

Meta Platforms, Inc. v. Jawbone Innovations LLC

IPR2023-01291
(U.S. Patent No. 8,467,543)

Petitioner's Demonstratives
December 4, 2024



Overview of Instituted Grounds

Ground	Prior Art	'543 Patent Challenged Claims
1	Andrea, Hussain, Holzrichter	<u>1</u> -2, 5- <u>26</u>
2	Andrea, Hussain, Holzrichter, Stevens	3
3	Andrea, Hussain, Holzrichter, Alcivar	4

Issues

Claim Construction

- “voice” is not unvoiced speech
- “generating” is not calculating

Prior Art

- Independent Claims
 - Combination meets the claimed “generating” and “subtracting” in [1H]
 - Combination meets the claimed “voice activity subsystem” in [1I]

Challenged Patent

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

Challenged Patent

(12) **United States Patent**
Burnett et al.

(10) **Patent No.:** **US 8,467,543 B2**
(45) **Date of Patent:** ***Jun. 18, 2013**

(54) **MICROPHONE AND VOICE ACTIVITY DETECTION (VAD) CONFIGURATIONS FOR USE WITH COMMUNICATION SYSTEMS**

(75) **Inventors:** **Gregory C. Burnett**, Livermore, CA (US); **Nicolas J. Petit**, San Francisco, CA (US); **Alexander M. Asselly**, San Francisco, CA (US); **Andrew E. Einaudi**, San Francisco, CA (US)

(73) **Assignee:** **AliphCom**, San Francisco, CA (US)

(*) **Notice:** Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 977 days.
This patent is subject to a terminal disclaimer.

(21) **Appl. No.:** **10/400,282**

(22) **Filed:** **Mar. 27, 2003**

(65) **Prior Publication Data**
US 2003/0228023 A1 Dec. 11, 2003

Related U.S. Application Data

(60) Provisional application No. 60/368,209, filed on Mar. 27, 2002.

(51) **Int. CL**
H04B 15/00 (2006.01)

(52) **U.S. CL**
USPC 381/94.1; 381/92; 381/94.3; 381/94.7; 704/226; 704/233

(58) **Field of Classification Search**
USPC 381/94.7; 71.6; 110; 71.7; 111; 11; 381/91-92; 94.1-94.3; 704/226; 233
See application file for complete search history.

(56) **References Cited**

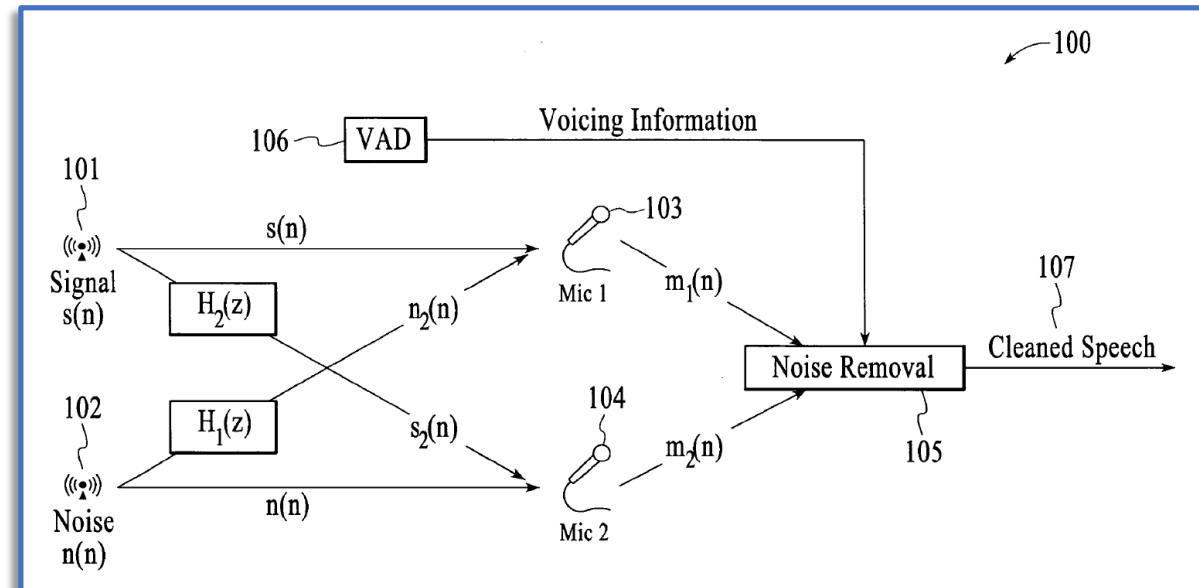
U.S. PATENT DOCUMENTS
3,789,166 A 1/1974 Sebesta
4,806,318 A 2/1977 Sebesta et al.
(Continued)

FOREIGN PATENT DOCUMENTS
EP 0 637 187 A 2/1995
EP 0 795 851 A2 9/1997
(Continued)

OTHER PUBLICATIONS
Zhao Li et al.: "Robust Speech Coding Using Microphone Arrays", Signals Systems and Computers, 1997. Conf. record of 31st Asilomar Conf., Nov. 2-5, 1997, IEEE Comput. Soc. Nov. 2, 1997. USA.
(Continued)

ABSTRACT
Communication systems are described, including both portable handset and headset devices, which use a number of microphone configurations to receive acoustic signals of an environment. The microphone configurations include, for example, a two-microphone array including two unidirectional microphones, and a two-microphone array including one unidirectional microphone and one omnidirectional microphone. The communication systems also include Voice Activity Detection (VAD) devices to provide information of human voicing activity. Components of the communications systems receive the acoustic signals and voice activity signals and, in response, automatically generate control signals from data of the voice activity signals. Components of the communication systems use the control signals to automatically select a denoising method appropriate to data of frequency subbands of the acoustic signals. The selected denoising method is applied to the acoustic signals to generate denoised acoustic signals when the acoustic signal includes speech and noise.

26 Claims, 29 Drawing Sheets



'543 Patent at Fig. 1

'543 Patent
Filed: Mar. 27, 2003
Issued: Jun. 18, 2013

Challenged Patent

1. A communications system, comprising:
a voice detection subsystem configured to receive voice activity signals that includes information associated with human voicing activity, the voice detection subsystem configured to automatically generate control signals using the voice activity signals; and
a denoising subsystem coupled to the voice detection subsystem, the denoising subsystem comprising a microphone array including a plurality of microphones, wherein a first microphone of the array is fixed at a first position relative to a mouth, wherein the first position orients a front of the first microphone towards the mouth, wherein a second microphone of the array is fixed at a second position relative to the mouth, wherein the second position orients a front of the second microphone away from the mouth such that the second position forms an angle relative to the first position, wherein the angle is greater than zero degrees, the microphone array providing acoustic signals of an environment to components of the denoising subsystem, components of

the denoising subsystem automatically selecting at least one denoising method appropriate to data of at least one frequency subband of the acoustic signals using the control signals and processing the acoustic signals using the selected denoising method to generate denoised acoustic signals, wherein the denoising method includes generating a noise waveform estimate associated with noise of the acoustic signals and subtracting the noise waveform estimate from the acoustic signal when the acoustic signal includes speech and noise,
wherein the voice detection subsystem is configured to receive the voice activity signals using a sensor independent from the microphone array and to output the control signals generated from the voice activity signals to the denoising system, the denoising system configured to use the control signals to denoise the acoustic signals from the microphone array.

'543 Patent, cl. 1.

Challenged Patent

26. A communications system, comprising:

a voice detection subsystem configured to receive voice activity signals that include information associated with human voicing activity, the voice detection subsystem configured to automatically generate control signals using the voice activity signals;

a denoising subsystem coupled to the voice detection subsystem, the denoising subsystem comprising a microphone array including a plurality of microphones, wherein a first microphone of the array is fixed at a first position relative to a mouth, wherein the first position orients a front of the first microphone towards the mouth, wherein a second microphone of the array is fixed at a second position relative to the mouth, wherein the second position orients a front of the second microphone away from the mouth such that the second position forms an angle relative to the first position, wherein the angle is greater than zero degrees, the microphone array providing acoustic signals of an environment to components of the denoising subsystem, components of the denoising subsystem automatically selecting at least one denoising method appropriate to data of at least one frequency subband of the acoustic signals using the con-

trol signals and processing the acoustic signals using the selected denoising method to generate denoised acoustic signals, wherein the denoising method includes generating a noise waveform estimate associated with noise of the acoustic signals and subtracting the noise waveform estimate from the acoustic signal when the acoustic signal includes speech and noise,

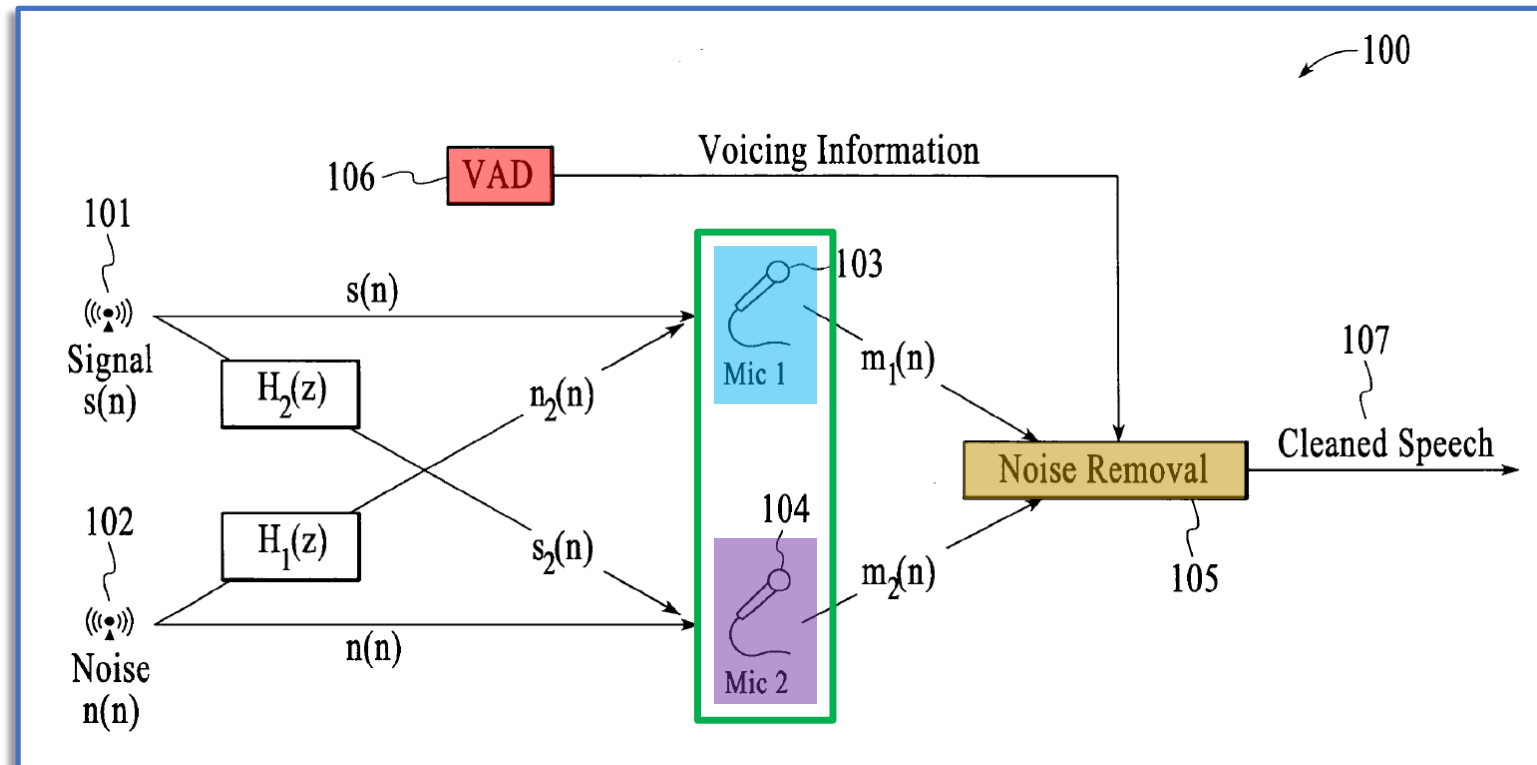
wherein the voice detection subsystem is configured to receive the voice activity signals using a sensor independent from the microphone array and to output the control signals generated from the voice activity signals to the denoising system, the denoising system configured to use the control signals to denoise the acoustic signals from the microphone array; and

a portable headset comprising the plurality of microphones and at least one speaker device, wherein the portable headset couples to at least one communication device selected from among cellular telephones, satellite telephones, portable telephones, wireline telephones, Internet telephones, wireless transceivers, wireless communication radios, personal digital assistants (PDAs), personal computers (PCs), and at least one of the voice detection subsystem and the denoising subsystem.

'543 Patent, cl. 26.

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

Claimed Embodiment



'543 Patent at Fig. 1 (annotated); Ex. 1003 ¶¶40-42.

Claim Construction

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

Parties' Proposed Constructions

**“voice detection subsystem” / “voice activity signals” /
“human voicing activity”**

Petitioner's Construction	Patent Owner's Construction
Plain and ordinary meaning	“a subsystem that detects the presence of speech, including voiced and <u>unvoiced</u> speech” “signals that indicate speech activity, including voice and <u>unvoiced</u> speech activity” “human speech activity, including voiced speech and <u>unvoiced</u> speech”

Pet. 1-6.

POR 14-16.

**Main dispute is whether “voice” must
include “unvoiced speech”**

“Voice” does not require unvoiced speech

The terms “speech” or “voice” as used herein generally refer to voiced, unvoiced, or mixed voiced and unvoiced human speech. Unvoiced speech or voiced speech is distinguished where necessary. However, the term “speech signal” or “speech”, when used as a converse to noise, simply refers to any desired portion of a signal and does not necessarily have to be human speech. It could, as an example, be music or some other type of desired acoustic information. As used in the Figures, “speech” is meant to mean any signal of interest, whether human speech, music, or anything other signal that it is desired to hear.

EX1001 ('543 Patent), 4:33-43; see also 1:47-49.

“[A] claim interpretation that excludes a preferred embodiment from the scope of the claim ‘is rarely, if ever, correct’”

On-Line Techs., Inc. v. Bodenseewerk Perkin-Elmer GmbH, 386 F.3d 1133, 1138 (Fed. Cir. 2004).

GEMS Detects “Voice”

General Electromagnetic Sensor (GEMS) VAD

The GEMS is a radiofrequency (RF) interferometer that operates in the 1-5 GHz frequency range at very low power, and can be used to detect vibrations of very small amplitude. The GEMS is used to detect vibrations of the trachea, neck, cheek, and head associated with the production of speech. These vibrations occur due to the opening and closing of the vocal folds associated with speech production, and detecting them can lead to a very accurate noise-robust VAD, as described in the Related Applications.

FIG. 10A shows an area of sensitivity 1002 on the human head appropriate for receiving a GEMS sensor, under an embodiment. The area of sensitivity 1002 further includes areas of optimal sensitivity 1004 near which a GEMS sensor can be placed to detect vibrational signals associated with voicing. The area of sensitivity 1002 along with the areas of optimal sensitivity 1004 is the same for both sides of the human head. Furthermore, the area of sensitivity 1002 includes areas on the neck and chest (not shown).

EX1001 ('543 Patent), 11:51-12:2.

GEMS Detects “Voice”

Figure 6 shows plots including a noisy audio signal (live recording) 602 along with a corresponding GEMS-based VAD signal 604, the corresponding GEMS output signal 612, and the denoised audio signal 622 following processing by the Pathfinder system using the VAD signal 604, under an embodiment. The GEMS-based VAD signal 604 was received from a trachea-mounted GEMS radiovibrometer from Aliph, San Francisco, California. The audio signal 602 was recorded using an Aliph microphone set in a babble noise environment inside a chamber measuring six (6) feet on a side and having a ceiling height of eight (8) feet. The Pathfinder system is implemented in real-time, with a delay of approximately 10 msec. The difference in the raw audio signal 602 and the denoised audio signal 622 clearly show noise suppression approximately in the range of 20-25 dB with little distortion of the desired speech signal. Thus, denoising using the GEMS-based VAD information is effective. It is clear that both the VAD signal and the denoising are effective, even though the GEMS is not detecting unvoiced speech. Unvoiced speech is normally low enough in energy that it does not significantly affect the convergence of $H_1(z)$ and therefore the quality of the denoised speech.

EX1016 ('162 Application), 16:17-31.

PO Argued GEMS is a “Voicing Sensor” in another proceeding

the patent is unaffected by environmental noise. DeJaco, ¶ 35. The preferred embodiment, for example, uses the General Electromagnetic Movement Sensor (GEMS). E.g., '058 Patent, 3:57-63. “The GEMS is a radio frequency device (2.4 GHz) that allows the detection of moving human tissue dielectric interfaces.” *Id.*, 3:64-66. As the '058 Patent explains, the GEMS “sends out weak electromagnetic waves (less than 1 milliwatt) that reflect off of whatever is around the sensor. The reflected waves are mixed with the original transmitted waves and the results analyzed for any change in position of the targets.” *Id.*, 4:1-5; *see also id.*, 3:64-4:13. DeJaco, ¶ 35. A POSITA would thus understand that the voicing sensor must be unaffected by environmental noise and the Board should construe the claim accordingly. *In re Suitco Surface, Inc.*, 603 F.3d 1255, 1260 (Fed. Cir. 2010) (“[C]laims should always be read in light of the specification and teachings in the underlying patent.”).

IPR2022-00623, Paper 17, p. 8

Reply, 4-5.

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

14

Dependent Claims Confirm “Voice” is “Voice”

PO Ignores Express Limitations in Dependent Claims 2-4

2. The system of claim 1, wherein the voice detection subsystem further comprises:

at least one glottal electromagnetic micropower sensor (GEMS) including at least one antenna for receiving the voice activity signals; and

at least one voice activity detector (VAD) algorithm for processing the GEMS voice activity signals and generating the control signals.

3. The system of claim 1, wherein the voice detection subsystem further comprises:

at least one accelerometer sensor in contact with skin for

receiving the voice activity signals; and
at least one voice activity detector (VAD) algorithm for processing the accelerometer sensor voice activity signals and generating the control signals.

4. The system of claim 1, wherein the voice detection subsystem further comprises:

at least one skin-surface microphone sensor in contact with skin for receiving the voice activity signals; and
at least one voice activity detector (VAD) algorithm for processing the skin-surface microphone sensor voice activity signals and generating the control signals.

EX1001 ('543 Patent), 24:38-59.

Perfect Speech Detector is Not Required

'512 Patent does not require perfect VAD for all speech types

“[we] normally do not interpret claim terms in a way that excludes embodiments disclosed in the specification”

Oatey Co. v. IPS Corp., 514 F.3d 1271, 1276 (Fed. Cir. 2008).

“there is a strong presumption against a claim construction that excludes a disclosed embodiment”

Nobel Biocare Servs. AG v. Intradent USA, Inc., 903 F.3d 1365, 1381 (Fed Cir. 2018)
citing *In re Katz Interactive Call Processing Patent Litig.*, 639 F.3d 1303, 1324 (Fed. Cir. 2011)

PO Quotes the Patent Out of Context

can be correctly calculated. The '543 Patent teaches that it is crucial to implement “a sufficiently accurate VAD . . . to adapt the coefficients of H_1 (noise *only*) and H_2 (if needed, when *speech* is being produced).” '543 Patent, 16:55-67 (emphasis added). Otherwise, if the voice detection system cannot detect certain types of speech, such as unvoiced speech, during the calculation of the coefficients of H_1 (noise only), it “would result in the removal of a large part of the speech energy from the signal of interest,” resulting in “signal distortion and reduction (de-signaling).”

Id. Based on these teachings, the “voice activity” relevant to the claims must include all speech, including voiced speech and unvoiced speech. If “voice activity”

POR, 15.

Regarding the use of VAD to control adaptation of the noise suppression system to the received signals, classical ANC uses no VAD information. Since, during speech production, there is signal in the reference microphone, adapting the coefficients of $H_1(z)$ (the path from the noise to the primary microphone) during the time of speech production would result in the removal of a large part of the speech energy from the signal of interest. The result is signal distortion and reduction (de-signaling). Therefore, the various methods described above use VAD information to construct a sufficiently accurate VAD to instruct the Pathfinder system when to adapt the coefficients of H_1 (noise only) and H_2 (if needed, when speech is being produced).

EX1001 ('543 Patent), 16:55-67.

Parties' Proposed Constructions

“generating a noise waveform estimate associated with noise of the acoustic signals”

Petitioner's Construction	Patent Owner's Construction
Plain and ordinary meaning	<u>“calculating</u> a noise estimate <u>based on</u> the noise associated with <u>all</u> the acoustic signals”

Pet. 6-10.

POR 16-23.

PO rewrites “generate” as “calculate”

1. A communications system, comprising:
a voice detection subsystem configured to receive voice activity signals that includes information associated with human voicing activity, the voice detection subsystem configured to automatically generate control signals using the voice activity signals; and
a denoising subsystem coupled to the voice detection subsystem, the denoising subsystem comprising a microphone array including a plurality of microphones, wherein a first microphone of the array is fixed at a first position relative to a mouth, wherein the first position orients a front of the first microphone towards the mouth, wherein a second microphone of the array is fixed at a second position relative to the mouth, wherein the second position orients a front of the second microphone away from the mouth such that the second position forms an angle relative to the first position, wherein the angle is greater than zero degrees, the microphone array providing acoustic signals of an environment to components of the denoising subsystem, components of

the denoising subsystem automatically selecting at least one denoising method appropriate to data of at least one frequency subband of the acoustic signals using the control signals and processing the acoustic signals using the selected denoising method to generate denoised acoustic signals, wherein the denoising method includes generating a noise waveform estimate associated with noise of the acoustic signals and subtracting the noise waveform estimate from the acoustic signal when the acoustic signal includes speech and noise,
wherein the voice detection subsystem is configured to receive the voice activity signals using a sensor independent from the microphone array and to output the control signals generated from the voice activity signals to the denoising system, the denoising system configured to use the control signals to denoise the acoustic signals from the microphone array.

'543 Patent, cl. 1.

PO's expert inconsistently interprets "generate"

1. A communications system, comprising:
a voice detection subsystem configured to receive voice activity signals that includes information associated with human voicing activity, the voice detection subsystem configured to automatically generate control signals using the voice activity signals; and
a denoising subsystem coupled to the voice detection subsystem, the denoising subsystem comprising a microphone array including a plurality of microphones, wherein a first microphone of the array is fixed at a first position relative to a mouth, wherein the first position orients a front of the first microphone towards the mouth, wherein a second microphone of the array is fixed at a second position relative to the mouth, wherein the second position orients a front of the second microphone away from the mouth such that the second position forms an angle relative to the first position, wherein the angle is greater than zero degrees, the microphone array providing acoustic signals of an environment to components of the denoising subsystem, components of

the denoising subsystem automatically selecting at least one denoising method appropriate to data of at least one frequency subband of the acoustic signals using the control signals and processing the acoustic signals using the selected denoising method to generate denoised acoustic signals, wherein the denoising method includes generating a noise waveform estimate associated with noise of the acoustic signals and subtracting the noise waveform estimate from the acoustic signal when the acoustic signal includes speech and noise,
wherein the voice detection subsystem is configured to receive the voice activity signals using a sensor independent from the microphone array and to output the control signals generated from the voice activity signals to the denoising system, the denoising system configured to use the control signals to denoise the acoustic signals from the microphone array.

'543 Patent, cl. 1.

PO's expert inconsistently interprets "generate"

1. A communications system, comprising:
a voice detection subsystem configured to receive voice activity signals that includes information associated with human voicing activity, the voice detection subsystem configured to automatically generate control signals using the voice activity signals; and
a denoising subsystem coupled to the voice detection subsystem, the denoising subsystem comprising a microphone array including a plurality of microphones, wherein a first microphone of the array is positioned at a first position relative to a mouth, where the first microphone orients a front of the first microphone towards the mouth, wherein a second microphone of the array is fixed at a second position relative to the mouth, where the second microphone orients a front of the second microphone away from the mouth such that the second microphone forms an angle relative to the front of the mouth, where the angle is greater than zero degrees, the microphone array providing acoustic signals to the voice detection subsystem, and the denoising subsystem comprising components of the denoising subsystem

the denoising subsystem automatically selecting at least one denoising method appropriate to data of at least one frequency subband of the acoustic signals using the control signals and processing the acoustic signals using the selected denoising method to generate denoised acoustic signals, wherein the denoising method includes generating a noise waveform estimate associated with noise of the acoustic signals and subtracting the noise waveform estimate from the acoustic signal when the acoustic

14 Q. Okay. To make the record clear, when the
15 claim -- Claim 1 refers to automatically generate
16 control signals, is that the same as automatically
17 calculate control signals?
18 A. Not necessarily. I think what I stated was
19 more in the context of generating a noise waveform
20 estimate.

Ex. 1018, 67:14-20; see Reply, 8-9.

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

21

PO's expert inconsistently interprets "generate"

1. A communications system, comprising:
a voice detection subsystem configured to receive voice activity signals that includes information associated with human voicing activity, the voice detection subsystem configured to automatically generate control signals using the voice activity signals; and
a denoising subsystem coupled to the voice detection subsystem, the denoising subsystem comprising a microphone array including a plurality of microphones, wherein a first microphone of the array is positioned relative to a mouth, where the array orients a front of the first microphone towards the mouth, wherein a second microphone of the array is fixed at a second position relative to the first position, the second position orients a front of the second microphone away from the mouth such that the microphone forms an angle relative to the front of the mouth, the angle is greater than zero degrees, the array providing acoustic signals and the denoising subsystem processing the acoustic signals to generate denoised acoustic signals.

the denoising subsystem automatically selecting at least one denoising method appropriate to data of at least one frequency subband of the acoustic signals using the control signals and processing the acoustic signals using the selected denoising method to generate denoised acoustic signals, wherein the denoising method includes generating a noise waveform estimate associated with noise of the acoustic signals and subtracting the noise waveform estimate from the acoustic signal when the acoustic signal is in the frequency subband.

21 Q. What about where Claim 1 recites, "Generate
22 denoised acoustic signals"? Is that the same as
23 calculate denoised acoustic signals?

24 A. Yes. Correcting that, yes, because that will
25 involve removing of a noise estimate, and that would
1 involve -- that would be a calculation as well.

Ex. 1018, 67:21-68:1; see Reply, 8-9.

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

22

PO Selectively Narrows

“[i]n the absence of any evidence to the contrary, we must presume that the use of these different terms in the claims connotes different meanings”

CAE Screenplates, Inc. v. Heinrich Fiedler GmbH & Co. KG, 224 F.3d 1308, 1317 (Fed. Cir. 2000).

“[w]hile the ‘protrusion’ and ‘end plate’ structures must be connected in some fashion, the use of these ‘two terms in a claim requires that they connote different meanings’”

Neville v. Found. Constructors, Inc., 972 F.3d 1350, 1357 (Fed. Cir. 2020).

PO Selectively Narrows

$$S(z) = \frac{M_1(z) - M_2(z)H_1(z)}{1 - M_2(z)H_1(z)}$$

$$S(z) \approx M_1(z) - M_2(z)H_1(z)$$

'543 Patent, 15:50-16:3.

PO Selectively Narrows

“[w]e...expressly rejected the contention that if a patent describes only a single embodiment, the claims of the patent must be construed as being limited to that embodiment”

Cont’l Circuits LLC v. Intel Corp., 915 F.3d 788, 797 (Fed. Cir. 2019).

“[e]ven when the specification describes only a single embodiment, the claims of the patent will not be read restrictively ...”

Liebel-Flarsheim Co. v. Medrad, Inc., 358 F.3d 898, 906 (Fed. Cir. 2004)
quoting *Teleflex, Inc. v. Ficosa N. Am. Corp.*, 299 F.3d 1313, 1327 (Fed. Cir. 2002).

PO's expert inconsistently interprets "the acoustic signals"

1. A communications system, comprising:
a voice detection subsystem configured to receive voice activity signals that includes information associated with human voicing activity, the voice detection subsystem configured to automatically generate control signals using the voice activity signals; and
a denoising subsystem coupled to the voice detection subsystem, the denoising subsystem comprising a microphone array including a plurality of microphones, wherein a first microphone of the array is fixed at a first position relative to a mouth, wherein the first position orients a front of the first microphone towards the mouth, wherein a second microphone of the array is fixed at a second position relative to the mouth, wherein the second position orients a front of the second microphone away from the mouth such that the second position forms an angle relative to the first position, wherein the angle is greater than zero degrees, the microphone array providing acoustic signals of an environment to components of the denoising subsystem, components of

the denoising subsystem automatically selecting at least one denoising method appropriate to data of at least one frequency subband of the acoustic signals using the control signals and processing the acoustic signals using the selected denoising method to generate denoised acoustic signals, wherein the denoising method includes generating a noise waveform estimate associated with noise of the acoustic signals and subtracting the noise waveform estimate from the acoustic signal when the acoustic signal includes speech and noise,
wherein the voice detection subsystem is configured to receive the voice activity signals using a sensor independent from the microphone array and to output the control signals generated from the voice activity signals to the denoising system, the denoising system configured to use the control signals to denoise the acoustic signals from the microphone array.

'543 Patent, cl. 1.

PO's expert inconsistently interprets "the acoustic signals"

1. A communications system, comprising:
a voice detection subsystem configured to receive voice activity signals that includes information associated with human voicing activity, the voice detection subsystem configured to automatically generate control signals using the voice activity signals; and
a denoising subsystem coupled to the voice detection subsystem, the denoising subsystem comprising a microphone array including a plurality of microphones wherein a first microphone of the array is fixed at a position relative to a mouth, wherein the first position orients a front of the first microphone towards the mouth, wherein a second microphone of the array is fixed at a second position relative to the mouth, wherein the second position orients a front of the second microphone away from the mouth such that the second position forms an angle relative to the first position, wherein the angle is greater than zero degrees, the microphone array providing acoustic signals of an environment, components of the denoising subsystem, components

the denoising subsystem automatically selecting at least one denoising method appropriate to data of at least one frequency subband of the acoustic signals using the control signals and processing the acoustic signals using the selected denoising method to generate denoised acoustic signals, wherein the denoising method includes generating a noise waveform estimate associated with noise

6 When Claim 1 recites "generate denoised
7 acoustic signals," is that referring to generating
8 denoised acoustic signals for all of the acoustic
9 signals provided by the microphone array?

10 MR. COWELL: Objection to form.

11 THE WITNESS: No. It was -- no, it's not
12 referring to all the signal in the microphone array. It
13 is specifically referring to the signals in the
14 different subbands, the noise signals of the different
15 subbands off the main signal, which includes primarily
16 the signal of interest.

PO rewrites “associated with” as “based on”

21 Q. Is it fair to say, then, that "based on" is
22 narrower than "associated with"?

23 MR. COWELL: Objection to form.

24 THE WITNESS: I think that's fair to say.

25 Maybe. I mean, that's so general that I -- yeah.

Ex. 1018, 101:21-25; see Reply, 10

Asserted Prior Art

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

29

Andrea Overview

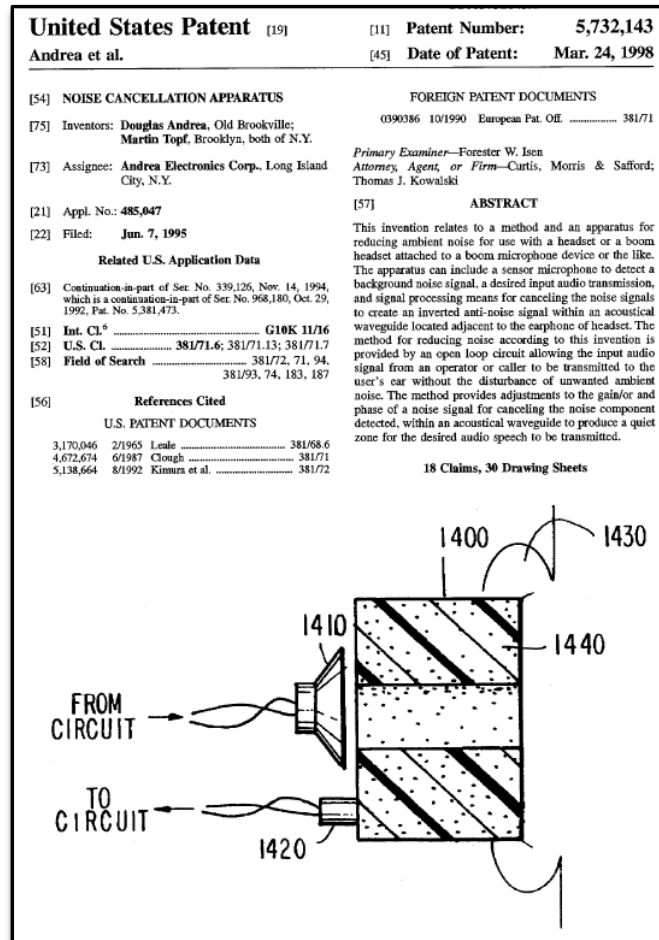


FIG. 2 is a block diagram of the noise cancellation apparatus used in the telephone of FIG. 1;

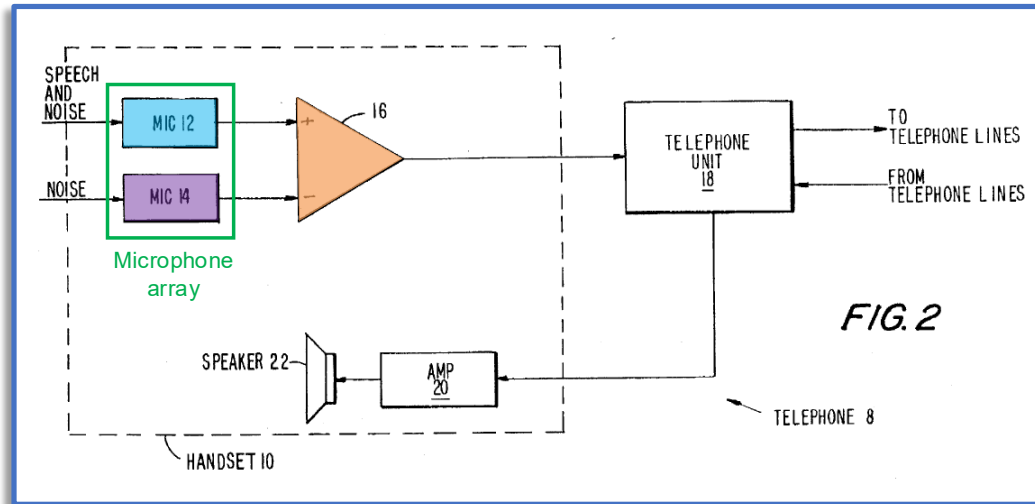
It is yet another object of the present invention to provide a relatively low-cost noise reduction and cancellation apparatus for use with telecommunication systems which is operable with standard available on-line power.

A still further object of the present invention is to provide a relatively low-cost noise cancellation apparatus which is readily adaptable to handsets of existing communication systems and which is operable with standard available on-line power.

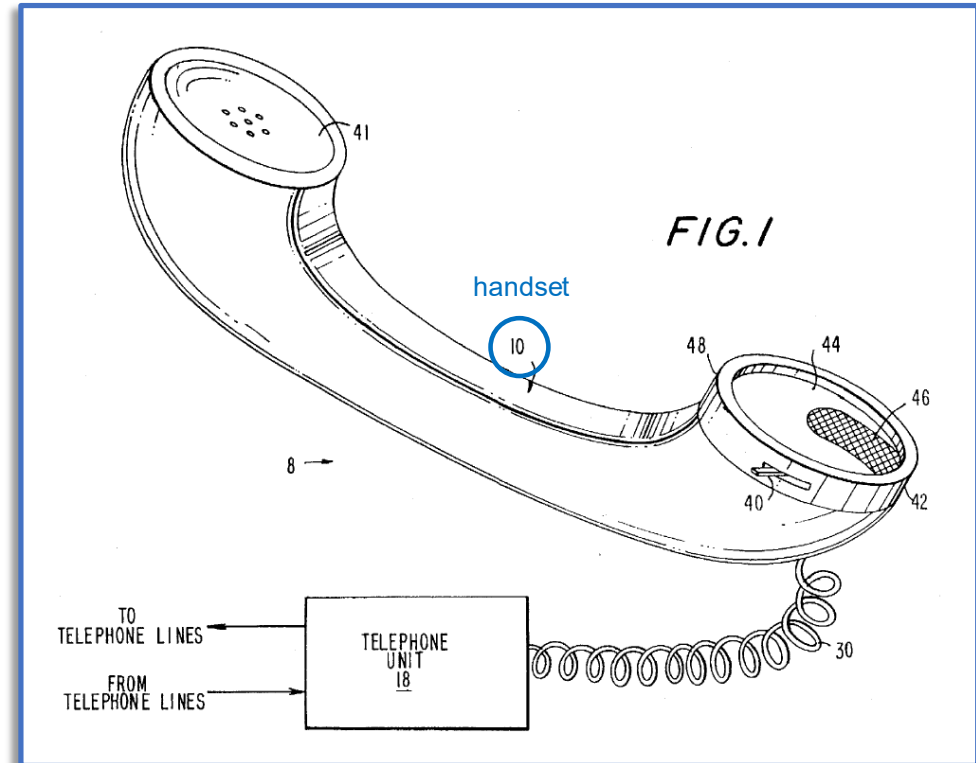
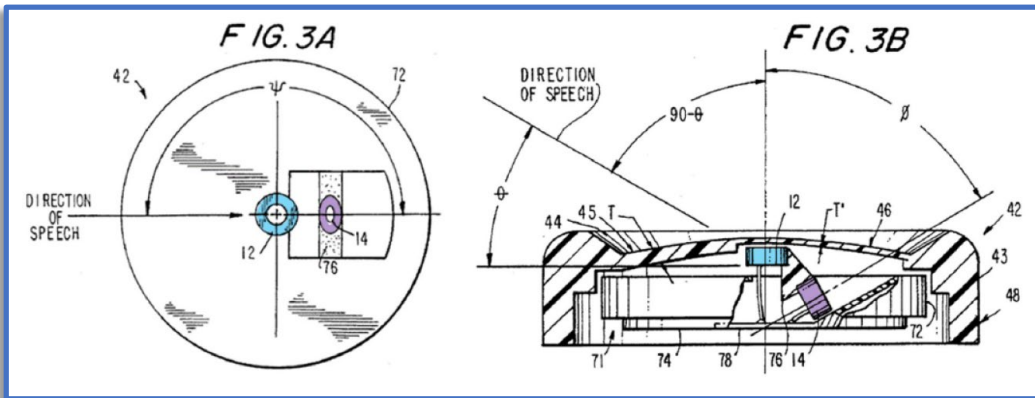
The active noise reduction apparatus as well as the noise cancellation apparatus can be used in any telecommunication systems that are used in flight (e.g., helicopter or airplane) or in other settings such as telephones, or voice recognition and/or verification systems for instance, for access to a physical facility or to a computer (either via direct or indirect interface or via telephone lines) or to an automatic teller machine or, in other recognition and/or verification systems.

EX1005 (Andrea) at 11:19-20; 6:58-61; 7:9-13; 11:19-20.

Andrea discloses a voice detection subsystem



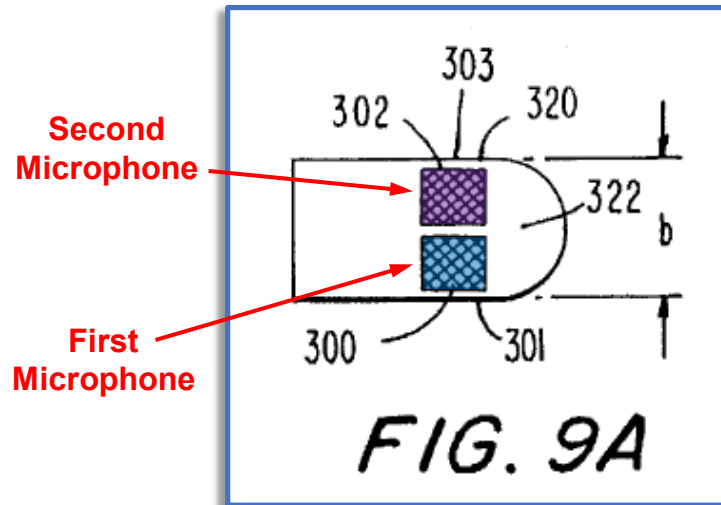
EX1005 (Andrea) at Fig. 2 .



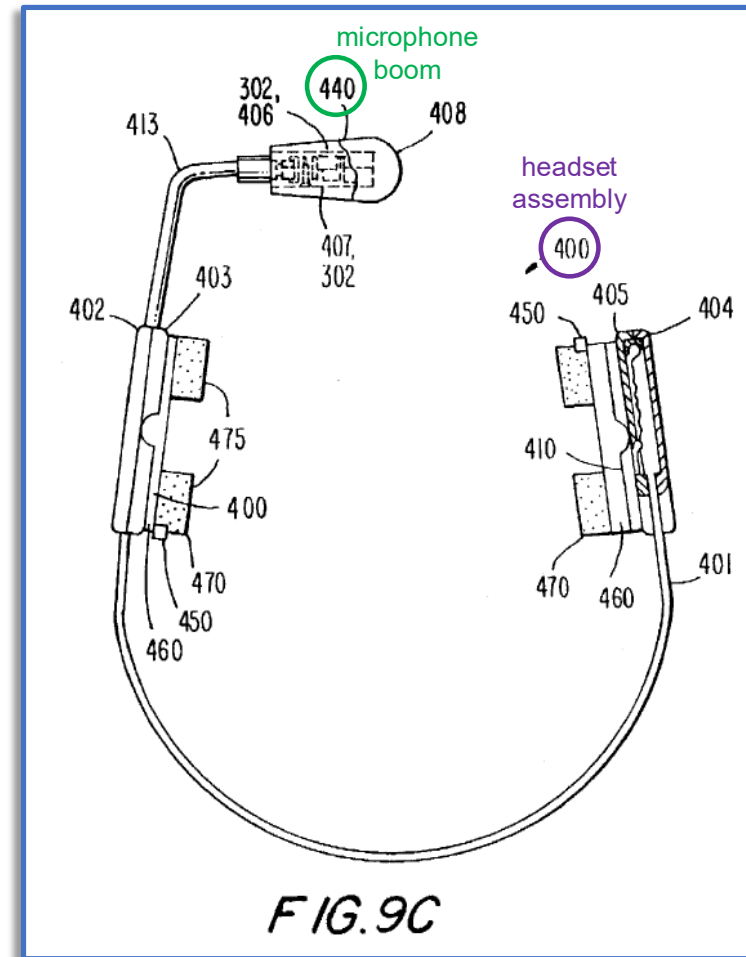
EX1005 (Andrea) at Fig. 1.

EX1005 (Andrea) at Figs. 3A, 3B.

Andrea discloses a voice detection subsystem



EX1005 (Andrea) at Fig. 9A.



EX1005 (Andrea) at Fig. 9C.

Hussain Overview

A NEW METRIC FOR SELECTING SUB-BAND PROCESSING IN ADAPTIVE SPEECH ENHANCEMENT SYSTEMS

Amir Hussain, Douglas R. Campbell and Thomas J. Moir
Department of Electronic Engineering and Physics,
University of Paisley, High St., Paisley PA1 2BE, Scotland U.K.
Corresponding author's email: huss_ae0@paisley.ac.uk

ABSTRACT

A multi-microphone adaptive speech enhancement system employing diverse sub-band processing is presented. A new robust metric is developed, which is capable of real-time implementation, in order to automatically select the best form of processing within each sub-band. It is based on an adaptively estimated inter-channel Magnitude Squared Coherence (MSC) relationship, which is used to detect the level of correlation between in-band signals from multiple sensors during noise-alone periods in intermittent speech. This paper reports recent results of comparative experiments with simulated anechoic data extended to include simulated reverberant data. The results demonstrate that the method is capable of significantly outperforming conventional noise cancellation schemes.

1. INTRODUCTION

Background noise contamination of speech signals reduces the Signal to Noise Ratio (SNR) of hands-free telephones, portable phones, and security screens. Speech recognition systems in particular, are known to experience problems due to levels of background noise that are in fact, considered quite acceptable to human listeners [7]. In addition to the noise level, the presence of multiple noise sources, reverberant environments, moving noise sources and statistically non-stationary noise sources considerably complicates the situation.

Classical speech enhancement methods based on full-band multi-microphone noise cancellation implementations which attempt to model acoustic path transfer functions can produce excellent results in anechoic environments with localized sound radiators [6], however performance deteriorates in reverberant environments. Multi-band processing has been found to be important in combating reverberation effects [10] [11]. Adaptation is necessary to compensate for changing noise fields [2] [10] due to for example, non-Gaussian sources, source/sensor motion, or time-varying acoustic paths. Multi-sensor methods are necessary to compensate for reverberation [13] and speech/noise spectral overlap [2].

Studies on noise in automobiles by various researchers [9] [10] [11] have found that the long-term noise correlation between two microphone locations was high for frequencies below 500Hz, and decreased gradually with virtually no correlation above 2kHz. Wallace and Goubran [12] applied the classical full-band linear FIR based correlated noise canceller, adapted using the Least Mean Square (LMS) algorithm, to recorded automobile noise and obtained significant noise

reduction in the low frequency range but at high frequencies, where both the correlation and noise energy were low, the noise increased. In some situations, as reflected in the measurements from office and automobile environments recently reported in [2], significant short term correlation may also occur at high frequencies depending upon the relative locations of the microphones and the nature of the noise sources. The above evidence implies that processing appropriate in one sub-band, may not be so in another. This therefore supports a general approach involving the use of diverse processing in frequency bands dependent on the correlation between the in-band signals from multiple sensors.

Van Compernelle et al [5] and Dabis et al [6] used closely spaced microphones in a full-band adaptive noise cancellation scheme involving the identification of a differential acoustic path transfer function during a noise only period in intermittent speech. A Multi-Microphone Sub-Band Adaptive (MMSBA) speech enhancement system has been described which extends this method by applying it within a set of sub-bands provided by a filterbank [2] [8] [11]. Even non-optimised MMSBA processing has shown the potential to yield more than 6dB SNR improvements over conventional full-band methods in real reverberant environments.

The MMSBA system has been further developed by employing diverse Sub-Band Processing (SBP) in order to allow inter-channel features within the sub-bands to influence the subsequent processing [4]. In order to realize this, a robust practical metric has been developed, which is capable of real-time implementation, based on the inter-channel Magnitude Squared Coherence (MSC) relationship in order to automatically select the best SBP option. The choice of SBP options include: (i) no processing; (ii) intermittent coherent noise canceller; and (iii) continuous incoherent noise canceller.

2. THE PROPOSED MMSBA SYSTEM INCORPORATING THE NEW METRIC

Two or more relatively closely spaced microphones may be used in an adaptive noise cancellation scheme [4] [5] [6] [11] to identify a differential acoustic path transfer function during a noise only period in intermittent speech. The extension of this work, termed the Multi-Microphone sub-band Adaptive (MMSBA) speech enhancement system, applies the method within a set of sub-bands provided by a filter bank as shown in Figure 1. The filter bank can be implemented using various orthogonal transforms or by a parallel filter bank approach. The sub-bands can be distributed either in a linear or a non-linear fashion.

ESCA, Eurospeech97, Rhodes, Greece. ISSN 1010-4074, Page 2611

2. Intermittent coherent noise canceller: If the noise power is significant and the noise between the two channels is significantly correlated in a sub-band, then perform adaptive intermittent noise cancellation, wherein an adaptive filter may be determined which models the differential acoustic-path transfer function between the microphones during the noise alone period. This can then be used in a noise cancellation format during the speech plus noise period to process the noisy speech signal. This scheme illustrated in Figure 1 can be described mathematically as follows. Assuming N , S , P , R represent the z-transforms of the noise signal, speech signal, primary signal and reference signal, respectively. The primary and reference signals in each sub-band are thus

$$P = B(S + H_1 N) \quad ; \quad R = B(S + H_2 N)$$

The transformed error signal is thus,

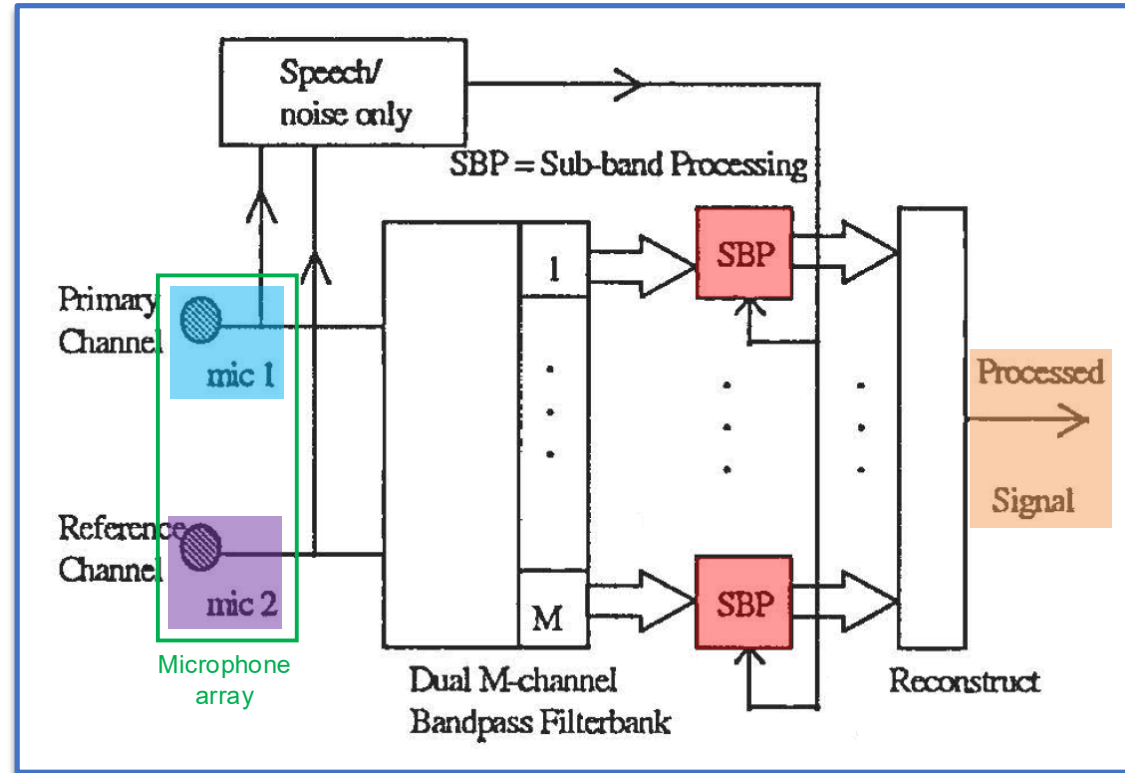
$$E = B[(1 - H_3)S + (H_1 - H_3 H_2)N]$$

which is a frequency domain error, weighted by the band-limiting transfer function B , and H_3 represents the sub-band adaptive filter. The Mean Squared Error (MSE) function is,

$$J_E = (2\pi j)^{-1} \oint_{|z|=1} E \cdot E^* \cdot z^{-1} dz$$

EX1006 (Hussain) at 11.

Hussain discloses adaptive noise cancellation



EX1006 (Hussain) at Fig. 1.

Holzrichter Overview

United States Patent [19]

Holzrichter

[11] Patent Number:

6,006,175

[45] Date of Patent:

Dec. 21, 1999

[54] METHODS AND APPARATUS FOR NON-ACOUSTIC SPEECH CHARACTERIZATION AND RECOGNITION

[75] Inventor: John F. Holzrichter, Berkeley, Calif.

[73] Assignee: The Regents of the University of California, Oakland, Calif.

[21] Appl. No.: 08/597,596

[22] Filed: Feb. 6, 1996

[51] Int. Cl.⁷ G10L 3/02

[52] U.S. Cl. 704/208; 704/205; 704/206; 704/207

[58] Field of Search 395/2.1, 2.16, 395/2.27, 2.37, 2.67, 2.17, 2.15, 2.14; 704/205-208, 201, 218, 228, 258

[56] References Cited

U.S. PATENT DOCUMENTS

2,193,102	3/1940	Koch	250/6
2,539,594	1/1951	Rines et al.	250/17
2,823,365	2/1958	Rines	340/6
3,555,188	1/1971	Meacham	381/115
3,699,856	10/1972	Chabot et al.	95/1.1
3,925,774	12/1975	Amlung	340/258
4,027,303	5/1977	Neuwirth et al.	340/258
4,092,493	5/1978	Rabiner et al.	179/1
4,260,229	4/1981	Bloomstein	352/50
4,461,025	7/1984	Franklin	381/56
4,621,348	11/1986	Tender	367/116
4,769,845	9/1988	Nakamura	381/43
4,783,803	11/1988	Baker et al.	381/42
4,803,729	2/1989	Baker	381/43
4,882,746	11/1989	Shimada	455/462
4,903,305	2/1990	Gillick et al.	381/41
4,914,703	4/1990	Gillick	381/43
5,008,941	4/1991	Sejnowa	388/143
5,027,406	6/1991	Roberts et al.	381/43
5,030,956	7/1991	Murphy	342/22
5,127,055	6/1992	Larkey	381/43
5,202,952	4/1993	Gillick et al.	395/2
5,227,797	7/1993	Murphy	342/22
5,280,563	1/1994	Ganong	395/2
5,337,394	8/1994	Sejnowa	395/2.5

OTHER PUBLICATIONS

Rabiner, L. R. "Applications of Voice Processing to Telecommunications", Proc. of the IEEE, 82(2), 199-228 (Feb. 1994).

Skolnik, M.J. (ed.) "Radar Handbook 2nd ed." McGraw-Hill, page v (1990).

Waynant, R. W. and Ediger, M. N. (eds.) "Electro-Optics Handbook", McGraw-Hill, p. 24.22 (1994).

Flanagan, J. L. "Speech Analysis Synthesis, and Perception", Academic Press NY, pp. 8, 16-20, 154-156 (1965).

Coker, C.H. "A Model of Articulatory Dynamics and Control", Proc. IEEE, 64(4), 452-459 (1976).

Javkin, H. et al "Multi-Parameter Speech Training System" Speech and Language Technology for Disabled Persons, Proceedings of a European Speech Communication Association (ESCA) Workshop, Stockholm, Sweden, 137-140 (May 31, 1993).

(List continued on next page.)

Primary Examiner—Tariq R. Hafiz

Attorney, Agent, or Firm—John P. Wooldridge

[57] ABSTRACT

By simultaneously recording EM wave reflections and acoustic speech information, the positions and velocities of the speech organs as speech is articulated can be defined for each acoustic speech unit. Well defined time frames and feature vectors describing the speech, to the degree required, can be formed. Such feature vectors can uniquely characterize the speech unit being articulated each time frame. The onset of speech, rejection of external noise, vocalized pitch periods, articulator conditions, accurate timing, the identification of the speaker, acoustic speech unit recognition, and organ mechanical parameters can be determined.

48 Claims, 31 Drawing Sheets

ii) The EM sensor signals provide explicit evidence of specific articulator activities which are very difficult to reliably extract from the acoustic signal. Many speech organ

information. By using a single EM sensor to measure the presence or absence of vocal fold motions, one can provide valuable information to a speech recognition system to determine if voiced speech is occurring. Similarly a single

FIG. 1. By comparison, a prior art CASR system forms a feature vector solely from acoustic information, while the NASR system of FIG. 2 makes EM sensor measurements of a number of speech organs (vocal fold, velum, tongue and lips are shown) and then combines feature vectors describing these signals with feature vectors describing acoustic information to produce a more complex, more information-rich feature vector describing the speech unit time frame.

FIG. 4 shows a view of a head with relevant speech organs and an illustrative NASR sensor 20. Three EM sensor transmit/receive modules 21, 22, 23 are shown although there can be as few as one or more than three. Also a microphone 24 is used to obtain simultaneous acoustic information. Sensor 20 includes the necessary electronics, increasingly less connected together. EM sensor ranging speech detection modules may be located against, near, separate from the body, and attached to or near microphones (either hand held, attached to harnesses, or as normally used on microphone stands), or included in head helmets (partial or full head covering), in mouth covers, and in other devices attached to or on parts of the body near to the head.

EX1007 (Holzrichter) at 10:35-37, 12:3-6, 16:42-49, 17:53-58, 18:42-48.

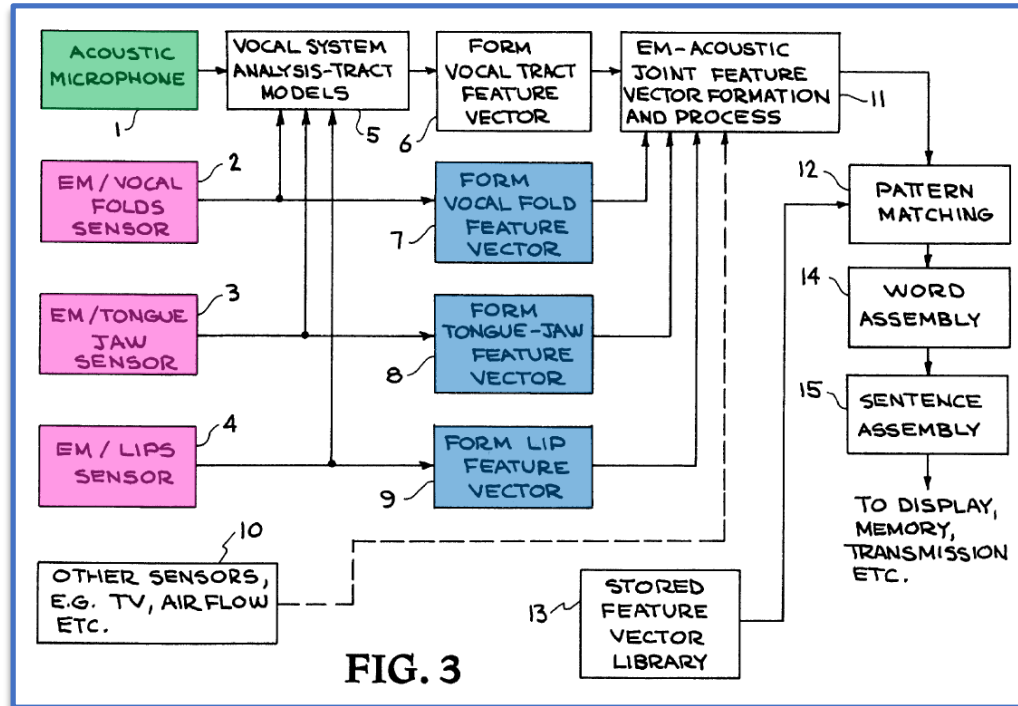
Pet, 10-12.

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

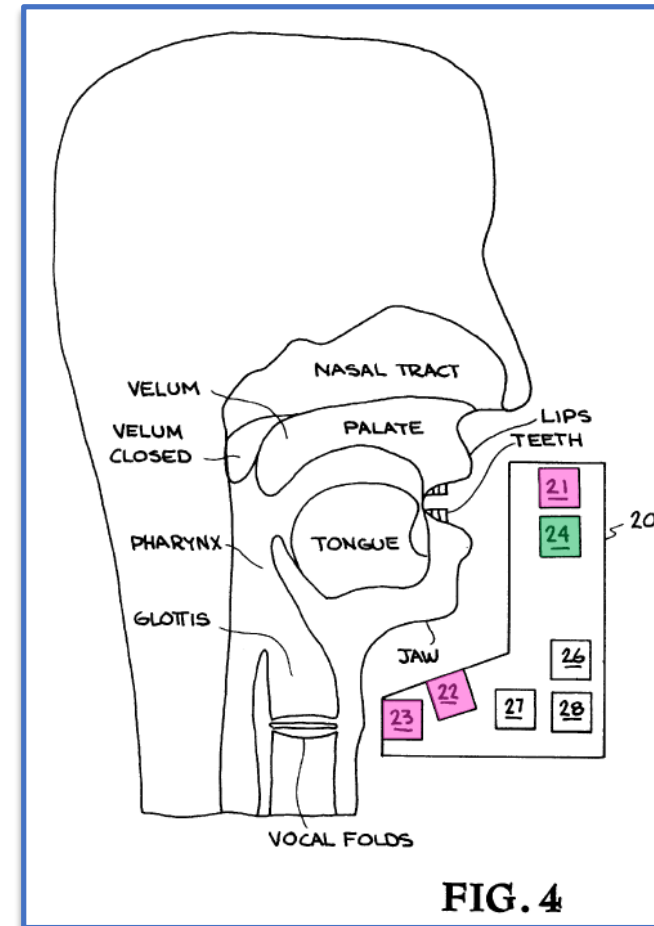
35

Meta Platforms, Inc. Exhibit 1019
Page 35 of 44

Holzrichter discloses EM sensors



EX1007 (Holzrichter) at Fig. 3.



EX1007 (Holzrichter) at Fig. 4.

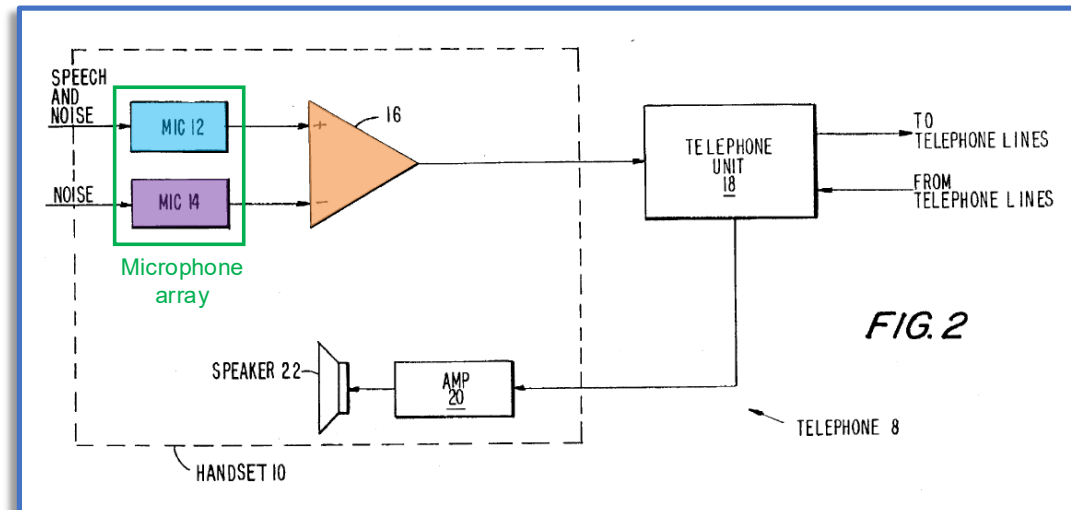
[1H]: Andrea renders obvious “generating” and “subtracting”

Acoustic signals composed of speech or the like and background noise are supplied to the first microphone 12 and converted therein into a corresponding electrical signal which is thereafter supplied to the plus terminal of the op-amp 16. The background noise is supplied to the second microphone 14 and converted therein into a corresponding electrical signal which is thereafter supplied to the minus terminal of the op-amp 16. The op-amp 16 is adapted to subtract the noise signal from the second microphone 14 from the speech and noise signal from the first microphone 12 and to supply therefrom an electrical signal representing substantially the speech to the telephone unit 18 whereupon the speech signal is transmitted therefrom through the tele-

180°. As a result of arranging the first and second microphones in this manner, the first microphone 12 receives both the speech from the operator and the background acoustic noise which is present in the vicinity, and the second microphone 14 essentially receives only the same background acoustic noise which is received by the first microphone.

op-amp 16. The op-amp 16 subtracts the signal received from the second microphone 14 from that received from the first microphone 12 and outputs the resulting subtracted signal. DC levels which may be present in the output signal

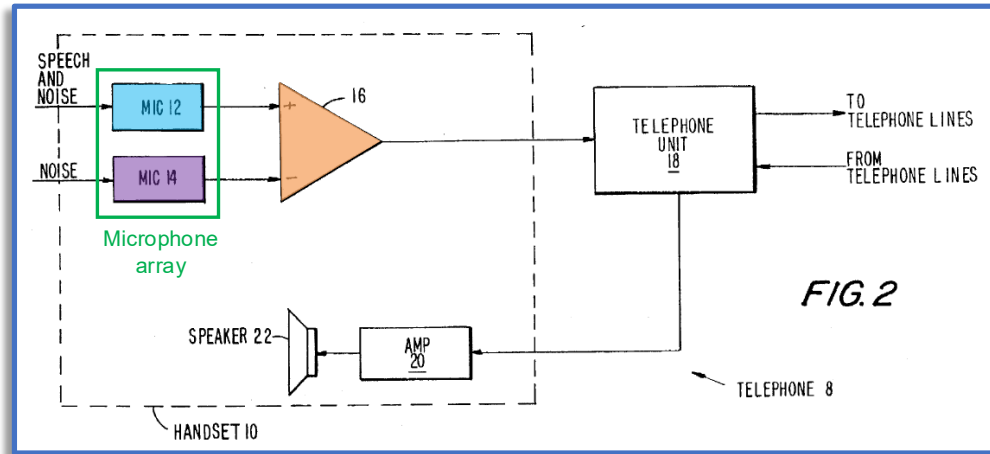
[1H]: “...generating a noise waveform estimate associated with noise of the acoustic signals and subtracting the noise waveform estimate from the acoustic signal when the acoustic signal includes speech and noise”



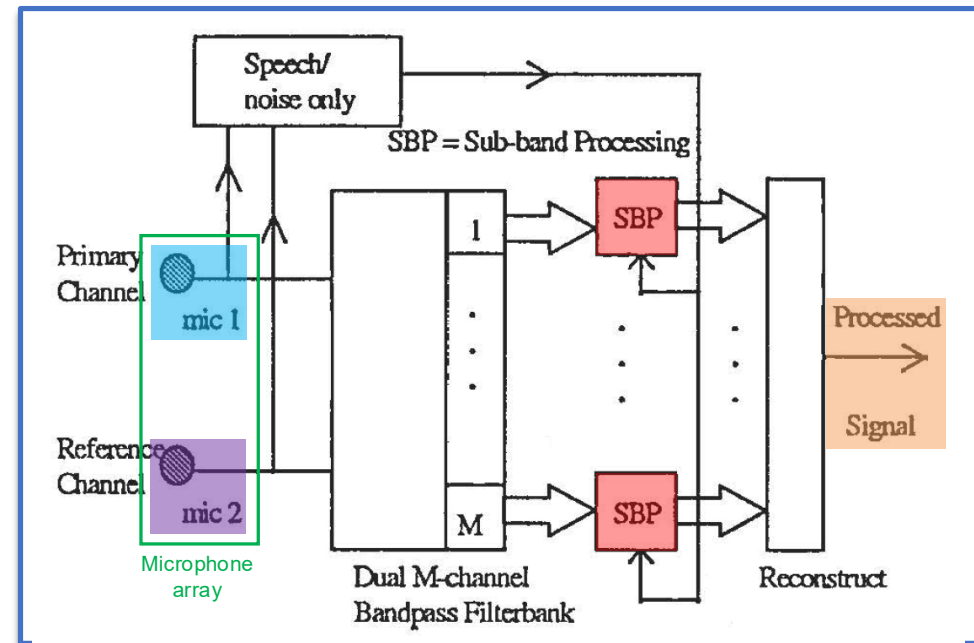
EX1005 (Andrea) at Fig. 2.

EX1005 (Andrea) at
12:55-66; 14:23-29; 16:20-23.

[1H]: Combination renders obvious “calculate” under PO’s construction



EX1005 (Andrea) at Fig. 2.



EX1006 (Hussain) at Fig. 1.

[1H]: Combination renders obvious “calculate” under PO’s construction

Abstract—Analog-to-digital converters (ADC’s) are ubiquitous, critical components of software radio and other signal processing systems. This paper surveys the state-of-the-art of ADC’s, including experimental converters and commercially available parts.

EX1017 (Walden) at Abstract.

[1H]: Combination renders obvious “calculate” under PO’s construction

3 Q. So your testimony is that the scope of the
4 independent claims of '543 patent are broad enough to
5 encompass calculation occurring in either the analog or
6 digital domain?

7 MR. COWELL: Objection to form.

8 THE WITNESS: Yeah, I think the basic concepts
9 are -- and the teachings would apply in both contexts.

EX1018 (Sayeed Tr.), 128:3-9.

[1H]: Combination renders obvious “calculate” under PO’s construction

14 Q. And, in your opinion, that op amp
15 then would subtract the signal coming from
16 mic 14 from the signal from mic 12; is that
17 right?
18 A. Yes.

EX2001 (Anderson Tr.) at 17:14-18.

3 Q. What does the actual subtracting
4 in your combination?

5 A. So there are two alternatives for
6 subtracting. The first, it's obvious, and
7 we've already discussed that Andrea does
8 subtracting.

9 The second is that we combine
10 Andrea with Hussain. Hussain does the same
11 subtracting except that he applies this
12 adaptive gain to the reference signal for
13 subtraction.

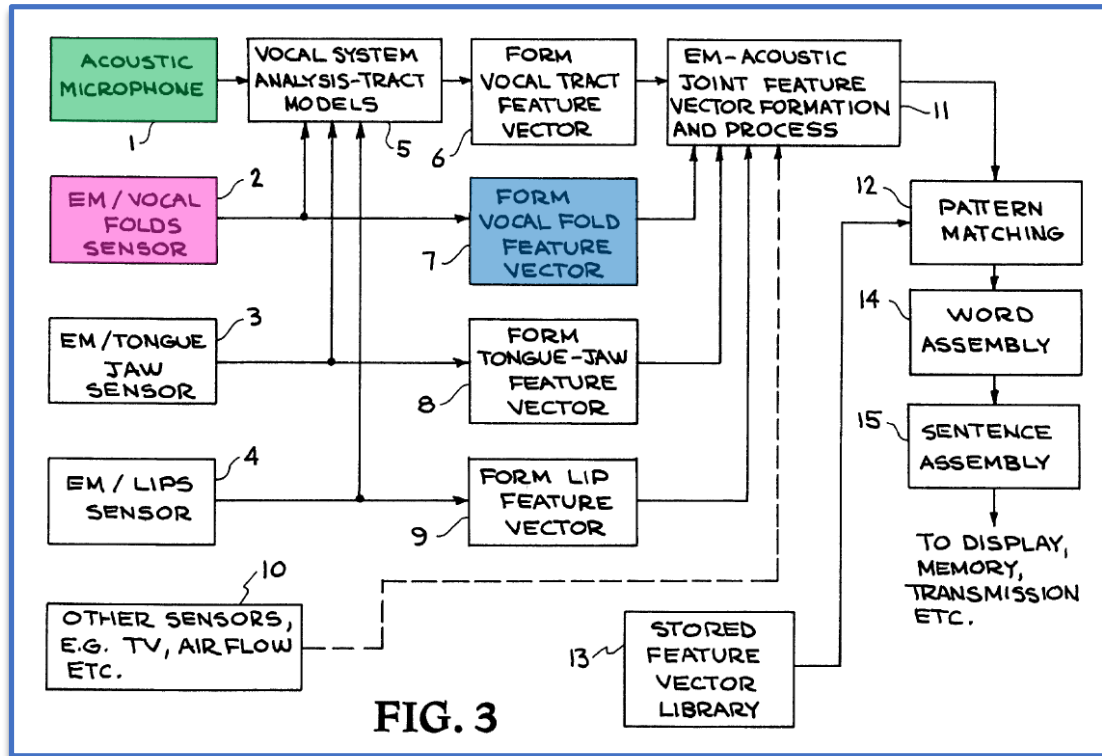
EX2001 (Anderson Tr.) at 31:3-13.

[1H]: Combination renders obvious “calculate” under PO’s construction

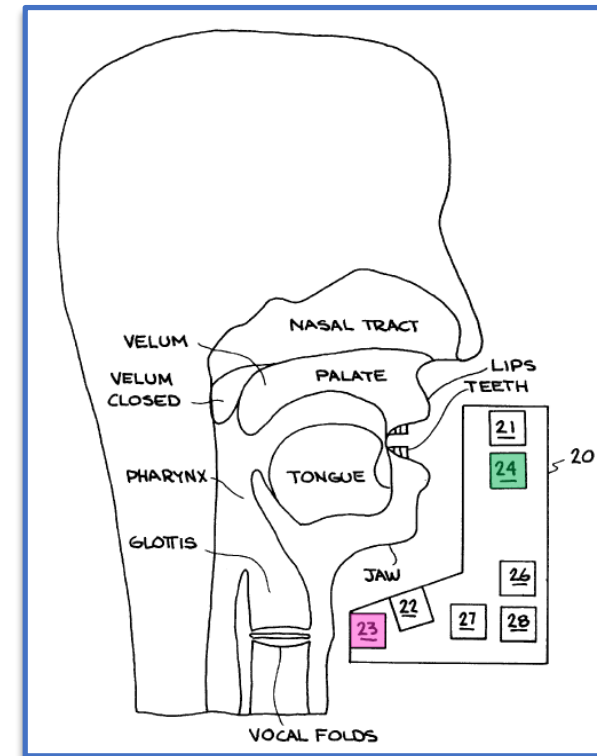
19 A. I chose to use Hussain. Hussain describes
20 sub-band processing system using two microphones as
21 an input that would fit beautifully with the
22 disclosed system of Andrea.

EX2005 (Anderson Tr.) at 115:19-22.

[1I]: Holzrichter renders obvious “a sensor independent from the [mic] array”



EX1007 (Holzrichter) at Fig. 3.

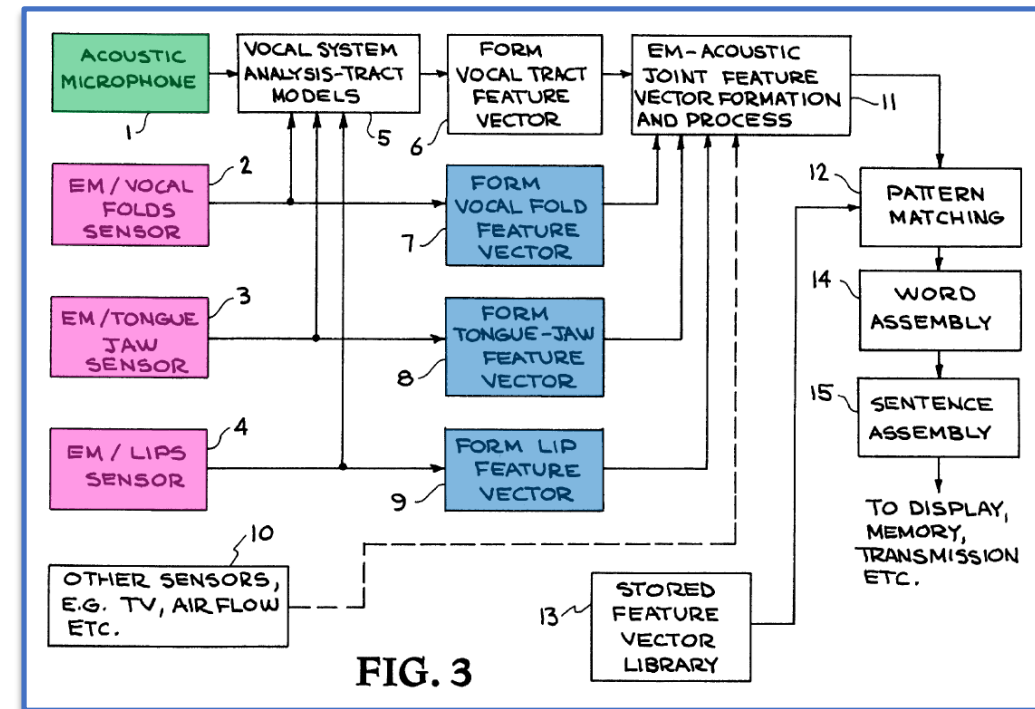


EX1007 (Holzrichter) at Fig. 4.

[1I]: Holzrichter renders obvious “a sensor independent from the [mic] array”

12 Q. Does the tongue move during production of
13 unvoiced speech?
14 A. Yes. The tongue can move to, you know -- you
15 know, to modulate that airflow, yes.
16 Q. Does the jaw move during production of
17 unvoiced speech?
18 A. You know, I think the jaw could also move,
19 yes.
20 Q. Do the lips move during production of unvoiced
21 speech?
22 A. Possibly. Yes.

EX1018 (Sayeed Tr.), 56:12-22.



EX1007 (Holzrichter) at Fig. 3.