No. 36,522 LERNER DAVID LLP 20 Commerce Drive Cranford, NJ 07016 Telephone: 908.645.5000 Facsimile: 908.654.7866 Email:ggewirtz@lernerdavid.com litigation@lernerdavid.com	Russell W. Faegenburg Registration No. 77,876 LERNER DAVID LLP 20 Commerce Drive Cranford, NJ 07016 Telephone: 908.645.5000 Facsimile: 908.654.7866 Email:rfaegenburg@lernerdavid.com litigation@lernerdavid.com

D. Service Information (37 C.F.R. § 42.8(b)(4))

Service information for lead and back-up counsel is provided in the designation of lead and back-up counsel above. Petitioner hereby consents to service by email at the following email address:

litigation@lernerdavid.com

E. Payment of Fees Pursuant to 37 C.F.R. § 42.103

The fee set forth in 37 C.F.R. § 42.15(a) for this petition has been paid. The undersigned further authorizes payment for any additional fees that may be due in connection with this petition to be charged to Deposit Account 121095.

F. Grounds for Standing Under 37 C.F.R. § 42.104(a)

Petitioner certifies that the '058 patent is available for IPR and that Petitioner is not barred or estopped from requesting IPR. This petition is being filed within one year of service of the original complaint against Petitioner in the district court litigation.

XII. CONCLUSION

Petitioner respectfully requests institution of this IPR and cancellation of the challenged claims.

Respectfully submitted,

LERNER DAVID LLP

Dated: June 23, 2023

/Gregory S. Gewirtz/

Gregory S. Gewirtz (Reg. No. 36,522)

Counsel for Petitioner Sony Electronics Inc.

**CERTIFICATE OF COMPLIANCE
WITH TYPE-VOLUME LIMITATION**

Pursuant to 37 C.F.R. § 42.24(d), the undersigned certifies that foregoing
**PETITION FOR *INTER PARTES* REVIEW OF UNITED STATES
PATENT NO. 7,246,058 PURSUANT TO 35 U.S.C. §§ 311-319, 37 C.F.R.
§ 42** exclusive of the parts exempted as provided in 37 C.F.R. § 42.24(a), contains
13,354 words and therefore complies with the type-volume limitations of 37
C.F.R. § 42.24(a).

Respectfully submitted,

LERNER DAVID LLP

Dated: June 23, 2023

/Gregory S. Gewirtz/

Gregory S. Gewirtz (Reg. No. 36,522)

Counsel for Petitioner Sony Electronics Inc.

CERTIFICATE OF SERVICE

The undersigned hereby certifies that a copy of the foregoing **PETITION FOR *INTER PARTES* REVIEW OF UNITED STATES PATENT NO. 7,246,058 PURSUANT TO 35 U.S.C. §§ 311-319, 37 C.F.R. § 42 and Exhibits 1001-1020** is being served on June 23, 2023, via FedEx Priority Overnight on counsel of record for U.S. Pat. 7,246,058 patent owner Jawbone Innovations, LLC at the address below:

Mark Leonardo
Nutter McClennen & Fish LLP
Seaport West
155 Seaport Blvd.
Boston, MA 02210-2604

Peter Lambrianakos
FABRICANT LLP
411 Theodore Fremd Avenue
Suite 206 South
Rye, NY 10580

LERNER DAVID LLP

Dated: June 23, 2023

/Gregory S. Gewirtz/
Gregory S. Gewirtz (Reg. No. 36,522)
Counsel for Petitioner Sony Electronics Inc.