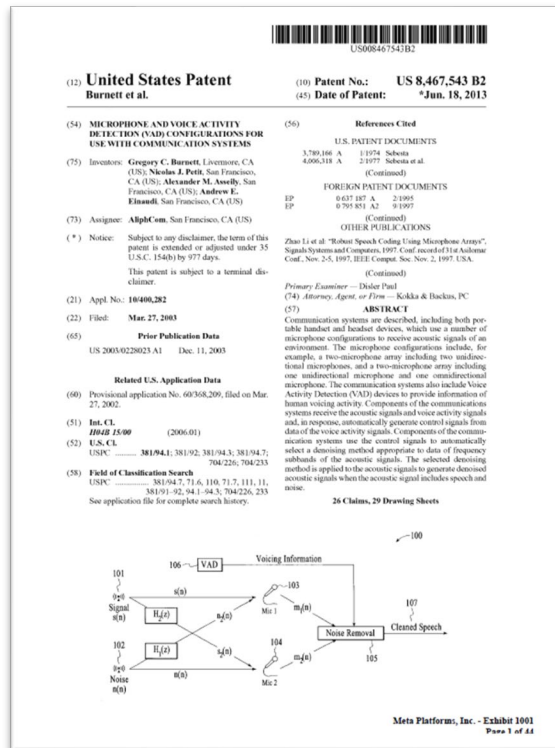


Before The Patent Trial And Appeal Board
Oral Hearing
December 4, 2024

Jawbone's Presentation

IPR2023-01291: U.S. 8,467,543

U.S. Patent 8,467,543



Ex. 1001

The '543 Patent:

- Filed: Mar. 27, 2003
- Issued: June 18, 2013
- 26 claims
- Independent claims 1 and 26

Claim 1

Claim 1

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.

Ex. 1001, cl. 1

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

Claim Construction

“voice detection subsystem,” “voice activity signals,” “human voicing activity”

Jawbone's Construction	Meta's Construction
“a subsystem that detects the presence of speech, including voiced and unvoiced speech”	No construction necessary
“signals that indicate speech activity, including voice and unvoiced speech activity”	Alternatively, “voice detection subsystem,” includes at least a “GEMS” (which solely detects voiced speech), an “accelerometer sensor in contact with skin,” or a “skin-surface microphone sensor,”
“human speech activity, including voiced speech and unvoiced speech”	

Claim Construction

“voice detection subsystem,” “voice activity signals,” “human voicing activity”

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.

Ex. 1001, 4:33-43

Jawbone’s Construction

“a subsystem that detects the presence of speech, including voiced and unvoiced speech”

“signals that indicate speech activity, including voice and unvoiced speech activity”

“human speech activity, including voiced speech and unvoiced speech”

POR, 14-16
Sur-Reply 7-10

Claim Construction

“voice detection subsystem,” “voice activity signals,” “human voicing activity”

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.

Jawbone's Construction

“a subsystem that detects the presence of speech, including voiced and unvoiced speech”

“signals that indicate speech activity, including voice and unvoiced speech activity”

“human speech activity, including voiced speech and unvoiced speech”

Ex. 1001, cl. 2-4

Sur-Reply 4-6

Claim Construction

“voice detection subsystem,” “voice activity signals,” “human voicing activity”

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.

Ex. 1001, 11:51-60

Jawbone's Construction

“a subsystem that detects the presence of speech, including voiced and unvoiced speech”

“signals that indicate speech activity, including voice and unvoiced speech activity”

“human speech activity, including voiced speech and unvoiced speech”

Claim 1

Claim 1

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.

wherein the denoising method includes
generating a noise waveform estimate
associated with noise of the acoustic signals,

Ex. 1001, cl. 1

Claim Construction

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

Jawbone's Construction	Meta's Construction
“calculating a noise estimate based on the noise associated with all the acoustic signals”	No construction necessary

POR, 16

Claim Construction

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

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

Jawbone's Construction

“calculating a noise estimate based on the noise associated with *all the acoustic signals*”

Ex. 1001, Claim 1

POR, 16-18
Sur-Reply 7-10

Claim Construction

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

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,

Ex. 1001, Claim 1

Jawbone's Construction

“*calculating* a noise estimate based on the noise associated with all the acoustic signals”

POR, 16-18
Sur-Reply 7-10

Claim Construction

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

Now, $H_1(z)$ can be calculated using any of the available system identification algorithms and the microphone outputs when only noise is being received. The calculation should be done adaptively in order to allow the system to track any changes in the noise.

Ex. 1001, 15-19

This calculation for $H_2(z)$ appears to be just the inverse of the $H_1(z)$ calculation, but remember that different inputs are being used as the calculation now takes place when speech is being produced. Note that $H_2(z)$ should be relatively constant,

Ex. 1001, 15:37-40

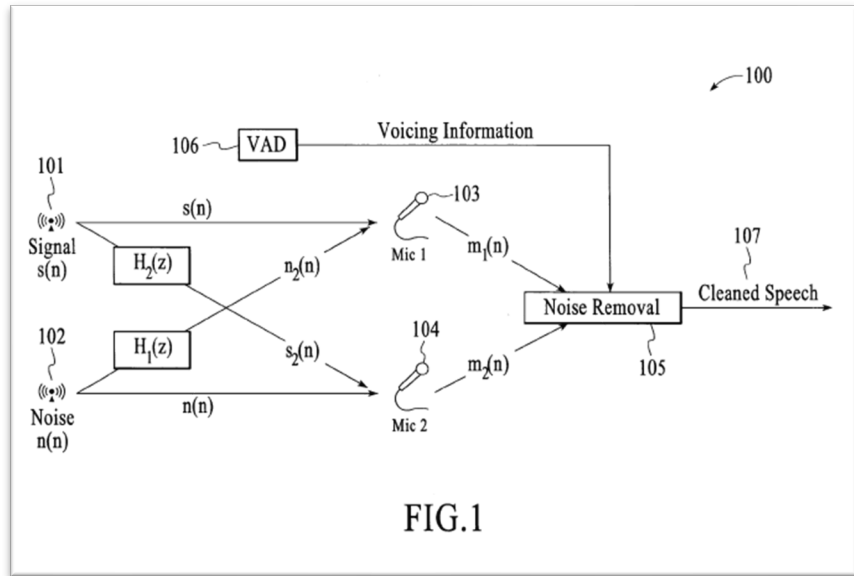
Jawbone's Construction

“*calculating* a noise estimate based on the noise associated with all the acoustic signals”

POR, 16-18
Sur-Reply 7-10

Claim Construction

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



Ex. 1001, Fig. 1

$$S(z) = M_1(z) - N(z)H_1(z)$$

$$N(z) = M_2(z) - S(z)H_2(z)$$

$$S(z) = M_1(z) - [M_2(z) - S(z)H_2(z)]H_1(z)$$

$$S(z)[1 - H_2(z)H_1(z)] = M_1(z) - M_2(z)H_1(z)$$

Ex. 1001, 15:50-55

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

Ex. 1001, 15:64-16:4

Noise Waveform
Estimate

Claim Construction

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

$$S(z) \approx M_1(z) - M_2(z)H_1(z), \text{ Ex. 1001, 15:64-16:4}$$

Noise Waveform
Estimate

$$H_1(z) = \frac{M_{1n}(z)}{M_{2n}(z)}$$

Ex. 1001, 15:10-15

Jawbone's Construction

“*calculating* a noise estimate based on the noise associated with all the acoustic signals”

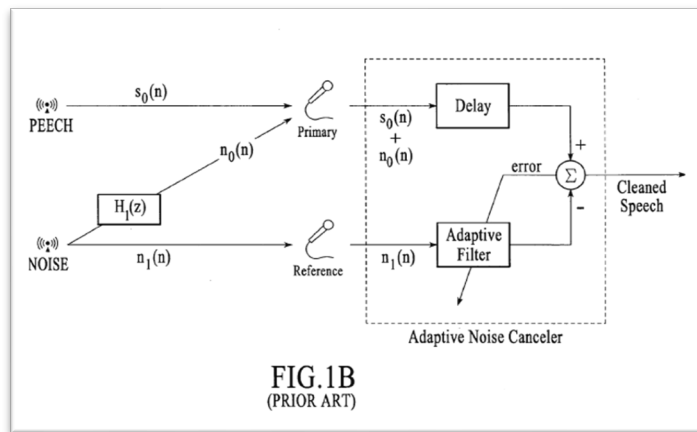
$$S(z) \approx M_1(z) - M_2(z) \frac{M_{1n}(z)}{M_{2n}(z)}$$

POR, 20

POR, 19-21
Sur-Reply 7-10

Claim Construction

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



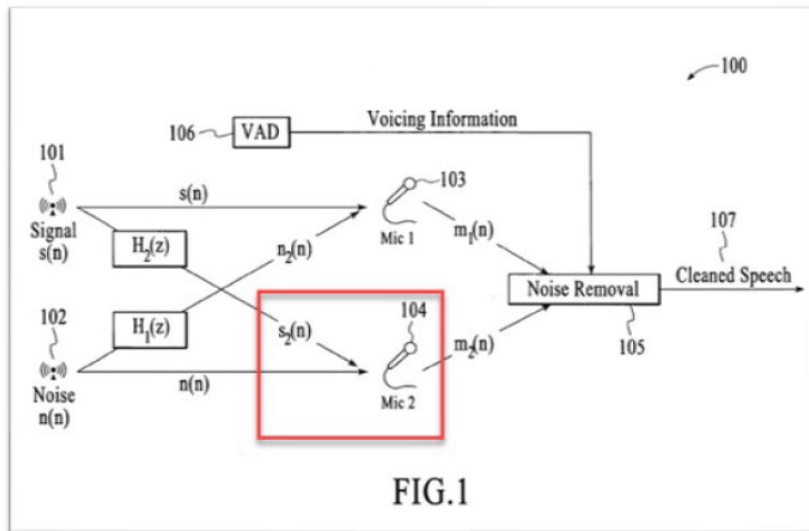
Ex. 1001, Fig. 1B

At first glance, it may seem as if the Pathfinder algorithm is very similar to other algorithms such as classical ANC (adaptive noise cancellation), shown in FIG. 1B. However, close examination reveals several areas that make all the difference in terms of noise suppression performance, including using VAD information to control adaptation of the noise suppression system to the received signals, using numerous subbands to ensure adequate convergence across the spectrum of interest, and supporting operation with acoustic signal of interest in the reference microphone of the system, as described in turn below.

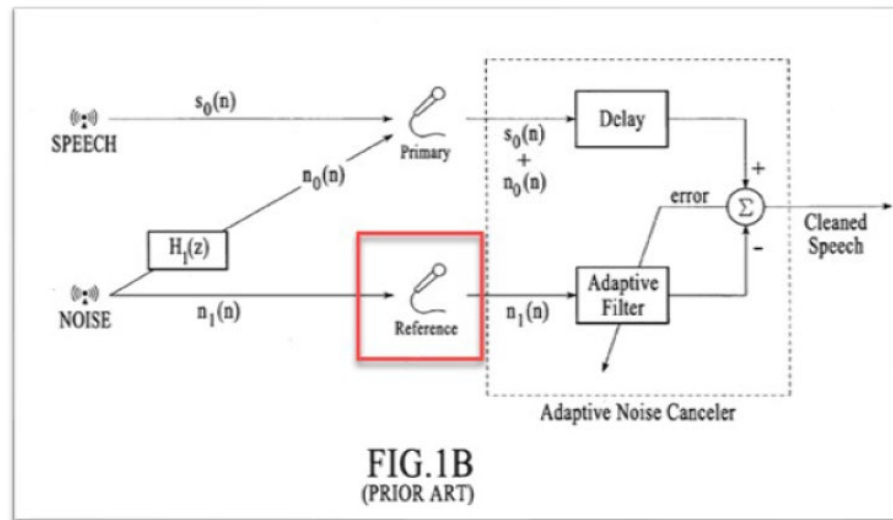
Ex. 1001, 16:44-54

Claim Construction

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



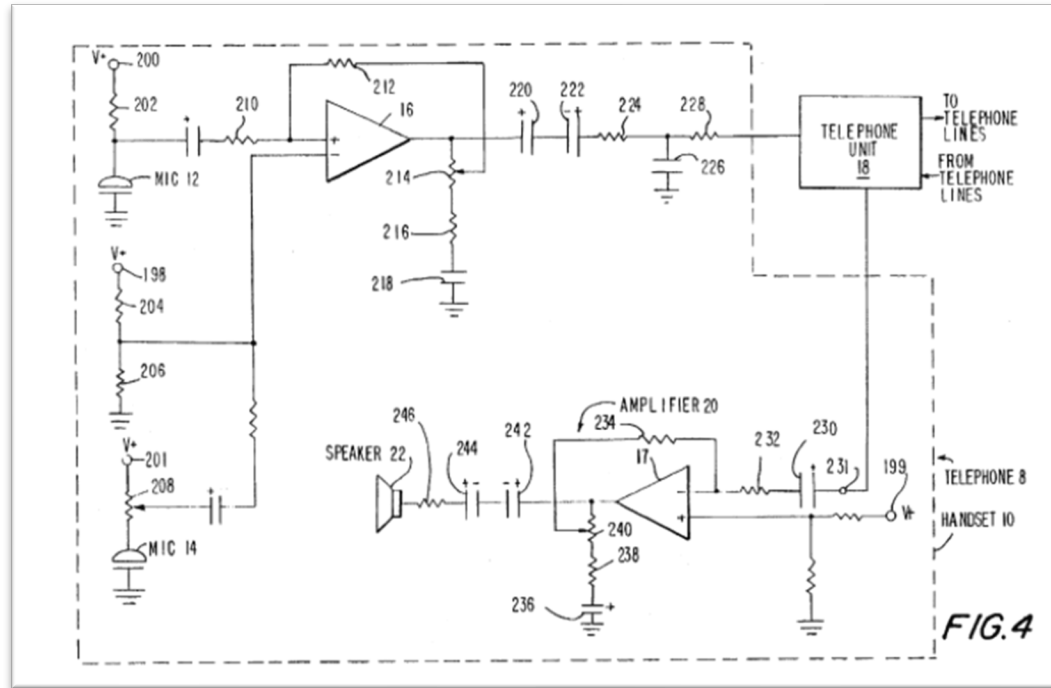
Ex. 1001, Fig. 1



Ex. 1001, Fig. 1B

GROUND 1: Meta's Combination

Andrea + Hussain



Ex. 1005, Fig. 4

GROUND 1: Meta's Combination

Andrea + Hussain

Andrea discloses or suggests [1H]. Ex. 1003, ¶ 131.

Pet., 40

Ex. 1003, ¶ 133. The “[a]coustic signals composed of speech or the like and background noise are supplied to the first microphone 12” are similarly converted into an electrical signal waveform and “supplied to the plus terminal of the op-amp 16.” Ex. 1005, 12:55-59. The op-amp 16 then “subtract[s] the noise signal from the second microphone 14 from the speech and noise signal from the first microphone 12” to provide an “electrical signal representing substantially the speech to the telephone unit 18” without the noise. *Id.*, 12:55-66, 16:20-23 (“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”).

Pet., 41

THE WITNESS: Andrea describes microphones and their configuration and in a system that's very effective. The acoustics are hard to design. Andrea's licked that.

What I'm scrapping is the differentiating amplifier. What I'm replacing it with is improved processing. So I'm scrapping the wide band differencing operation with a sub-band operation that can do both estimation and differencing and other methods.

Ex. 2005, 58:9-18

GROUND 1: Meta's Combination

Holzrichter

The algorithm detects the onset of speech by using both an acoustic and one or more EM sensor modules at the same time. The issue for this algorithm is that some words begin with unvoiced sounds such as “s” in “sam” or in “sixteen”; thus the algorithm must be able to back up to catch sounds that may be missed by the vocal fold motion onset detection by the EM sensor. In two speech processing systems, shown in FIGS. 12 and 13, the onset algorithm described here would be used primarily in box 66. The unvoiced duration timer T is initialized to zero at system turn-on.

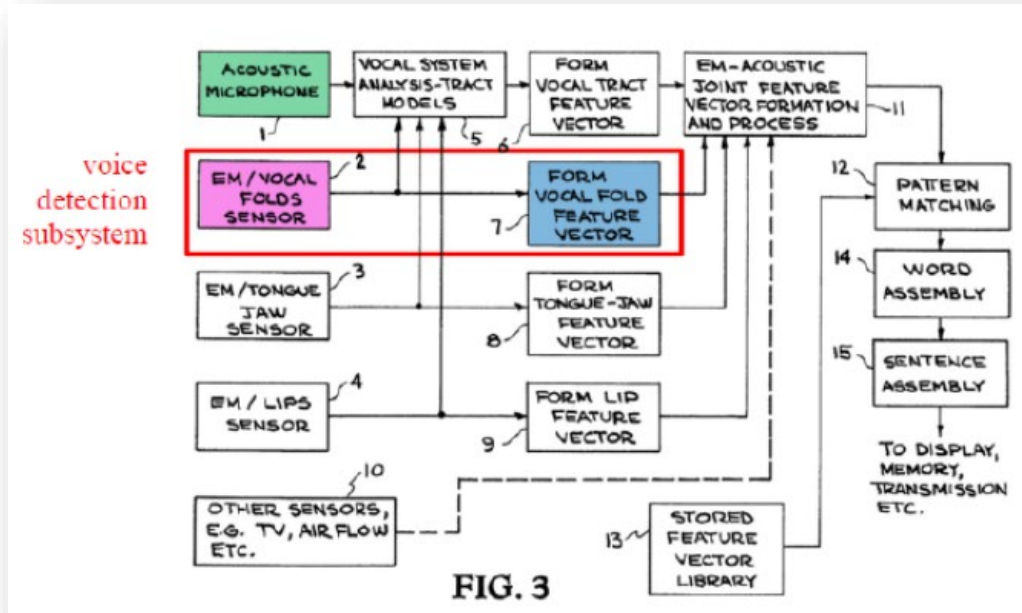
Ex. 1007, 31:50-60

within the same speech time period in box 79. If acoustics and vocal fold motion are present, then a voiced speech unit has occurred and the unvoiced timer, T, is set to zero in box 78, and processing on frame t_i continues. If satisfactory acoustics, but no vocal fold motion occurs in box 72, the sound in frame t_i is labeled as nonvoiced speech in box 79 and processing continues. If the acoustic signal is too loud

Ex. 1007, 32:4-6

GROUND 1: Meta's Combination

Holzrichter



Pet., 43

Holzrichter



Ex. 1006, 2614