

(43) International Publication Date
30 January 2014 (30.01.2014)

(51) International Patent Classification:

H04R 1/46 (2006.01) **G10L 21/02** (2013.01)
H04R 1/08 (2006.01) **G10L 25/93** (2013.01)
H04R 3/00 (2006.01) **G10L 25/84** (2013.01)
H04R 5/027 (2006.01) **H04M 1/60** (2006.01)
H04M 1/05 (2006.01) **G02C 11/00** (2006.01)

(21) International Application Number:

PCT/FI2013/050766

(22) International Filing Date:

16 July 2013 (16.07.2013)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

13/557,895 25 July 2012 (25.07.2012) US

(71) Applicant: **NOKIA CORPORATION** [FI/FI]; Keilalahdentie 4, FI-02150 Espoo (FI).

(72) Inventors: **XIE, Peng**; 3475 Granada Avenue #393, Santa Clara, California 95051 (US). **WHITE, Sean**; 903 Plumtree Lane, Mountain View, California 94040 (US). **KAEWITIP, Kantapon**; 623 Kelton Street, Los Angeles, California 90024 (US). **WILLIAMS, Lance**; 4321 Talofa Avenue, Toluca Lake, California 91602-2917 (US).

(74) Agents: **NOKIA CORPORATION** et al.; IPR Department, Jussi Jaatinen, Keilalahdentie 4, FI-02150 Espoo (FI).

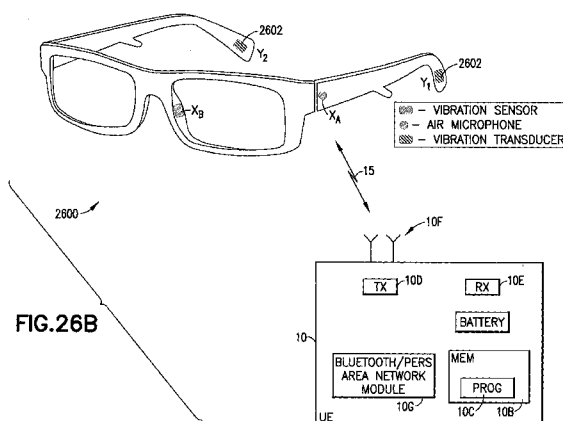
(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— with international search report (Art. 21(3))

(54) Title: HEAD-MOUNTED SOUND CAPTURE DEVICE



(57) Abstract: A platform (2600) that is configured to be removably placed symmetrically on or about a user's head has at least a first transducer (X_B , 2602) configured to capture vibration of the user's skull or facial movement generated by the user's voice activity and to detect the user's speaking activity. This first transducer (X_B , 2602) converts the vibration or facial movement into a first electrical audio signal. The electrical audio signal from the first transducer (X_B , 2602) is processed by circuitry or embodied software as voiced frames and/or as unvoiced frames, in which the voiced frames and/or the unvoiced frames are defined based at least on the first electrical audio signal. Multiple embodiments follow from this: where the first transducer (X_B , 2602) is a vibration sensor; where voice is captured by an air microphone (X_A) and filtering adaptation differs for the voiced versus unvoiced frames as defined by the first transducer (X_B , 2602), and another with at least three air microphones (X_A).

HEAD-MOUNTED SOUND CAPTURE DEVICE

TECHNICAL FIELD:

The exemplary and non-limiting embodiments of this invention relate generally to
5 arrays of sensors (and in some embodiments also one or more sensors) for voice
capture which are mounted to spectacles, helmets, headbands or other devices
mounted symmetrically about a user's head, and the specific examples concern the
arrangement of such arrays as well as noise reduction, multiple sensors, bone
conduction, adaptive filtering, and binaural and directional recording of ambient
10 sound captured at such a device.

BACKGROUND:

Hands-free mobile communication is often facilitated with wireless headsets having a
15 Bluetooth link with the mobile handset, which allows the user to conveniently and
safely communicate while engaged in another activity such as driving. But the use of
headsets raises new issues of styling, comfort, and most of all, ambient noise, which
often causes voice quality to be sacrificed.

20 Speech signals corrupted by additive ambient noise has long been a major challenge
for voice communication and speech user interfaces for mobile terminals and speech
recorders, particularly in highly noisy environments. End retail users often prefer that
the microphone is not in close proximity to their mouths, which makes the challenge
of capturing quality vocal signals even more pronounced. Such a headset's
25 microphone is prone to pick up more ambient sounds than one disposed closer to the
speaker's mouth, resulting in lower signal-to-noise ratio (SNR) and lower speech
intelligibility.

Users have also placed a high value on comfort and appearance of these headsets.
30 Typically such headsets are mounted in or around the ear, which causes discomfort
from prolonged use. These sometimes also interfere with eyeglasses or sunglasses
which the user might be wearing. When a headset speaker is implemented as an
earbud or similar in-the-ear arrangement, they can sometimes block out ambient

sound to an extent that is unacceptable. At least one recent survey found a large minority of those surveyed considered conventional Bluetooth headsets unattractive.

Research in improving the above concerns has been ongoing. For example, noise-reduction techniques to suppress ambient noise in ear-mounted Bluetooth headsets are detailed for example at US Patent Nos. 7,983,907 entitled HEADSET FOR SEPARATION OF SPEECH SIGNALS IN A NOISY ENVIRONMENT; and 8,098,844 entitled DUAL-MICROPHONE SPATIAL NOISE SUPPRESSION; as well as US Patent Application Publications 2006/0120537 entitled NOISE SUPPRESSING MULTI-MICROPHONE HEADSET; 2007/0088544 entitled CALIBRATION BASED BEAMFORMING, NON-LINEAR ADAPTIVE FILTERING, AND MULTI-SENSOR HEADSET; and 2011/0010172 entitled NOISE REDUCTION SYSTEM USING A SENSOR BASED SPEECH DETECTOR.

There are also commercially available Bluetooth headsets implemented as spectacles/eyeglasses. Existing techniques can be seen at US Patent Nos. 7,289,767 entitled WIRELESS COMMUNICATION GLASSES; 7,792,552 entitled EYEGLASSES FOR WIRELESS COMMUNICATIONS; and 7,684,571 entitled SYSTEM AND METHOD OF GENERATING AN AUDIO SIGNAL. Further example includes Temco Vibraudio model GL41AB which is a bone-conduction Bluetooth arrangement on sunglasses and currently available in Japan (online product literature at <http://www.temco-j.co.jp/web/webAction.do?method=productDetail&productId=fcaaf8b4227210b501227952bd4a0032>).

In a somewhat related vein are hearing aids implemented in eyeglasses, some of which also incorporate a microphone or multiple microphones to pick up ambient sound for amplification via the hearing aid portion. See for example US Patent Nos. 6,629,076 entitled METHOD AND DEVICE FOR AIDING SPEECH; 7,031,483 entitled HEARING AID COMPRISING AN ARRAY OF MICROPHONES; 7,609,842 entitled SPECTACLE HEARING AID; and 8,139,801 entitled HEARING AID GLASSES USING ONE OMNI MICROPHONE PER TEMPLE.

Further references which might be relevant to the teachings below include US Patent Nos. 4,520,238 entitled PICKUP DEVICE FOR PICKING UP VIBRATION TRANSMITTED

THROUGH BONES; 7,555,136 entitled NASAL BONE CONDUCTION WIRELESS COMMUNICATION TRANSMITTING DEVICE; 7,447,630 entitled METHOD AND APPARATUS FOR MULTI-SENSORY SPEECH ENHANCEMENT; 7,346,504 entitled MULTI-SENSORY SPEECH ENHANCEMENT USING A CLEAN SPEECH PRIOR; 7,406,303 entitled MULTI-SENSORY SPEECH ENHANCEMENT USING SYNTHESIZED SENSOR SIGNAL; 7,499,686 entitled METHOD AND APPARATUS FOR MULTI-SENSORY SPEECH ENHANCEMENT ON A MOBILE DEVICE; and 7,680,656 entitled MULTI-SENSORY SPEECH ENHANCEMENT USING A SPEECH-STATE MODEL. Additionally, reference is made to the following research papers: Z. Liu, Z. Zhang, A. Acero, J. Droppo, and X. D. Huang, DIRECT FILTERING FOR AIR- AND BONE-CONDUCTIVE MICROPHONES, [*IEEE International Workshop on Multimedia Signal Processing (MMSP)*, Siena, Italy, 2004]; and Z. Liu, A. Subramanya, Z. Zhang, J. Droppo, and A. Acero, LEAKAGE MODEL AND TEETH CLACK REMOVAL FOR AIR- AND BONE-CONDUCTIVE INTEGRATED MICROPHONES, [*Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, Philadelphia, PA, USA, 2005.]

SUMMARY:

According to a first exemplary aspect the invention there is an apparatus comprising:
a platform, configured to be removably placed symmetrically on or about a user's head; at least a first transducer configured to capture vibration of the user's skull or facial movement generated by the user's voice activity and detect the user's speaking activity, in which the at least first transducer converts the vibration or facial movement into a first electrical audio signal; and one or more circuitries and/or one or more memories including computer program code for processing the at least the electrical audio signal from the at least first transducer received at the platform as voiced frames and/or as unvoiced frames, in which the voiced frames and/or the unvoiced frames are defined based at least on the first electrical audio signal.

According to a second exemplary aspect the invention there is a method comprising:
determining from at least a first electrical audio signal from a first transducer voiced frames during which a user is speaking and unvoiced frames during which the user is not speaking, in which the first transducer is disposed on a platform configured to be

removably placed symmetrically on or about a user's head and is configured to capture vibration of the user's skull or facial movement generated by the user's voice activity and detect the user's speaking activity; and processing the first electrical audio signal received at the platform as the voiced frames and/or as the unvoiced frames.

According to a third exemplary aspect the invention there is a computer readable memory storing a program of computer readable instructions which when executed by at least one processor result in actions comprising: determining from at least a first electrical audio signal from a first transducer voiced frames during which a user is speaking and unvoiced frames during which the user is not speaking, in which the first transducer is disposed on a platform configured to be removably placed symmetrically on or about a user's head and is configured to capture vibration of the user's skull or facial movement generated by the user's voice activity and detect the user's speaking activity; and processing the first electrical audio signal received at the platform as the voiced frames and/or as the unvoiced frames.

According to a fourth exemplary aspect the invention there is an apparatus comprising a platform, configured to be removably placed symmetrically on or about a user's head; at least two transducers disposed about opposed sides of the platform; and circuitry or embodied software for resolving a spatial signal from signals output from the at least two transducers.

According to a fifth exemplary aspect the invention there is a method comprising: receiving signals from at least two transducers disposed about opposed sides of a platform, configured to be removably placed symmetrically on or about a user's head; and resolving a spatial signal from the received signals.

According to a sixth exemplary aspect the invention there is a memory storing a program of computer readable instructions which when executed by at least one processor result in actions comprising: in response to receiving signals from at least two transducers disposed about opposed sides of a platform, configured to be removably placed symmetrically on or about a user's head, resolving a spatial signal from the received signals.

These and other aspects are detailed further below.

5 BRIEF DESCRIPTION OF THE DRAWINGS:

Figures 1A-B illustrate two implementations of a first embodiments/single channel configuration of an eyeglass mounted voice capture arrangement according to any exemplary embodiment of these teachings.

- 10 Figure 2 illustrates a prior art arrangement of an air microphone disposed on eyeglasses.

Figures 3A-B and 4A-B show experimental results from the air microphone only of Figure 2 and the vibration sensor only of Figure 1A in the time and the frequency
15 domains respectively in an environment with minimal ambient noise.

Figures 5A-B and 6A-B are similar to Figures 3A-B and 4A-B respectively but for a noisy environment.

- 20 Figures 7 and 8 illustrate block diagrams for processing the single channel audio input according to a first solution of the first embodiment of these teachings shown by example at Figures 1A-B, according to exemplary embodiments of these teachings.

Figures 9A-B and 10A-B show experimental results from the first solution (without
25 applying the optional module in Figure 8) of the single channel configuration (Figures 1A-B) in the time and the frequency domain, respectively.

- Figure 11 illustrates a block diagram for processing the single channel audio input according to a second solution of the first embodiment of these teachings shown by
30 example at Figures 1A-B, according to exemplary embodiments of these teachings.

Figures 12A-B and 13A-B show experimental results from the second solution of the single channel configuration (Figures 1A-B) in the time and the frequency domain,

respectively.

Figures 14A-B extend Figures 1A-B for a second embodiment/dual channel configuration with one vibration sensor and one air microphone capturing a user's voice according to another aspect of these teachings.

Figures 15A-B are functional modules used in further Figures which demonstrate the first and second solutions for the second embodiment/dual-channel configuration of Figure 14A-B according to these teachings.

Figures 16A-B illustrate block diagrams according to a first solution for the second embodiment/dual-channel configuration (Figure 14A-B) according to an exemplary embodiment of these teachings.

Figures 17A-B and 18A-B show experimental results from the first solution (Figure 16A-B) of the second embodiment/dual-channel configuration (Figure 14A-B) in the time and the frequency domain, respectively.

Figures 19A-B illustrate block diagrams according to a second solution for the second embodiment/dual-channel configuration (Figure 14A-B) according to an exemplary embodiment of these teachings.

Figures 20A-B and 21A-B show experimental results from the second solution (Figure 19A-B) of the second embodiment/dual-channel configuration (Figure 14A-B) in the time and the frequency domain, respectively.

Figure 22A is a perspective view of spectacles with three microphones and a nose-mounted vibration sensor with the microphones capturing a user's voice in a third embodiment/multi-channel configuration according to an exemplary embodiment of these teachings.

Figure 22B is similar to Figure 22A but with the vibration sensor mounted on the spectacle arm above the user's ear.

Figure 23 is a schematic block diagram for processing the signals received at the spectacles of Figures 22A-C according to an exemplary embodiment of these teachings.

5

Figures 24A-B and 25A-B show experimental results from the diagram of Figure 23 in the time and the frequency domains, respectively.

Figure 26A-B is similar to the embodiment of Figure 14A but with additionally a vibration transducer mounted to contact a user's skull behind each ear for providing sound to the user for a fully hands free and ear free two way communication device, and further illustrating a simplified block diagram of a mobile terminal to which the head-set platform may be wirelessly linked through Bluetooth or some other personal areas network.

15

DETAILED DESCRIPTION:

Below are detailed three distinct embodiments for utilizing a platform configured to removably rest symmetrically on a user's head such as eyeglasses (including sunglasses), helmets, hats, headbands and the like, for primary use with capturing voice signals from a user wearing the platform. While these are all detailed particularly as being disposed in eyeglasses, that is only one of the possible platforms in which they may be implemented and which is used to explain further particulars of the circuitry and the signal processing. In the first embodiment there is a bone vibration sensor, sometimes termed a 'bone microphone', which is in one embodiment a mechanical vibration to electric transducer and senses vibrations in the skull of the speaker. In the first embodiment these sensed mechanical vibrations are used for the voice signal itself. This is alternatively referred to as a single channel configuration. This single channel configuration differs from other prior art in-the-ear type devices which may pick up vibrations from the user's cheekbone or ear canal, which prior art devices are of course not symmetric about a user's head when worn.

25
30

In the second embodiment there is a vibration sensor and also an air microphone

disposed in the eyeglasses. The signals from both the vibration sensor and the air microphone are taken as the voice signals. In the third embodiment there are multiple air microphones arranged in a spatial array, and also a vibration sensor which is used differently as compared to the second embodiment, and only to detect the user's voice activities and distinguish those voiced frames from unvoiced frames where the user is not speaking. For the third embodiment, below are also detailed alternatives for the vibration sensor and so it is more generally referred to as a voice activity sensor. Such alternatives include a downward facing camera, an ultrasonic sensor and an infrared sensor, to name a few non-limiting alternatives to the vibration sensor. Multiple laboratory analyses are presented below to show just how effective is the vibration sensor for the various detailed embodiments.

Certain of the exemplary and non-limiting embodiments detailed below present an unobtrusive head-mounted device which laboratory testing has proven to capture clear voice from the wearer for voice communication and speech user interface (UI) in a hands-free, mouth-free, and ear-free manner. The examples below and the related description utilize conventional eyeglasses/spectacles as the structure of the removable head wearable platform but this is simply for clarity of the examples and is not limiting; other types of platforms are noted above.

Figure 1A illustrates one embodiment of eyeglasses 100 according to the first embodiment of these teachings. In this first embodiment there is only the vibration sensor mounted on the spectacle frame 100, and the signal from this vibration sensor is wirelessly sent to the user's mobile handset, such as via a Bluetooth link. The signal $x_B(t)$ may be sent directly, or the signal processing described herein may be done in the eyeglass platform itself and the 'system output' signal sent wirelessly to the user's mobile handset for framing, amplification, and transmission over a cellular or other link to the other party of some ongoing voice call.

One implementation of the first embodiment shown at Figure 1A disposes the vibration sensor x_B in the nose pad 106. Another implementation of the first embodiment shown at Figure 1B disposes the vibration sensor x_B near the distal end and on the inboard side of one of the spectacle arms 104A-B. In the Figure 1B

implementation the vibration sensor x_B would lay just behind the ear so as to be in physical contact with the user's skull at the mastoid. Alternatively the vibration sensor x_B may be disposed along one of the arms 104A-B nearer to the front frame 102 so as to contact the user's skull in the temple area. A headband or hat or helmet
5 are effective implementations of the platform with the vibration sensor disposed along the user's temple or forehead.

The experimental data collected and presented below utilized resampling to 8 kHz and also high-pass filtered to remove the very low frequencies (below 200Hz) so as to fit
10 in the standard telephony band. Quantitative experimental results from the Figure 1A-B single-channel configuration (no air microphone) operating on the auxiliary channel $x_B(t)$ from the vibration sensor x_B are shown at Figures 3B and 5B which are in the time domain, and also at Figures 4B and 6B which shows results of the same underlying data in the frequency domain. For comparison, Figures 3A, 4A, 5A and
15 6A an input/forward microphone signal from a conventional air microphone configuration such as that of Figure 2.

Evident in Figures 3A through 6B is a digital on/off signal line (see also Figure 9A) which shows a square wave superimposed over the actual data of the microphone
20 signal, and is relevant for other embodiments which filter voiced frames differently from unvoiced frames.

In the test arrangement for gathering the data of Figures 3A-B through 6A-B, as well as for all of the experimental data presented herein, there were eight loudspeakers
25 disposed about an office environment to simulate practical ambient noise environments, including babble, subway, street, etc. Data was collected from different individuals wearing the same testing spectacles to normalize for different voice characteristics, each of whom spoke as if for a normal phone conversation.

30 Figures 3A-B are in the time domain and represent the signal captured in same recordings by the air microphone x_A alone (representing the conventional single air microphone configuration of prior art Figure 2) and by the bone microphone x_B alone, respectively, from the Figure 1A-B arrangement. Figures 4A-B are similar results in

the frequency domain. All four of these represent sound captured in a quiet environment with little ambient noise to serve as a baseline. Figures 5A-B are similar to Figures 3A-B, and Figures 6A-B are similar to Figures 4A-B, but for a noisy environment.

5

From these four figures the following characteristics of the two different types of microphones becomes clear.

10

Table 1: Comparison between air microphone and vibration sensor for Figure 9A.

	Air Microphone	Vibration Sensor
Advantages	• Flat frequency response • Wide spectrum	• Insensitive to air-conducted sound (The leakage of ambient sound is captured through the spectacles)
Disadvantages	• Capture “everything” in the air • Low sound quality and intelligibility in noisy environments	• Rough frequency response (can be solved by equalization) • Limited spectrum (but could still meet telephony requirements) • Requires larger amplification

In the quiet environment both types of sensors x_A and x_B are able to capture clear voice, with very high signal-to-noise ratio (SNR). The air microphone x_A outperforms the vibration sensor x_B with a flatter frequency and a wider spectrum. But in the highly noisy environment, the user’s voice from the air microphone x_A is completely buried in the ambient noise. Here it can be clearly noticed that the signal captured by the vibration sensor x_B has a higher SNR than that captured by the air microphone x_A . In this recording, the SNR of the air microphone x_A input is 4.5 dB, and the SNR of the vibration sensor x_B input is 12.7 dB. Thus the vibration sensor x_B alone provides an 8.2 dB improvement in SNR without any further processing. Although the vibration sensor x_B itself is almost immune to air-conducted sound, it still captures a certain amount of ambient noise through the spectacles due to the frame (and more significantly the lenses) vibrating when sound waves impinge upon them. This SNR improvement can be further improved by the addition of a vibration dampening

mechanism between the vibration sensor x_B and spectacle frame; for example foam or flexible adhesive to attach the vibration sensor x_B to the frame.

The vibration sensor x_B has a much higher SNR at lower frequencies, especially
5 below 1500 Hz. The circuit of Figure 7 exploits this characteristic by applying a low-pass filter (LPF) 702 with the cutoff frequency at 1500 Hz, from which can be obtained a relatively clear voice signal. The LPF 702 can be imposed in conjunction with the optional module 800 shown at Figure 8. The low-pass filtering may be performed in the time domain with a simple finite impulse response (FIR) filter. But
10 for ready comparison with other experimentation results the LPF 702 was implemented in the frequency domain by setting the value of corresponding frequency bins to zero. In this case the system output using Figure 7 had an SNR at 24.0 dB, which means an improvement of 19.5 dB over a conventional air microphone input, but at some sacrifice of sound quality. The voice is of course band limited to below
15 1500 Hz, but its intelligibility was still acceptable.

Improved performance can be obtained by employing an optional module 800 shown more particularly at Figure 8. The low pass filtering 702 was again performed in the frequency domain to facilitate further processing. The additional processing in Figure
20 8 over the bare LPF 702 of Figure 7 includes an artificial bandwidth extension (ABE) block 806 which takes as its input the output signal of a voice activity detector 804. The voice activity detector 804 distinguishes the voiced frames from the unvoiced frames, which the vibration sensor x_B distinguishes easily since it is not susceptible to ambient noise (though it is noted below that there might be some ambient vibration
25 noise from vibrations of the eyeglass lenses or frame). The concept of artificial bandwidth extension is the focus of ongoing research by Nokia, Oy (the assignee of the rights herein), as well as other companies. For completeness Figure 8 also illustrates a framing/discrete Fourier transform block 802 to put the input time-domain signal into the frequency domain, and an overlap/add block 808 following the
30 bandwidth expansion block 306 to turn the signal back into the time domain. If the LPF 702 is implemented in the time domain the output of the LPF 702 will feed the discrete Fourier transform (DFT) 802, else the LPF 702 can be implemented directly in the frequency domain by setting the value of corresponding frequency bins to zero

after the DFT 802.

Quantitative experimental results from the Figure 1A-B single-channel configuration (no air microphone) operating on the signal $x_B(t)$ from the vibration sensor x_B are shown at Figure 9B which is in the time domain, and 10B which shows results of the same underlying data in the frequency domain. For comparison, Figures 9A and 10A show an input/forward microphone signal from a conventional air microphone configuration such as that of Figure 2, whereas Figures 9B and 10B show the system output $z(t)$ from the vibration microphone x_B of Figures 1A-B.

Evident in Figures 9A-B is a digital on/off signal line 902 labeled at Figure 9A, which like that in Figures 3A through 6B shows a square wave 902 superimposed over the actual data of the microphone signal. Vertical transitions of that square wave distinguish the voiced frames 904 from the unvoiced frames 906 which are considered to be noise only.

Figure 11 illustrates a second circuit for the first embodiment (single channel) of Figures 1A-B, as an alternative to the circuit of Figures 7-8. The signal $x_B(t)$ from the vibration sensor x_B is input to a framing/DFT block 1102, and the voice activity detector 1104 and overlap/add 1108 blocks are similar to those shown and described with respect to Figure 8. The voice activity detector block 1104 separates the voiced signal from the unvoiced (noise-only) frames. The voice activity detector 1104 can be implemented in either the frequency domain as shown at Figure 11 or directly in the time domain in an alternative circuit arrangement. In only the noise-only frames the noise profile (the estimated noise spectrum) is updated at update block 1106. In all frames, voiced and unvoiced/noise-only, spectral subtraction is performed at subtraction block 1107 to remove the noise spectrum output from the update block 1106. After the overlap-add process 1108 the data is transformed back into the time domain to form the output.

Experimental results from the Figure 1A-B single-channel configuration using the circuitry of Figure 11 operating on the signal $x_B(t)$ from the vibration sensor x_B are shown at Figures 12B (time domain) and 13B (frequency domain). For comparison

an input/forward air microphone signal is shown at Figures 12A (time domain) and 13A (frequency domain) for the conventional air microphone configuration of Figure 2, and the system output $z(t)$ is shown at Figures 12B and 13B. The square wave shows the voiced versus the unvoiced frames as noted earlier.

5

The circuit of Figure 11 provides a system output with a 20.5 dB SNR, which is a 16.0 dB improvement over a conventional air microphone input. Though this improvement is less than that provided by the circuit of Figure 7, it offers better sound quality because it preserves some of the high frequency components of the voice.

10 The spectral subtraction method used in this case is not too complex, but with further complex processing such as for example Attenuation Limited Wiener Estimation (ALWE) noise filtering, which is ongoing research by Nokia, Oy (the assignee of the rights herein), the signal to noise ratio SNR and sound quality could be increased even further.

15

In the first embodiment of Figures 1A-B there was a single vibration sensor. This is expanded in the second embodiment the eyeglasses incorporate an additional air conduction microphone x_A which operates with the bone vibration sensor x_B as dual channels from which the output voice signal is derived. This second embodiment also
20 can be implemented on spectacles, sports helmets, headbands, and other such head-mounted devices. Figures 14A-B illustrate two exemplary but non-limiting locations for these two sensors which drive the auxiliary channel for the second embodiment in which there is both an air conduction microphone x_A and a bone vibration sensor x_B in the eyeglasses. This second embodiment may be considered a dual channel
25 configuration.

Figures 14A-B illustrate exemplary placements for the bone microphone x_B (same as shown at Figures 1A-B) and with an additional air microphone x_A for this second embodiment. Alternatively, the vibration sensor x_B could be installed on the inboard
30 side of one arm 104A, 104B nearer the front face to contact the user's temple, but this placement is more effective with a helmet or headband type of platform instead of eyeglasses. For a helmet or headband type of platform, the user's forehead is another effective location.

Now consider a more quantitative treatment of the second embodiment as shown in Figure 14A where inputs from both the vibration sensor x_B and from the air microphone x_A are used to generate the output. As noted above, two solutions are presented for this second embodiment/dual-channel configuration, both of which are based on the following time domain system modeling:

$$x_A(t) = a(t) + n_A(t) \quad [1]$$

$$x_B(t) = b(t) + n_B(t) + n_e(t) \quad [2]$$

where $x_A(t)$ and $x_B(t)$ are the signal observed by the air microphone and the vibration sensor, respectively. The signal $a(t)$ represents the clean voice content captured by the air microphone, and $n_A(t)$ represents the ambient noise content. Correspondingly, $b(t)$ and $n_B(t)$ represent the clean voice and the ambient noise content picked up by the vibration sensor x_B , respectively. The electronic noise content in the air microphone x_A is ignored due to its high SNR level on self-noise, but the corresponding content in the vibration sensor x_B cannot be ignored and is represented by $n_e(t)$.

Turning the system into the frequency domain yields:

$$X_A(\omega) = A(\omega) + N_A(\omega) \quad [3]$$

$$X_B(\omega) = B(\omega) + N_B(\omega) + N_e(\omega) \quad [4]$$

where capital letters represent the Fourier transforms of the corresponding lowercase time-domain signals. Since $A(\omega)$ and $B(\omega)$ are both clean voice contents from the user, there exists an “equalizing transfer function” between them, which is $H(\omega) = \frac{A(\omega)}{B(\omega)}$.

Then equation [4] can be written as:

$$X_B(\omega) = H^{-1}(\omega)A(\omega) + N_B(\omega) + N_e(\omega) \quad [5]$$

25

If we define $X_B'(\omega) = X_B(\omega) - N_e(\omega)$, then equation [5] can be written as:

$$H(\omega)X_B'(\omega) = A(\omega) + H(\omega)N_B(\omega) \quad [6]$$

Defining $\beta(\omega) = \frac{N_B(\omega)}{N_A(\omega)}$, a transfer function on the ambient noise content, and $\alpha(\omega) = H(\omega)\beta(\omega)$, then equation [6] can be written as:

30

$$H(\omega)X_B'(\omega) = A(\omega) + \alpha(\omega)N_A(\omega) \quad [7]$$

Figures 15A-B are function modules 1500, 1550 which solve the above equations and

which are used in the first and second solutions for the second dual-channel configuration (Figure 14A). Both modules work in the frequency domain and take $X_A(n, \omega)$ and $X_B(n, \omega)$ as inputs. The frame number n is added into the input signals $X_A(\omega)$ and $X_B(\omega)$ because they are obtained by framing and DFT.

5

When a voice recording is done in a quiet environment, both $N_A(\omega)$ and $N_B(\omega)$ are close to zero and so equation [3] and [5] can be simplified as:

$$X_A(\omega) = A(\omega) \quad [8]$$

$$X_B(\omega) = H^{-1}(\omega)A(\omega) + N_e(\omega) \quad [9]$$

10

From equations [8] and [9] the transfer function $H(\omega)$ and the sensor electronic noise $N_e(\omega)$ can be directly computed, using the general block diagram shown at the module 1500 of Figure 15A. At each frequency ω , a frame selector 1502, 1504 determines and chooses reliable frames in $X_A(n, \omega)$ and $X_B(n, \omega)$ to minimize the effect of interference such as electronic noise, aliasing, etc. In one non-limiting example the set of reliable frames is derived as follows:

15

- Determine voiced frames at block 1502.
- Pick the highest 10% of $X_A(n, \omega)$ and $X_B(n, \omega)$ in amplitude among those.
- Select the frames that happen to be in both categories at block 1504. If the number of mutual frames are too low (for example, less than 30), then relax the percentile constraint to allow more frames.

20

Block 1506 then uses the above reliable frames to compute $H(\omega)$ by $H(\omega) = \frac{\sum X_A(n, \omega)}{\sum (X_B(n, \omega) - N_e(n, \omega))}$. And in $X_B(n, \omega)$, an estimator block 1508 determines unvoiced frames,

25

which yields an estimate of $N_e(\omega)$.

Figure 15B shows a block diagram of an alternative way to solve the above equations. This module 1550 uses Equation [3] and [7] to solve for ambient noise content $N_A(\omega)$, and then the voice content $A(\omega)$ in the input from the air microphone x_A , of which the estimate is denoted by $\bar{A}(n, \omega)$. After the transfer function $H(\omega)$, noise $N_e(\omega)$, and $\alpha(\omega)$ have been obtained in a separate training process (see Figure 19A below) and stored in a local memory 1554, the general process of the Figure 15B module is:

30

- Smooth $X_A(n, \omega)$ and $X_B(n, \omega)$ at block 1552 by averaging each of them with the

corresponding k previous frames, i.e., $X_{A,k}(n, \omega) = \frac{\sum_{i=0}^k X_A(n-i, \omega)}{k+1}$ and $X_{B,k}(n, \omega) = \frac{\sum_{i=0}^k X_B(n-i, \omega)}{k+1}$.

- Solve for $N_A(n, \omega)$ at block 1556 using $N_A(n, \omega) = \frac{X_{A,k}(n, \omega) - H(\omega)(X_{B,k}(n, \omega) - N_e(\omega))}{1 - \alpha(\omega)}$.
- For unreliable α (when $|1 - \alpha| < \alpha_t$, where α_t is a threshold), $N_A(n, \omega)$ is obtained by interpolating from its reliable neighbors at block 1558.
- Estimate $A(n, \omega)$ at block 1560 using $A(n, \omega) = X_{A,k}(n, \omega) - N_A(n, \omega)$.

The two solutions for the second configuration (Figure 14A-B) take advantage of the modules 1500 (Figure 15A) and 1550 (Figure 15B). Both these solutions require a short training step, which utilizes module 1500 (Figure 15A) in different ways. First consider the training step which is shown at Figure 16A, and with Figure 16B representing the diagram in operation after training. The equalizing transfer function $H(\omega)$, generated by the additional air microphone x_A and the training step, is applied after the spectral subtraction stage 1610 (Figure 16B). So in the training stage of Figure 16A the process changes into:

- Record the user's voice in a quiet environment.
- Frame both sensor inputs and do DFT at block 1602.
- Apply Module 1500 (Figure 15A).
- Store $H(\omega)$ and $N_e(\omega)$ to the memory 1604.

20

Once the training stage is complete and the initial transfer function $H(\omega)$ and noise $N_e(\omega)$ are stored, operation in practice for the first solution of the second configuration follows Figure 16B:

- Frame the vibration sensor input and do DFT at block 1602.
- Send the data into VAD 1606, which separates the voiced and noise-only frames.
- In those noise-only frames, update the noise profile (estimated noise spectrum) at block 1608.
- In all frames, perform spectral subtraction at block 1610 to remove the noise spectrum.
- Apply the equalizing transfer function from the memory 1604 on the data at block 1612.

30

- Apply an overlap-add process at block 1614 and transform the data back into the time domain to form the output.

The training step of Figure 16A could be either pre-programmed or during the daily use. For the pre-programmed option the manufacturer can store a set of equalizing transfer functions in the device and allow the user to choose the one that offers the best sound quality. For the at-use training option, when the user is speaking the air microphone x_A can detect the ambient sound level. Once it is below a certain level for a certain period of time, meaning that the environment is quiet, the training process would start automatically. Furthermore, since the air microphone x_A is only used in the training step, in practice the device could offer two modes: “noise-free” (x_B only) and “all-in” (x_A only), in which the user is then free to choose either of them.

The results from Figure 16B are presented at Figures 17A-B (time domain) and 18A-B (frequency domain). Figures 17A and 18A give the signal from the air microphone x_A (representing the conventional single air microphone configuration shown at Figure 2) and Figures 17B and 18B give the system output. This embodiment generated an output with 25.2 dB SNR, which means an improvement of 20.7 dB over the x_A input. Comparing Figures 17B and 18B with Figures 12B and 13B from the single channel configuration, the transfer function $H(\omega)$ not only brings in another 5 dB improvement in SNR but also amplifies the voice frequencies, thus improving the overall sound quality as shown by comparing those figures.

The block diagram to implement the second solution for the dual channel configuration is shown at Figures 19A (training) and 19B (usage). The training step in this solution uses recordings in two different sound environments: speaking in a quiet field and silence in a highly noisy field. The training procedure for Figure 19A is then as follows:

- When the user is speaking in a quiet environment (upper branch of Figure 19A), apply Module 1500 (Figure 15A), and compute $H(\omega)$ and $N_e(\omega)$.
- When the user is silent with a high-amplitude background noise (lower branch of Figure 19A), apply Module 1500 (Figure 15A), and compute the noise transfer function $\beta(\omega)$.

- Compute $\alpha(\omega)$ at block 1906 and store it to the memory 1908 together with $H(\omega)$ and $N_e(\omega)$.

Usage:

- 5
 - Frame both sensor inputs and do DFT at block 1902.
 - Send the data into voice activity detector 1904, which separates the voiced and noise-only frames.
 - Apply Module 1550 (Figure 15B).
 - Further enhance the spectrum by smoothing at block 1910.
- 10
 - Apply overlap-add process at block 1912 and transform the data back into time domain to form the output.

Similar with the solution of Figures 16A-B, data from the training step could be either pre-installed or generated during the user's daily use. For the at-use training, the air microphone x_A can classify the sound environments at block 1914 and enable/disable the training automatically, or alternatively the user could manually control it and start/stop the training when he/she feels it might be necessary.

The results from Figure 19B are presented at Figures 20A-B (time domain) and 21A-B (frequency domain). Figures 20A and 21A give the signal from the air microphone x_A (representing the conventional single air microphone configuration) and Figures 21B and 21B give the system output which has an SNR at 27.4 dB. That is a 22.9 dB improvement over the raw air microphone signal x_A .

Table 2 below shows the SNR improvement results from the first and second solutions for both the first embodiment (Figures 1A-B) and the second embodiment (Figures 14A-B) detailed above. Other multi-channel noise reduction technologies and acoustical treatments can also be applied to improve the performance.

Table 2: Performance comparison of the 4 solutions in SNR improvement

	Air Microphone x_A	Contact Sensor x_B	Single channel		Dual channel	
			Solution #1	Solution #2	Solution #1	Solution #2

SNR (dB)	4.5	12.7	24.0	20.5	25.2	27.4
SNRI (dB)	0	8.2	19.5	16.0	20.7	22.9

Now is described the third embodiment which is shown by example at Figures 22A-C as implemented in eyeglasses 2200. In this third embodiment there are two different types of sensors mounted on the spectacle frame 2200. Specifically, Figure 22A includes three air microphones and one vibration sensor in the eyeglasses 2200 or other platform. The vibration sensor is in one embodiment a mechanical vibration to electric transducer as with Figures 1A-B and Figures 14A-B. The vibration sensor may be considered more generally as a voice activity sensor in this third embodiment, and therefore could be a simplified version of the sensor used in previous embodiments. Another implementation of that is a downward facing camera to detect when the user's lips are moving, disposed for example along the underside of one or both eyepiece portions of the front frame 2202 at the illustrated alternate location 2208. Other embodiments of the voice activity sensor is an ultrasonic sensor or an infrared sensor.

15

Air microphone x_1 is at the center of the front frame 2202 on the bridge 2206 and above the user's nose when the spectacles are worn. Air microphones x_2 and x_3 are each on opposed spectacle arms 2204A, 2204B, directly above the user's ears when the spectacles are worn. Other embodiments detailed further below may have only one air microphone mounted along one of the arms 2204A, 2204B in addition to the air microphone mounted at the front frame 2202; or still further embodiments may employ multiple microphones along the front frame 2202.

20

The vibration sensor x_B is installed in the nose pad 2206 in the Figure 22A embodiment, similar to Figure 1A. Or in another implementation of this third embodiment shown at Figure 22B, the vibration sensor x_B may be disposed along the inboard side of one of the spectacle arms 2204A-B near the distal end and adjacent to the user's ear similar to that shown in Figure 1B so as to be in physical contact with the user's skull at the mastoid.

25

In the embodiments for other types of head-mounted devices, such as headbands and helmets, the air microphones could remain at the same locations as shown for Figures 22A-B, and the vibration sensor x_B could be deployed on the user's forehead, which is an ideal location for bone conducted voice pickup. For a helmet the primary voice microphone(s) x_1 may be deployed along the underside of the lip running across the user's forehead. In other spectacle embodiments, the primary voice microphone x_1 may be disposed along the underside of the front frame 2202 beneath the eyepiece, or two may be so deployed with one under each eyepiece.

For any of the implementations at Figures 22A-B, air microphone x_1 is placed closest to the user's mouth to serve its primary purpose of capturing the user's voice. Of course the signal from this microphone x_1 would still be corrupted by ambient noise that may be present. The primary purpose of microphones x_2 and x_3 is to capture ambient noise, which is the reason for their location away from the user's mouth which helps assure the signals which they capture and output are predominantly noise.

Linear combinations of these three air microphone inputs form the two data channels for signal processing as is shown by the block circuit diagram of Figure 23, which is an exemplary noise cancellation circuit 2300 though it may be implemented by a tangibly embodied signal processing software program. There is also a third (auxiliary) channel input to that circuit from the vibration sensor x_B which detects the vibration in the skull generated by the user's speaking activity

Four signals are input to the circuit 2300 of Figure 23. A forward microphone signal $x_1(t)$ is from the microphone x_1 mounted on the front frame 2202; two lateral microphone signals $x_2(t)$ and $x_3(t)$ are from the respective microphones x_2 and x_3 mounted on either of the spectacle arms 2204A-B, and a contact signal $x_B(t)$ is input from the vibration sensor x_B or is the signal combined from multiple such sensors. These are manipulated in the noise reduction circuit 2300 which is a modified adaptive interference canceller (AIC), having an adaptive filter w as its kernel (reference #2304). The input signal $x_1(t)$ from the forward microphone x_1 is weighted at a combiner 2302 to form a primary channel y_1 :

$$y_1(t) = a_1 x_1(t) \quad [10]$$

The three microphone signals $x_1(t)$, $x_2(t)$, and $x_3(t)$ are also combined at a combiner 2302 and the weighted difference between them forms a secondary channel $y_2(t)$, which may be considered as a noise reference channel:

$$y_2(t) = b_2 x_2(t) + b_3 x_3(t) - b_1 x_1(t) \quad [11]$$

Then the noise reference channel $y_2(t)$ is fed into the adaptive filter w and processed to match the noise contained in the primary channel $y_1(t)$ as close as possible. The output of the adaptive filter w is then a filtered noise $y_r(t)$ which in the ideal represents noise in the primary channel $y_1(t)$. There are various algorithms known in the art which can be used to implement the adaptive filter w , including but not limited to least mean squares (LMS), recursive least squares (RLS), etc. The inventors' experiments have found that normalized least mean squares (NLMS) works quite effectively as the error criterion.

The filtered noise $y_r(t)$ is then subtracted from the primary channel $y_1(t)$, as delayed $y_d(t)$ by block 2306 to account for processing delay in the adaptive filter w so as to match in time the filtered noise channel $y_r(t)$ to the delayed primary channel $y_d(t)$. In this manner the noise component is canceled from the primary channel $y_1(t)$, ideally leaving only the user's voice signal which was captured at the microphone x_1 mounted on the front frame 2202. But the cancellation is not ideal in a practical system and so the result from the combiner that follows the delay block 2306 is termed an error signal, obtained as:

$$e(t) = y_d(t) - y_r(t) \quad [12]$$

This error signal $e(t)$ is then applied to two branches. One branch is immediately fed back into the adaptive filter w . The other branch is input to a high-pass filter (HPF) 2308 whose output forms the final output $z(t)$. While it is an error signal for the adaptive filter w , in fact it represents the voice signal captured at the microphone x_1 with some of the ambient noise removed; it is error only so far as some noise might still remain, thus the final output $z(t)$. The HPF 2308 is not required in all implementations, but is applied to remove the low frequency components outside the

regular telephone audio band. The HPF 2308 may be implemented to also filter out part of the noise residue and even improve the performance.

5 An important part of the overall system is the contact signal $x_B(t)$ from the vibration sensor x_B , which is input into a voice activity detector (VAD) 2310. The vibration sensor x_B is a bone microphone and so it only picks up vibrations from the user's skull and thus its signal is free of ambient sound. Though its sound quality is not as good as a regular air microphone, its purpose is to detect precisely when the user is vocalizing. The similar principle is in effect when the voice activity sensor is
10 implemented as a downward facing camera or a motion detector; it identifies those frames in which the user is talking.

Since the adaptive filter 2304 would not converge if the target signal (speech) and the interfering signal (noise) exist simultaneously, it can only adapt during noise-only
15 periods. Therefore, the VAD 2310 operates as a switch to suspend the filter 2304 adaptation as soon as the user starts speaking, and to enable the filter 2304 to continue adapting in the unvoiced frames when he/she is not speaking. During the suspension periods, the filter w will be fixed to the final value from the last adaptation iteration. As noted above, the vibration sensor x_B is not the only way to implement the VAD
20 switching of filter adaptation; other types of sensors can be used such as for example a downward facing camera or a motion detector mounted on the front frame 2202 to detect facial movement during speech. In general these and other implementations may be referred to more generally as voice activity sensors, distinguished from conventional over-the-air microphones in that voice activity sensors are immune to
25 ambient noise.

Ideally, after the adaptive filter w converges the additive noise in the primary channel $y_1(t)$ would be completely removed by the subtraction, and therefore the system output $z(t)$ would be completely noise-free. One effective measure of the system's
30 performance is signal-to-noise ratio improvement (SNRI), which is the difference between the output SNR and the input SNR. This is shown at Figures 24A-B and 25A-B.

The data for Figures 24A-B and 25A-B were gathered from experiments directed toward evaluating the performance of an apparatus somewhat similar to that shown at Figure 22A, except the experimental spectacles had five omni-directional air microphones on the frame rather than three. The test apparatus for the data of Figures 24A-B and 25A-B had three air microphones mounted at the locations shown at Figure 22A and two additional microphones mounted to the spectacles near the front end of the spectacle arms 2204A, 2204B near the user's temples, with one additional air microphone on each arm 2204A, 2204B. One of these additional air microphones was used to capture a reference channel, representing the conventional single-microphone configuration for head-mounted voice capture such as at Figure 2. In the test arrangement there were eight loudspeakers disposed about an office environment to simulate practical ambient noise environments, including babble, subway, street, etc. Data was collected from different individuals wearing these testing spectacles to normalize for different voice characteristics, each of whom spoke as if for a normal phone conversation.

The output of the experimental system is compared with that of the conventional single-microphone setup of Figure 2 having only one air microphone disposed on the left spectacle arm. Figures 24A and 25A give results for the single-microphone reference (Figure 2), while Figures 24B and 25B give results for the tested spectacles which utilized the vibration sensor x_B , three air-microphones shown at Figure 22A, and filter adaptation using the VAD 2310. Figures 24A and 24B show the comparison in the time domain, while Figures 25A and 25B show the same underlying data compared in the frequency domain.

Evident in Figures 24A-B is a digital on/off signal line 2402 labeled at Figure 24A, which is generated by the VAD 2310 and is superimposed as the square wave over the actual data of the microphone signal. Vertical transitions of that VAD signal line 2402 divide the microphone signal into voiced 2404 and unvoiced (noise-only) 2406 sections, those sections being specifically labeled at Figure 24B. In this evaluation the experimental implementation of the inventive spectacles reached roughly 18 dB in SNRI. This was only a proof of concept evaluation and the basic concept presented at the circuit 2300 of Figure 23 could be further improved using other multi-channel

noise reduction technologies and with more extensive post processing and acoustical treatments.

Figure 26A-B is a perspective view of an example implementation of a two-way
5 (capture & playback) audio spectacles being worn by a user, according to the second
embodiment/dual-channel configuration. This is similar to Figure 14A, except there
are additionally two bone vibration transducers 2602 to provide sound from the
spectacles to the user, rendering the spectacles a completely ear-free two-way audio
interface. In this implementation there is a bone vibration transducer mounted
10 towards the distal ends of each spectacle arm, furthest from the front frame so as to
lie in contact with the user's temporal bone portion of the skull, either above the ear
(squama temporalis) or behind the ear on the mastoid (pars mastoidea). Figure 26A-B
provides an example solution on how to achieve two-way audio, and the same
vibration transducer deployment can be applied to the first and the third embodiment.
15 From the above test results the inventors conclude that the two-way implementation in
Figure 26A-B is the most effective.

For completeness, also shown at Figure 26B is a schematic block diagram of a user
20 equipment UE 10 for illustrating a mobile terminal to which the head-mounted device
2600 may be in contact via Bluetooth or other personal area network wireless link 15.
While only one wireless link 15 is shown there may be more, each representing
multiple logical and physical channels.

25 The UE 10 includes a controller, such as a computer or a data processor (DP) 10A, a
computer-readable memory medium embodied as a memory (MEM) 10B that stores a
program of computer instructions (PROG) 10C, and a suitable radio frequency (RF)
transmitter 10D and receiver 10E for bidirectional wireless communications over
cellular links via one or more antennas 10F (two shown). The UE 10 is also shown as
30 having a Bluetooth module 10G, whose antenna may be inbuilt into the module.
While not particularly shown at Figure 26, the head-set device 2600 also includes a
Bluetooth or other personal area network module for wireless communications at
short range with the UE 10, as well as a processor/DP and a memory/MEM storing a

program/PROG for implementing the circuits and functional blocks detailed above.

More practical deployments would dispose the circuitry and modules detailed above into the head-set device 2600 and only the system output signal is sent wirelessly to the UE 10, but it is possible to send the raw microphone signals to the UE 10 in which the described circuitry and processing blocks are disposed, assuming the wireless link 15 supports more than one bi-directional audio channel. At least one of the PROGs in the headset device 2600 or in the UE 10 is assumed to include program instructions that, when executed by the associated DP, enable the device to operate in accordance with the exemplary embodiments of this invention, as detailed above. That is, the exemplary embodiments of this invention may be implemented at least in part by computer software executable by the DP of the UE 10 and/or by the DP of the head-set device 2600, or by hardware, or by a combination of software and hardware (and firmware).

15

In general, the various embodiments of the UE 10 can include, but are not limited to, cellular telephones, personal digital assistants (PDAs) having wireless communication capabilities, portable computers having wireless communication capabilities, image capture devices such as digital cameras having wireless communication capabilities, gaming devices having wireless communication capabilities, music storage and playback appliances having wireless communication capabilities, Internet appliances permitting wireless Internet access and browsing, as well as portable units or terminals that incorporate combinations of such functions.

25 The computer readable MEM in the UE 10 and in the head-set device 2600 may be of any type suitable to the local technical environment and may be implemented using any suitable data storage technology, such as semiconductor based memory devices, flash memory, magnetic memory devices and systems, optical memory devices and systems, fixed memory and removable memory. The DPs may be of any type suitable to the local technical environment, and may include one or more of general purpose computers, special purpose computers, microprocessors, digital signal processors (DSPs) and processors based on a multicore processor architecture, as non-limiting examples.

30

Some of the above embodiments may be generalized as follows. There is determined, from at least a first electrical audio signal from a first transducer, voiced frames during which a user is speaking and unvoiced frames during which the user is not speaking. In this case the first transducer is disposed on a platform configured to be removably placed symmetrically on or about a user's head, and the first transducer is further configured to capture vibration of the user's skull or facial movement generated by the user's voice activity and to detect the user's speaking activity. The first electrical audio signal received at the platform is then processed as the voiced frames and/or as the unvoiced frames.

In the first solution for the single channel embodiment, the first electrical audio signal received at the platform is from the first transducer which is a vibration sensor, and the one or more circuitries and/or one or more memories including computer program code processes the voiced frames by low-pass filtering and artificially extending a bandwidth thereof.

In the second solution for the single channel embodiment, the first electrical audio signal received at the platform is from the first transducer which is a vibration sensor, and the one or more circuitries and/or one or more memories including computer program code processes the unvoiced frames as noise-only frames for updating a noise profile and processes the voiced frames by spectrally subtracting the noise profile therefrom.

For the dual or multi channel embodiment, there is further at least a second transducer which is an air microphone, and second audio electrical signals received at the platform are differentially processed by the one or more circuitries and/or one or more memories including computer program code as voiced and unvoiced frames.

In the first solution for the dual channel embodiment, an equalizing function, which is the transfer function between clean voice content captured by the first and the second transducer, is computed by a separate training process of the one or more circuitries and/or one or more memories including computer program code. In this case the one

or more circuitries and/or one or more memories including computer program code processes the unvoiced frames as noise-only frames for updating a noise profile obtained only from the first electrical audio signal from the first transducer which is a vibration sensor, and processes the voiced frames by spectrally subtracting the noise
5 profile therefrom and applying the pre-computed equalizing transfer function to enhance the output spectrum.

In the second solution for the dual channel embodiment, two equalizing functions [which are a) the transfer function between the clean voice content captured by the first and the second transducer, and b) a transfer function between ambient noise
10 content captured by the two transducers and an estimate of electronic noise of the first transducer] are computed by a separate training process of the one or more circuitries and/or one or more memories including computer program code. In this case the ambient noise content captured by the second transducer is estimated by utilizing
15 results of the training process. Additionally, the circuitry or embodied software processes the voiced frames to estimate a speech signal by separating therefrom the ambient noise content estimated from the output signals of the second transducer.

For the multi-channel embodiment, the platform comprises at least three air
20 microphones spatially disposed about the platform; and the one or more circuitries and/or one or more memories including computer program code is configured to output an adaptively filtered noise signal from at least inputs from side-mounted ones of the air microphones. This adaptive filtering produces an error signal which remains after subtracting the filtered noise signal from an output signal of a forward-
25 mounted one of the air microphones, and the adaptive filtering is dynamically adaptive only during the unvoiced frames and static during the voiced frames.

As was detailed above, for the multi-channel embodiment the first transducer is one of a vibration sensor, a downward facing camera, an ultrasonic sensor and an infrared
30 sensor. And also in any of the above embodiments the platform comprises one of eyeglasses, sunglasses, a helmet and a headband.

Consider again the above multi-channel embodiments at Figure 22A-B with the array

of three air microphones. While the described vibration sensor may be alternatively implemented as a downward facing camera, an ultrasonic sensor, an infrared sensor, and the like, the air microphone array by itself provides quite a few advantages independent of any voice activity sensor. For example, this array of air microphones
5 can be used to capture high-fidelity directional/surround sound in which the signals of all three air microphones are post-processed by some existing solutions, including the surround sound capture algorithm of Nokia, Oy (the assignee of the rights herein); and can be used for binaural recording in which the signals of only the side-mounted air microphones are post-processed by existing solutions like the binaural recording
10 algorithm of Nokia, Oy (the assignee of the rights herein).

In general, the various exemplary embodiments may be implemented in hardware or special purpose circuits, software, logic or any combination thereof. For example, some aspects may be implemented in hardware, while other aspects may be
15 implemented in embodied firmware or software which may be executed by a controller, microprocessor or other computing device, although the invention is not limited thereto. While various aspects of the exemplary embodiments of this invention may be illustrated and described as block diagrams, flow charts, or using some other pictorial representation, it is well understood that these blocks, apparatus, systems,
20 techniques or methods described herein may be implemented in, as non-limiting examples, hardware, embodied software and/or firmware, special purpose circuits or logic, general purpose hardware or controller or other computing devices, or some combination thereof, where general purpose elements may be made special purpose by embodied executable software.

25 The above embodiments are quite varied and so not all will exhibit the same technical advantages. Some of them such as those described with reference to Figure 26A-B enable a completely hands-free and mouth-free and ear-free solution for two-way voice communication, and with greatly enhanced voice quality and speech
30 intelligibility for its use in highly noisy environments. In some embodiments this setup is multi-functional; for example the three microphone arrays of Figures 22A-B can also be used to do binaural recording. Other embodiments provide some but not all of these technical advantages and may provide other advantages as detailed above

for the specific configurations. As already mentioned above, the technology detailed for the eyeglass platform can easily be implemented on other forms of head-mounted devices such as headbands and helmets.

- 5 Another advantage of certain of the above examples is that a spectacle-mounted Bluetooth accessory is more likely to be widely adopted and garner wider acceptance than current Bluetooth earpiece/microphone boom arrangements. The head-set mounted multi-microphone platform is flexible enough to offer the potential for multichannel audio technologies to become commercially available in the future, such
10 as for example integrated with a spectacle-based see-through display, gaze tracker, camera, and other types of sensors.

- It should thus be appreciated that at least some aspects of the exemplary embodiments of the inventions may be practiced in various components such as integrated circuit
15 chips and modules, and that the exemplary embodiments of this invention may be realized in an apparatus that is embodied as an integrated circuit. The integrated circuit, or circuits, may comprise circuitry (as well as possibly firmware) for embodying at least one or more of a data processor or data processors, a digital signal processor or processors, and circuitry described herein by example.

- 20 Furthermore, some of the features of the various non-limiting and exemplary embodiments of this invention may be used to advantage without the corresponding use of other features. As such, the foregoing description should be considered as merely illustrative of the principles, teachings and exemplary embodiments of this
25 invention, and not in limitation thereof.

CLAIMS:

We claim:

1. An apparatus comprising:
a platform, configured to be removably placed symmetrically on or about a
5 user's head;
at least a first transducer configured to capture vibration of the user's skull or
facial movement generated by the user's voice activity and detect the
user's speaking activity, in which the at least first transducer converts
the vibration or facial movement into a first electrical audio signal;
10 and
one or more circuitries and/or one or more memories including computer
program code for processing the at least the electrical audio signal
from the at least first transducer received at the platform as voiced
frames and/or as unvoiced frames, in which the voiced frames and/or
15 the unvoiced frames are defined based at least on the first electrical
audio signal.
2. The apparatus according to claim 1, in which the first electrical audio signal
received at the platform is from the first transducer which is a vibration sensor, and
20 the one or more circuitries and/or one or more memories including computer program
code processes the voiced frames by low-pass filtering and artificially extending a
bandwidth thereof.
3. The apparatus according to claim 1, in which the first electrical audio signal
25 received at the platform is from the first transducer which is a vibration sensor, and
the one or more circuitries and/or one or more memories including computer program
code processes the unvoiced frames as noise-only frames for updating a noise profile
and processes the voiced frames by spectrally subtracting the noise profile therefrom.
- 30 4. The apparatus according to claim 1, in which the apparatus further comprises
at least a second transducer which is an air microphone, and second audio electrical
signals received at the platform are differentially processed by the one or more

circuitries and/or one or more memories including computer program code as voiced and unvoiced frames.

5. The apparatus according to claim 4, in which an equalizing function, which is
5 the transfer function between clean voice content captured by the first and the second transducer, is computed by a separate training process of the one or more circuitries and/or one or more memories including computer program code;
and in which the one or more circuitries and/or one or more memories including
computer program code processes the unvoiced frames as noise-only frames for
10 updating a noise profile obtained only from the first electrical audio signal from the first transducer which is a vibration sensor, and processes the voiced frames by spectrally subtracting the noise profile therefrom and applying the pre-computed equalizing transfer function to enhance the output spectrum.

15 6. The apparatus according to claim 4, in which two equalizing functions, which are the transfer function between the clean voice content captured by the first and the second transducer, and a transfer function between ambient noise content captured by the two transducers and an estimate of electronic noise of the first transducer, are
computed by a separate training process of the one or more circuitries and/or one or
20 more memories including computer program code;
in which the ambient noise content captured by the second transducer is estimated by utilizing results of the training process;
and in which the circuitry or embodied software processes the voiced frames to
estimate a speech signal by separating therefrom the ambient noise content estimated
25 from the output signals of the second transducer.

7. The apparatus according to claim 1, in which the apparatus further comprises
at least three air microphones spatially disposed about the platform; and
the one or more circuitries and/or one or more memories including
30 computer program code is configured to output an adaptively filtered noise signal from at least inputs from side-mounted ones of the air microphones,
in which the adaptive filtering produces an error signal which remains
after subtracting the filtered noise signal from an output signal of a

forward-mounted one of the air microphones;
and in which the adaptive filtering is dynamically adaptive only during the unvoiced frames and static during the voiced frames.

- 5 8. The apparatus according to claim 7, in which the first transducer is one of a vibration sensor, a downward facing camera, an ultrasonic sensor and an infrared sensor.
9. The apparatus as claimed in any previous claim, in which the platform
10 comprises one of: eyeglasses, sunglasses, a helmet and a headband.
10. A method comprising:
determining from at least a first electrical audio signal from a first transducer
voiced frames during which a user is speaking and unvoiced frames during which the
15 user is not speaking, in which the first transducer is disposed on a platform configured to be removably placed symmetrically on or about a user's head and is configured to capture vibration of the user's skull or facial movement generated by the user's voice activity and detect the user's speaking activity; and
processing the first electrical audio signal received at the platform as the
20 voiced frames and/or as the unvoiced frames.
11. The method according to claim 10, in which the first electrical audio signal received at the platform is from the first transducer which is a vibration sensor, and processing the first electrical audio signal comprises processing the voiced frames by
25 low-pass filtering and artificially extending a bandwidth thereof.
12. The method according to claim 10, in which the first electrical audio signal received at the platform is from the first transducer which is a vibration sensor, and processing the signals comprises processing the unvoiced frames as noise-only frames
30 for updating a noise profile and processing the voiced frames by spectrally subtracting the noise profile therefrom.

13. The method according to claim 10, in which the processing is further on a second audio electrical signal received from at least a second transducer disposed on the platform which is an air microphone.
- 5 14. The method according to claim 13, in which an equalizing function, which is a transfer function between clean voice content captured by the first and the second transducers, is computed by a separate training process;
and in which processing the first and second audio electrical signals comprises
processing the unvoiced frames as noise-only frames for updating a noise profile
10 obtained only from the first audio electrical signal from the vibration sensor, and
processing the voiced frames by spectrally subtracting the noise profile therefrom and
applying the pre-computed equalizing transfer function to enhance the output
spectrum.
- 15 15. The method according to claim 13, in which two equalizing functions, which
are a transfer function between clean voice content captured by the first and the
second transducers and a transfer function between the ambient noise content
captured by the two transducers, and an estimate of the electronic noise of the first
transducer, are computed by a separate training process;
20 in which the ambient noise content captured by the second transducer is estimated by
utilizing results of the training process;
and in which processing the first and second audio electrical signals comprises
processing the voiced frames to estimate a speech signal by separating therefrom the
ambient noise estimated from the output signals of only the second transducer.
- 25
16. The method according to claim 10, in which the platform comprises at least
three air microphones spatially disposed about the platform; and
processing the first audio electrical signal received at the platform
comprises outputting an adaptively filtered noise signal from at least
30 inputs from side-mounted ones of the air microphones, in which the
adaptive filtering produces an error signal which remains after subtracting
the filtered noise signal from an output signal of a forward-mounted one of
the air microphones;

and in which the adaptive filtering is dynamically adaptive only during the unvoiced frames and static during the voiced frames.

17. The method according to claim 16, in which the first transducer is one of a vibration sensor, a downward facing camera, an ultrasonic sensor and an infrared sensor.

18. The method as claimed in claims 10 to 17, in which the platform comprises one of: eyeglasses, sunglasses, a helmet and a headband.

19. A memory storing a program of computer readable instructions which when executed by at least one processor result in actions comprising:

determining from at least a first electrical audio signal from a first transducer voiced frames during which a user is speaking and unvoiced frames during which the user is not speaking, in which the first transducer is disposed on a platform configured to be removably placed symmetrically on or about a user's head and is configured to capture vibration of the user's skull or facial movement generated by the user's voice activity and detect the user's speaking activity; and

processing the first electrical audio signal received at the platform as the voiced frames and/or as the unvoiced frames.

20. An apparatus comprising:
a platform, configured to be removably placed symmetrically on or about a user's head;
at least two transducers disposed about opposed sides of the platform; and
circuitry or embodied software for resolving a spatial signal from signals output from the at least two transducers.

21. The apparatus according to claim 20, in which the resolved spatial signal is a binaural audio signal.

22. The apparatus according to claim 20, in which the platform comprises at least three transducers and the resolved spatial signal is a three dimensional surround sound signal.
- 5 23. A method comprising:
receiving signals from at least two transducers disposed about opposed sides of a platform, configured to be removably placed symmetrically on or about a user's head; and
resolving a spatial signal from the received signals.
- 10 24. The method according to claim 23, in which the resolved spatial signal is a binaural audio signal.
- 15 25. The method according to claim 23, in which the platform comprises at least three transducers and the resolved spatial signal is a three dimensional surround sound signal.

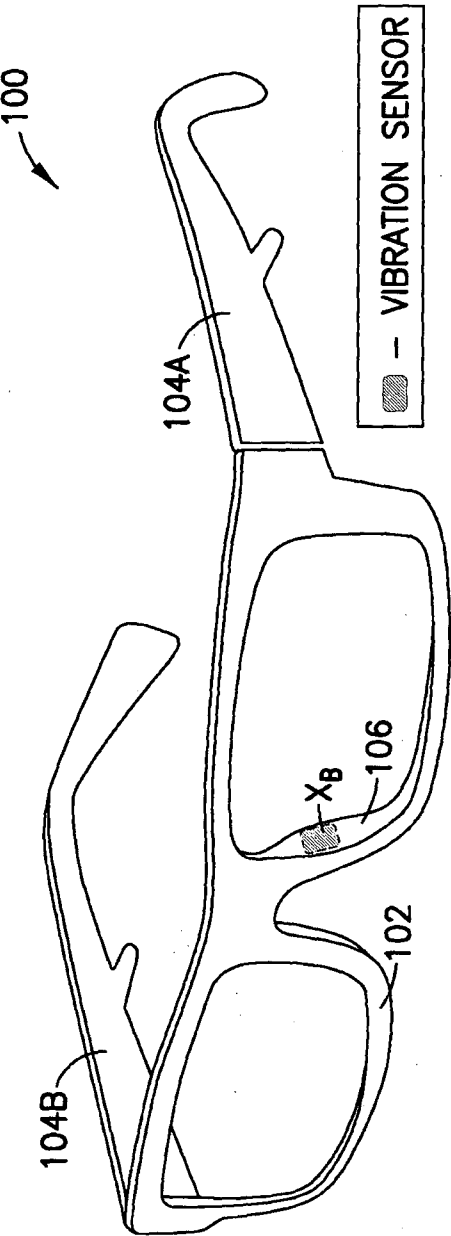


FIG. 1A

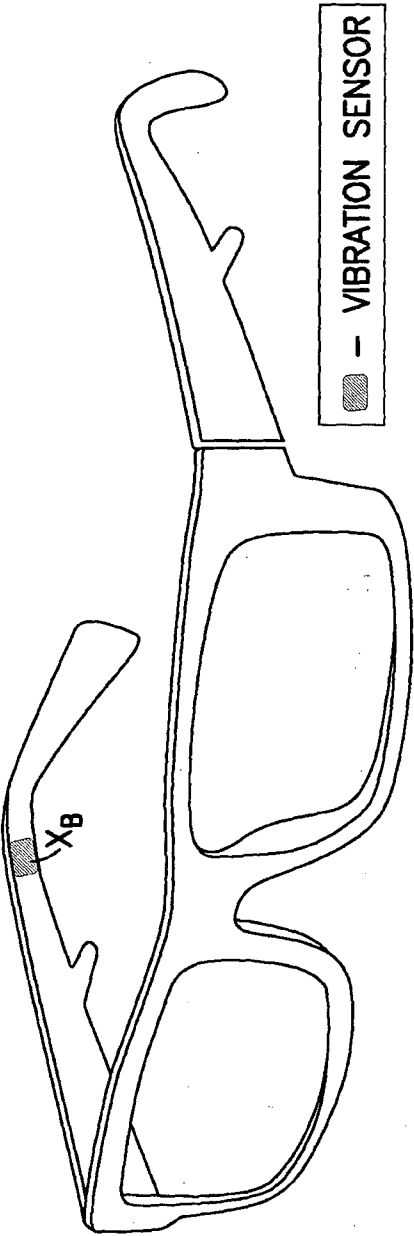


FIG. 1B

2/26

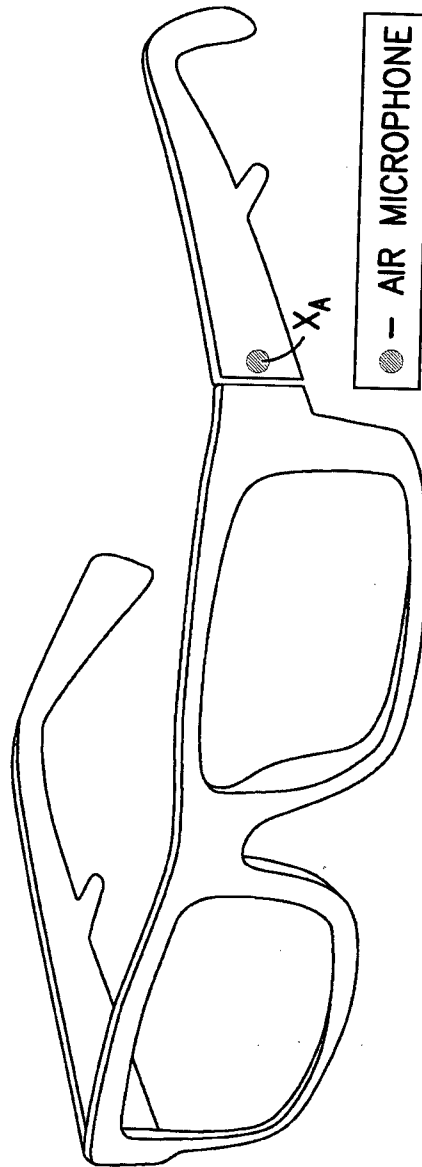


FIG. 2
PRIOR ART

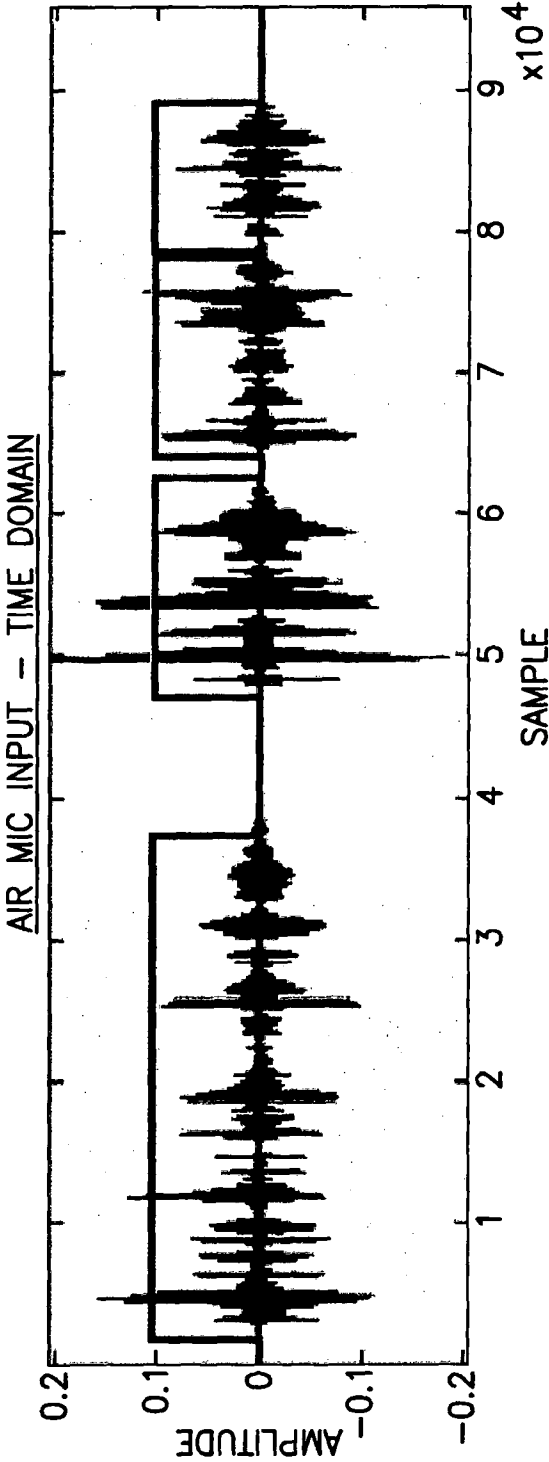


FIG. 3A

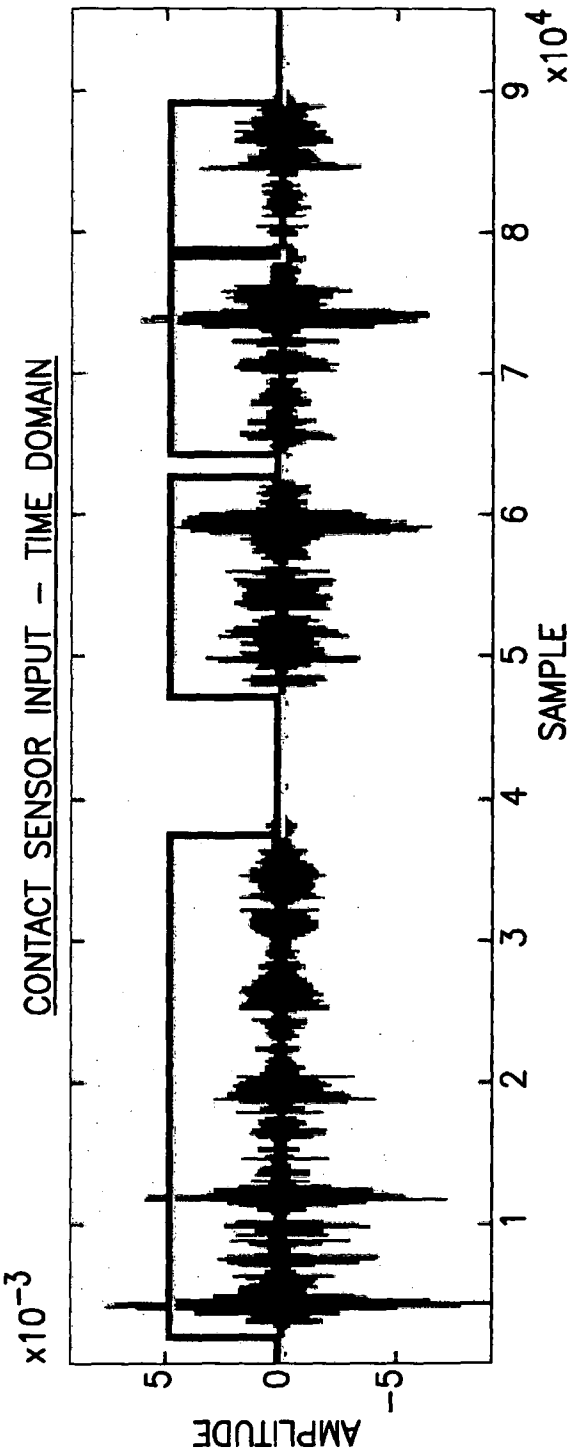


FIG. 3B

4/26

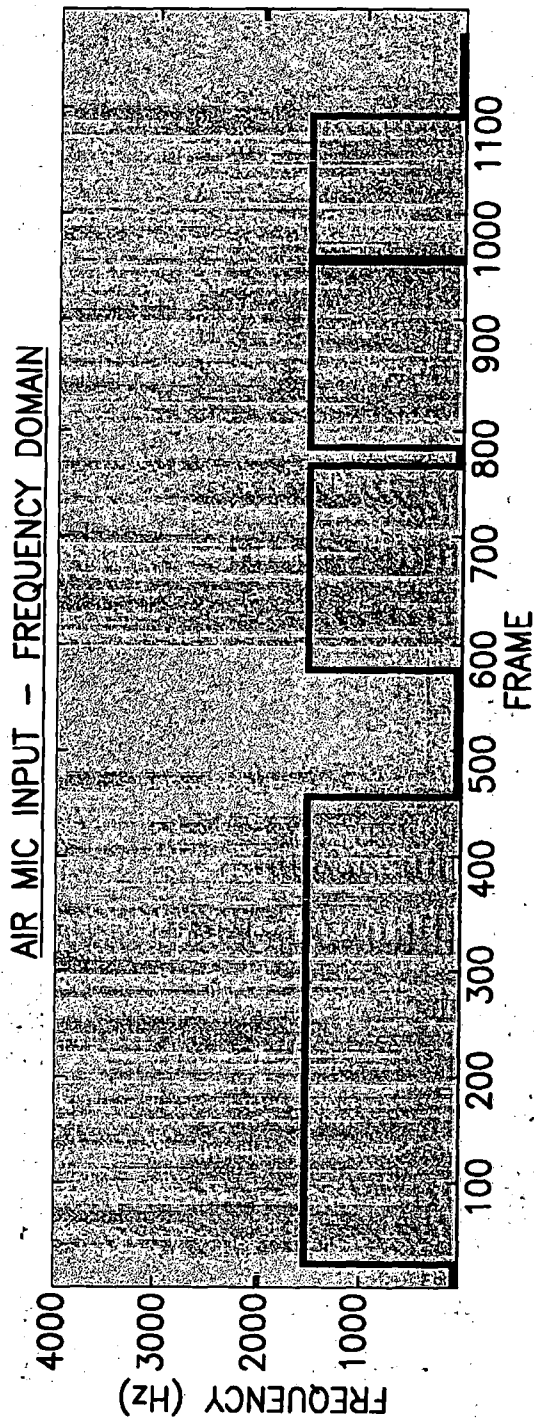


FIG.4A

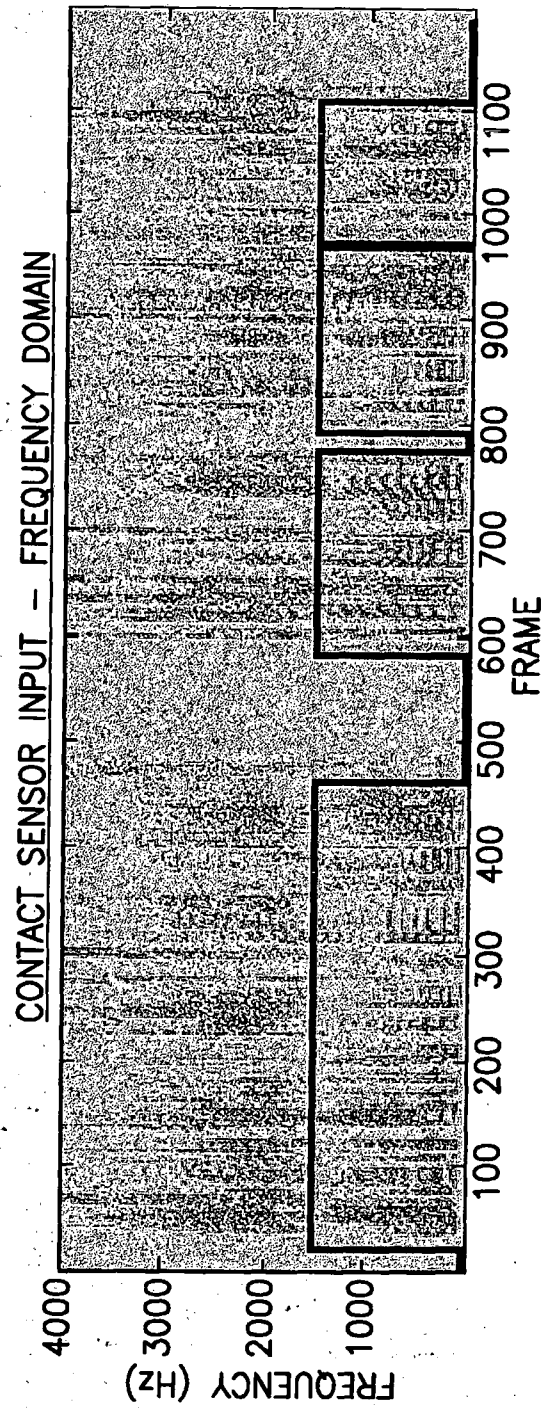


FIG.4B

5/26

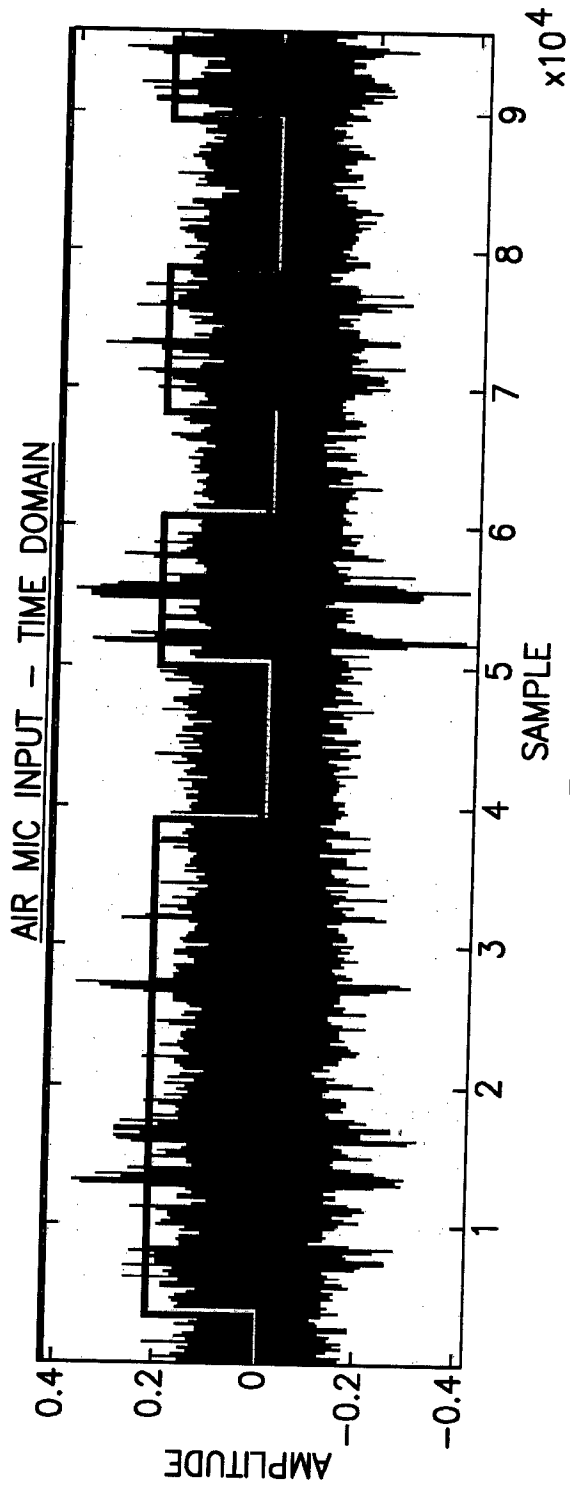


FIG. 5A

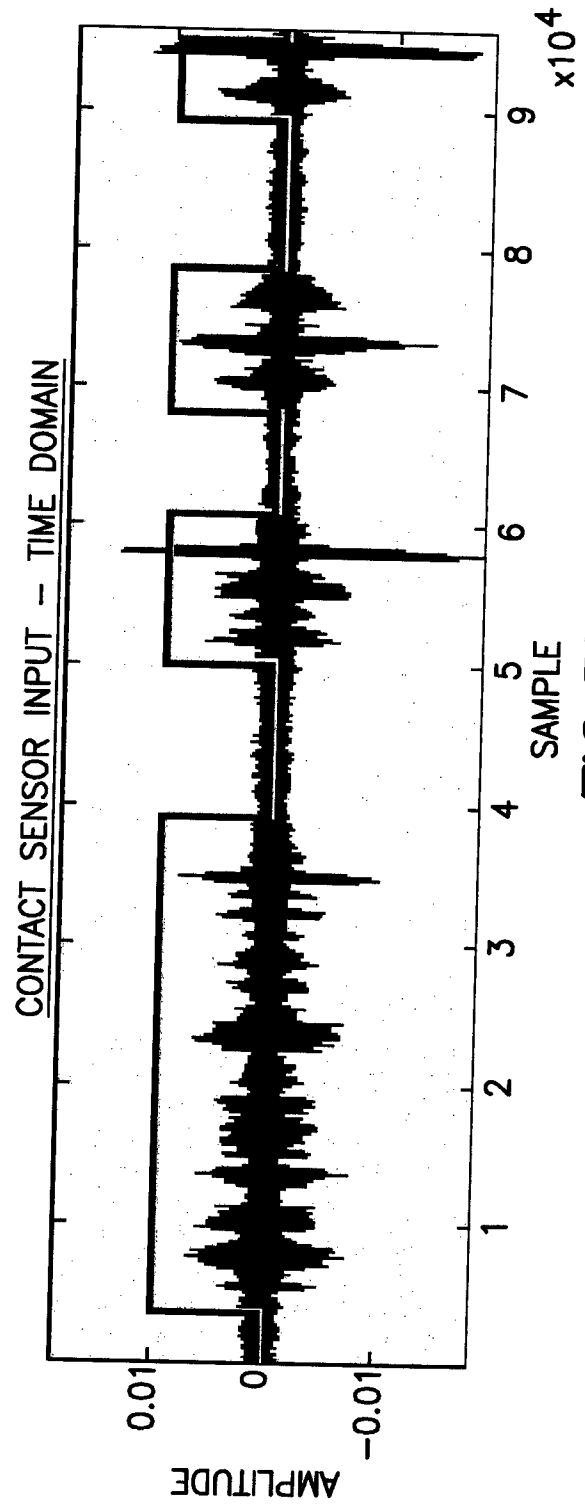


FIG. 5B

6/26

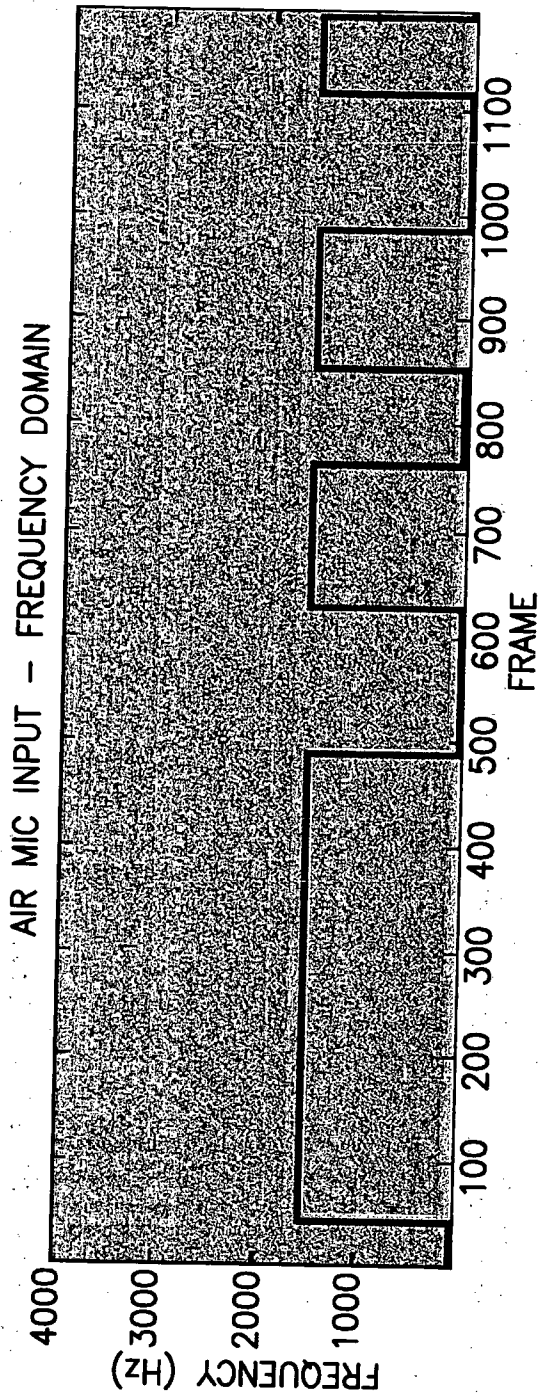


FIG. 6A

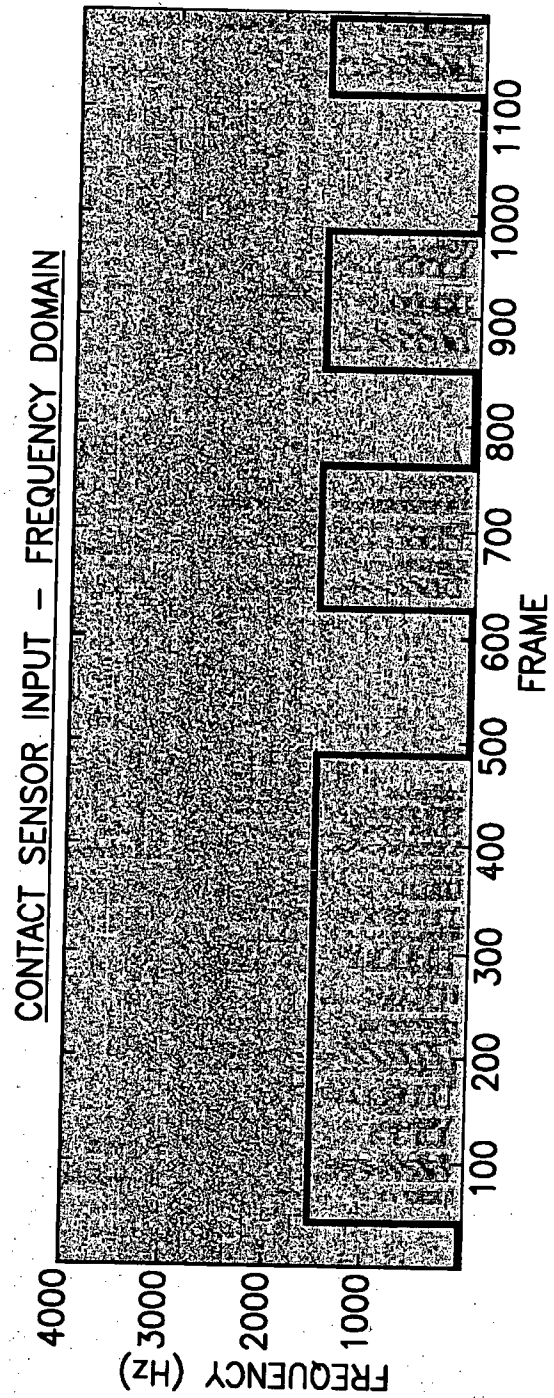


FIG. 6B

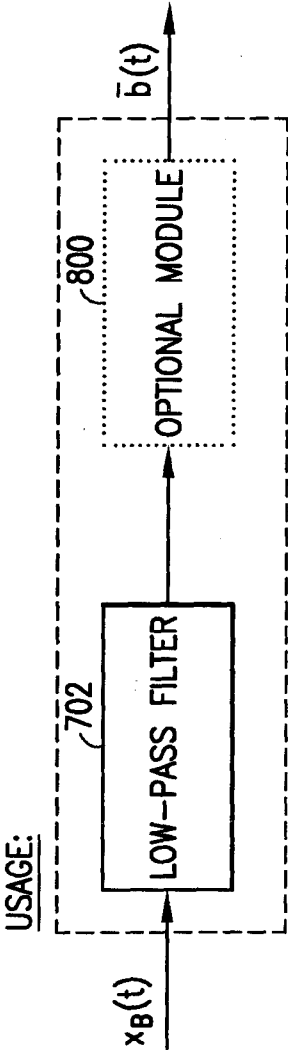


FIG. 7

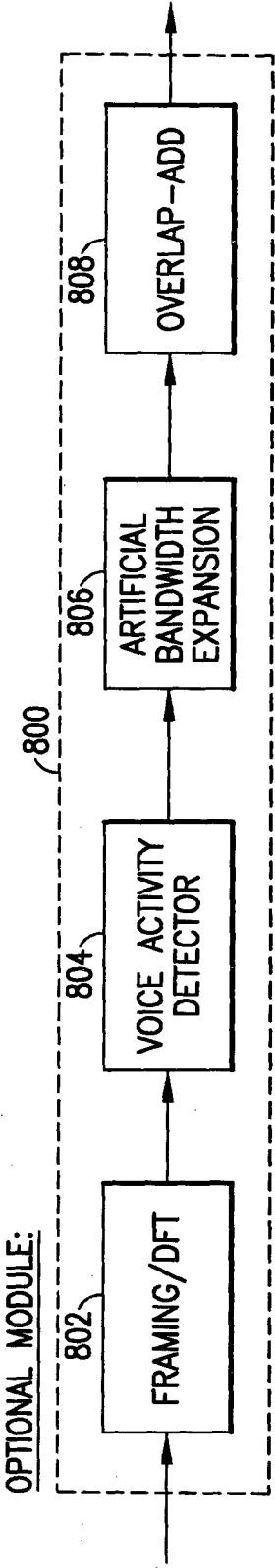


FIG. 8

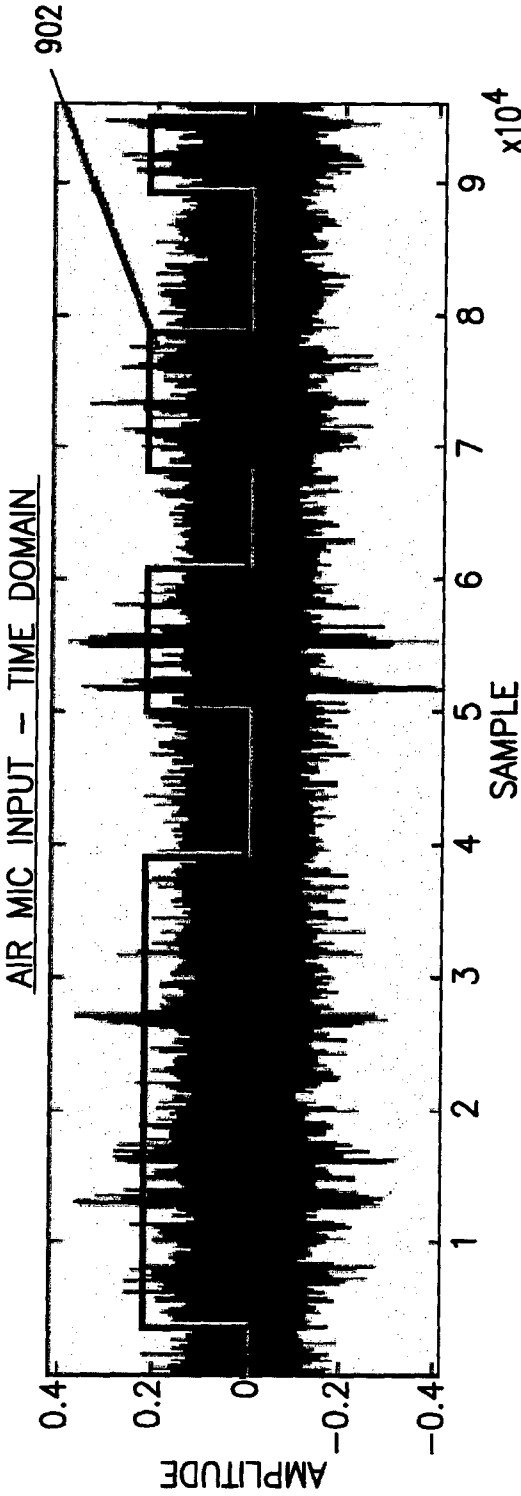


FIG. 9A

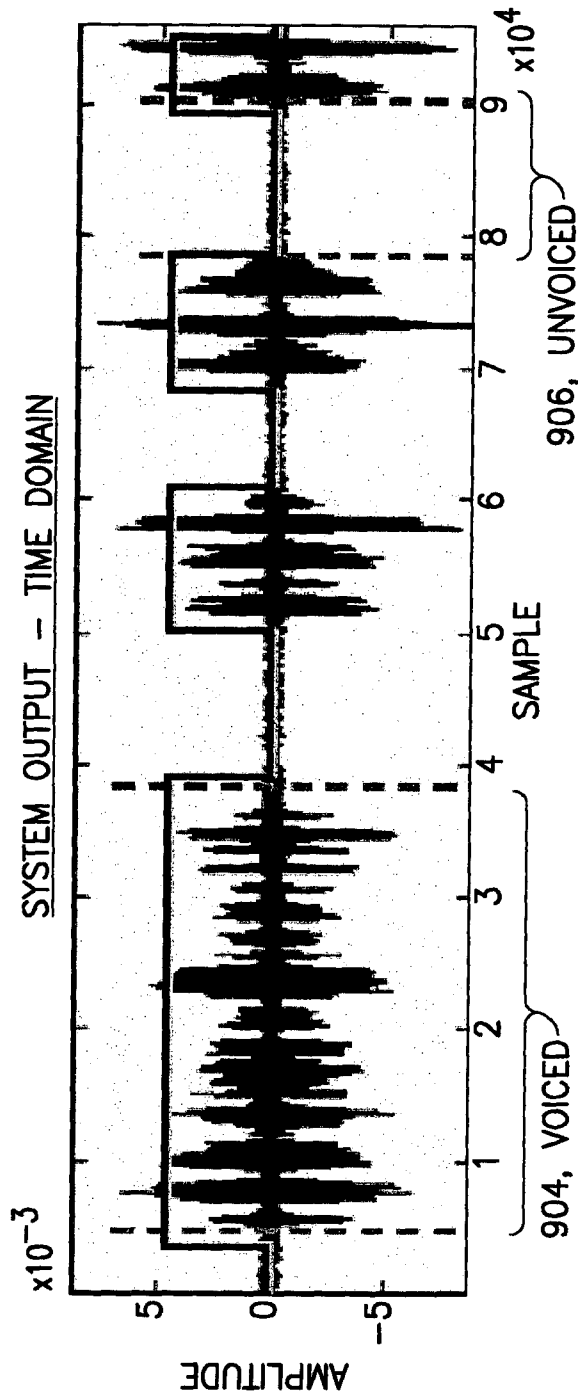


FIG. 9B

9/26

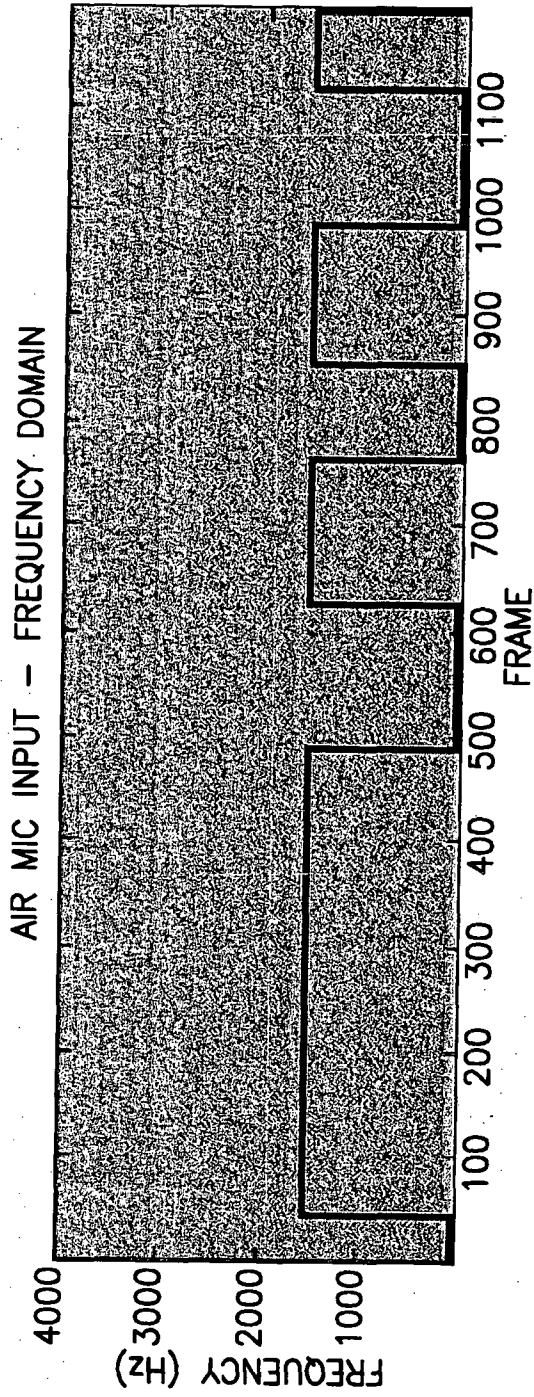


FIG.10A

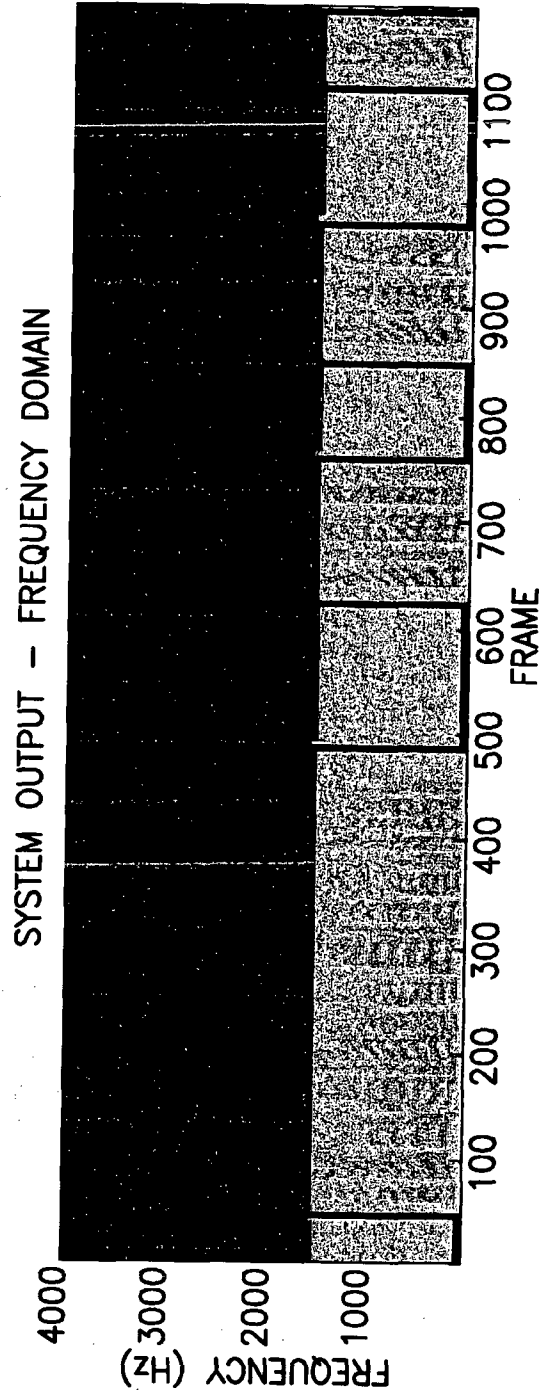


FIG.10B

10/26

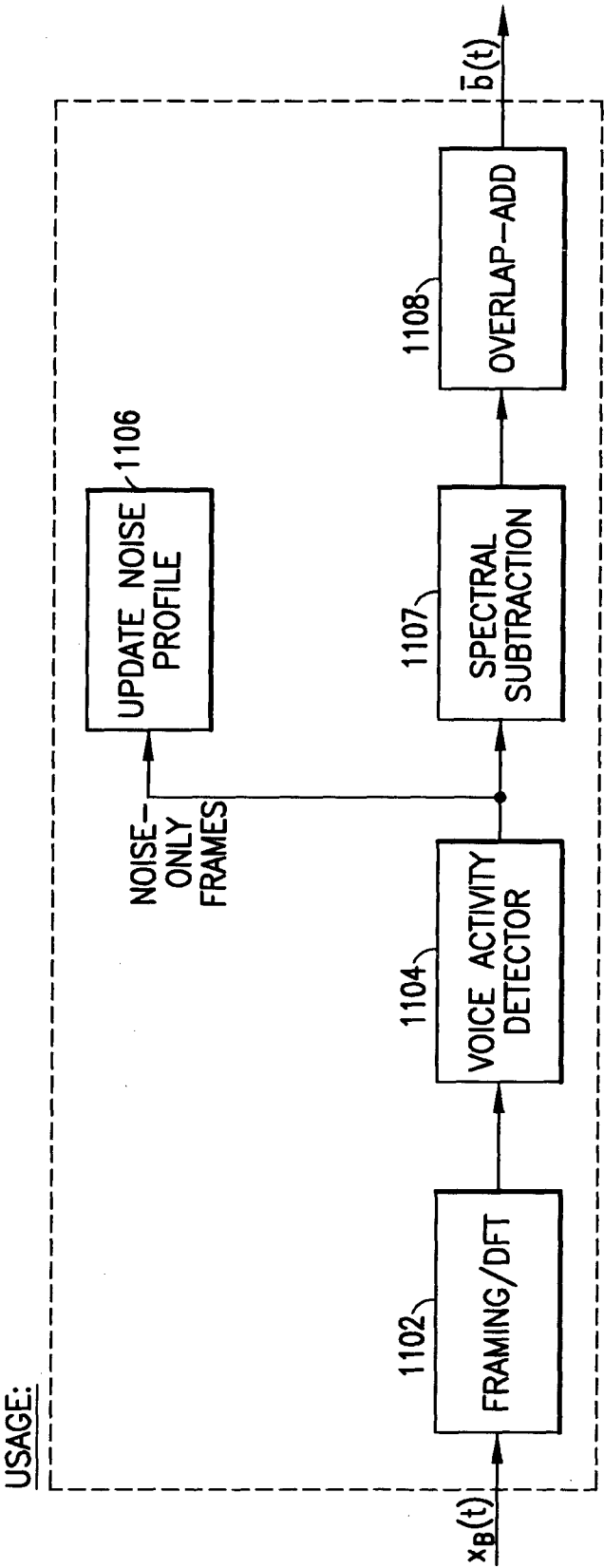


FIG.11

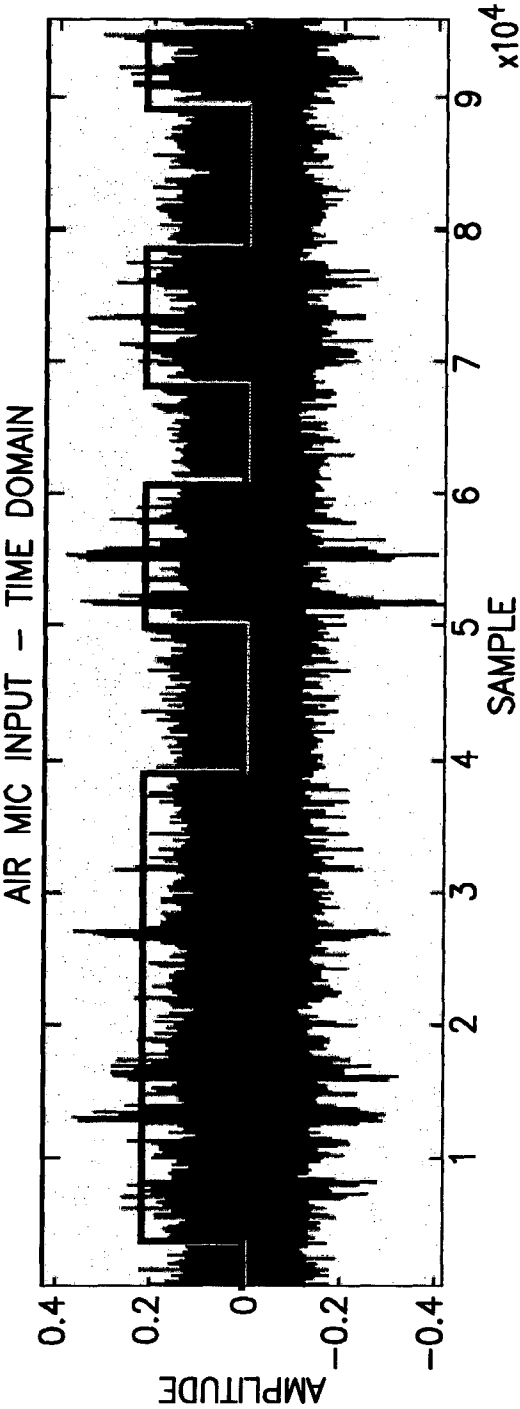


FIG. 12A

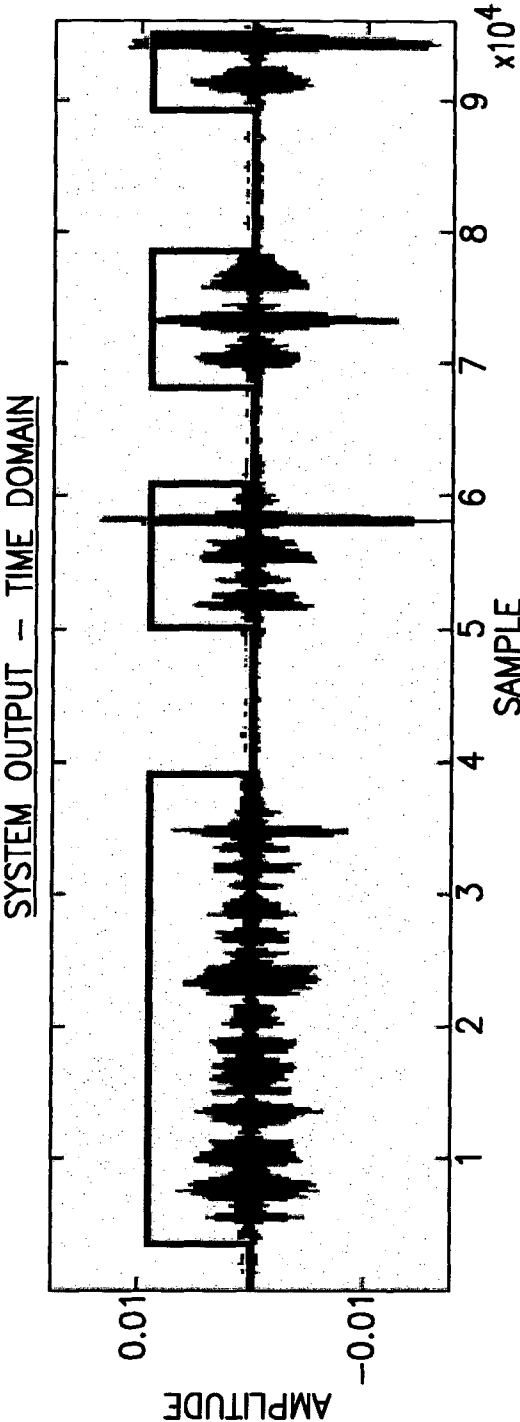


FIG. 12B

12/26

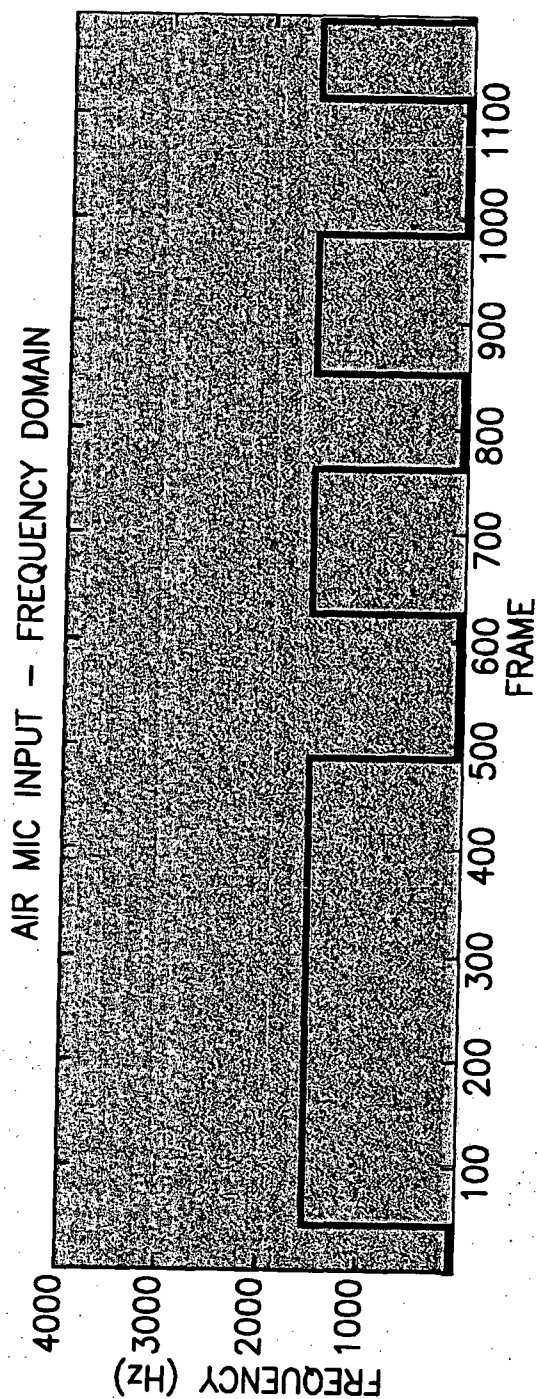


FIG.13A

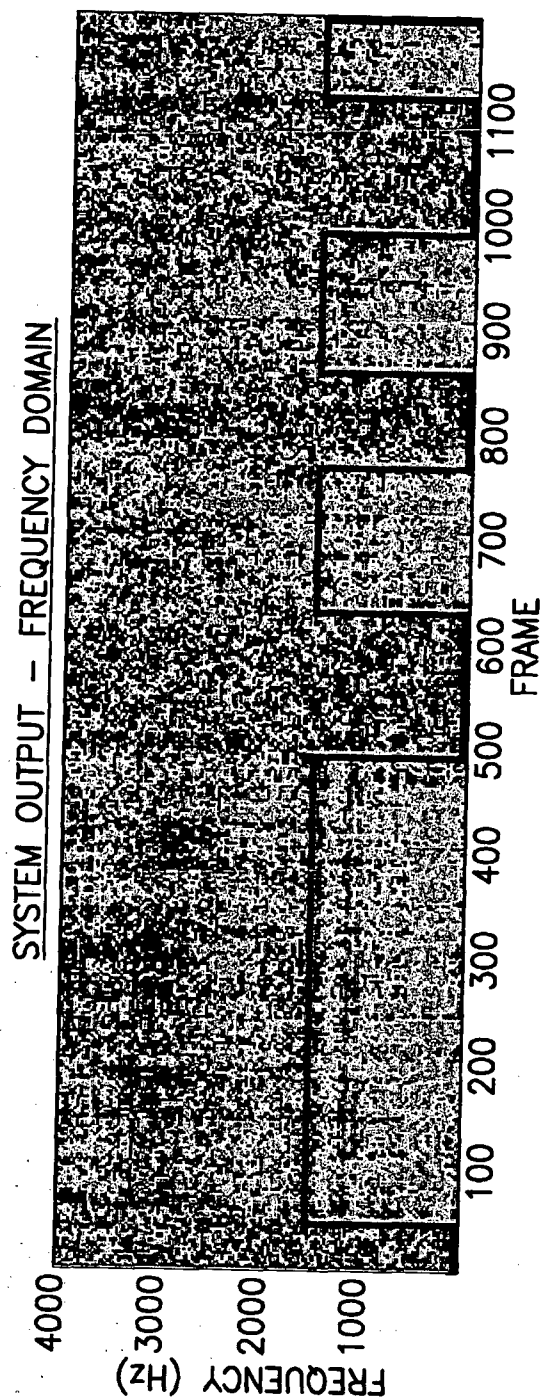


FIG.13B

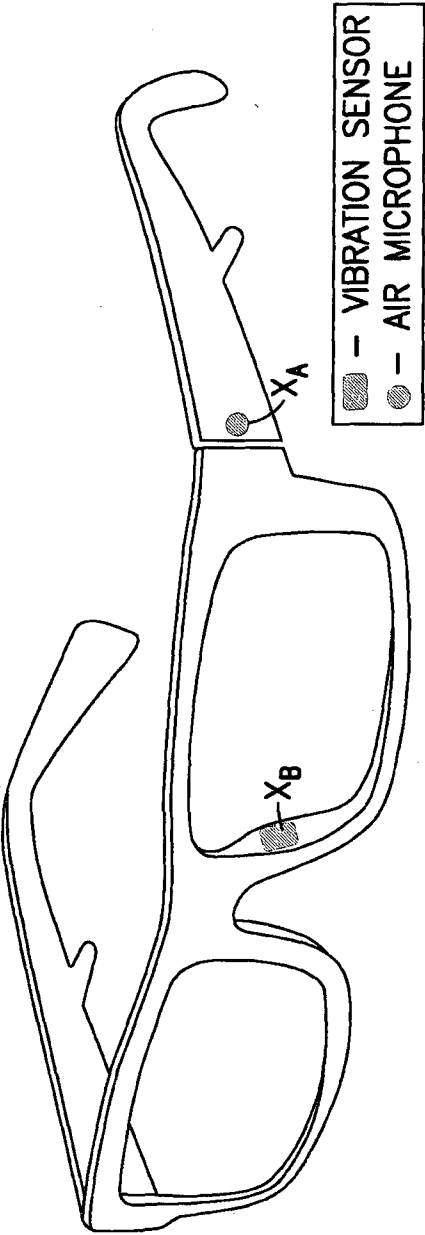


FIG. 14A

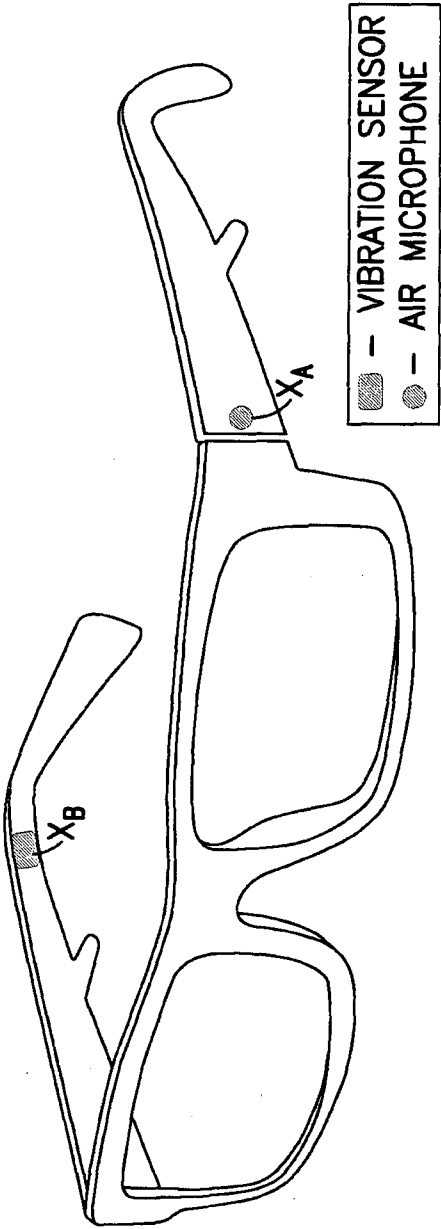


FIG. 14B

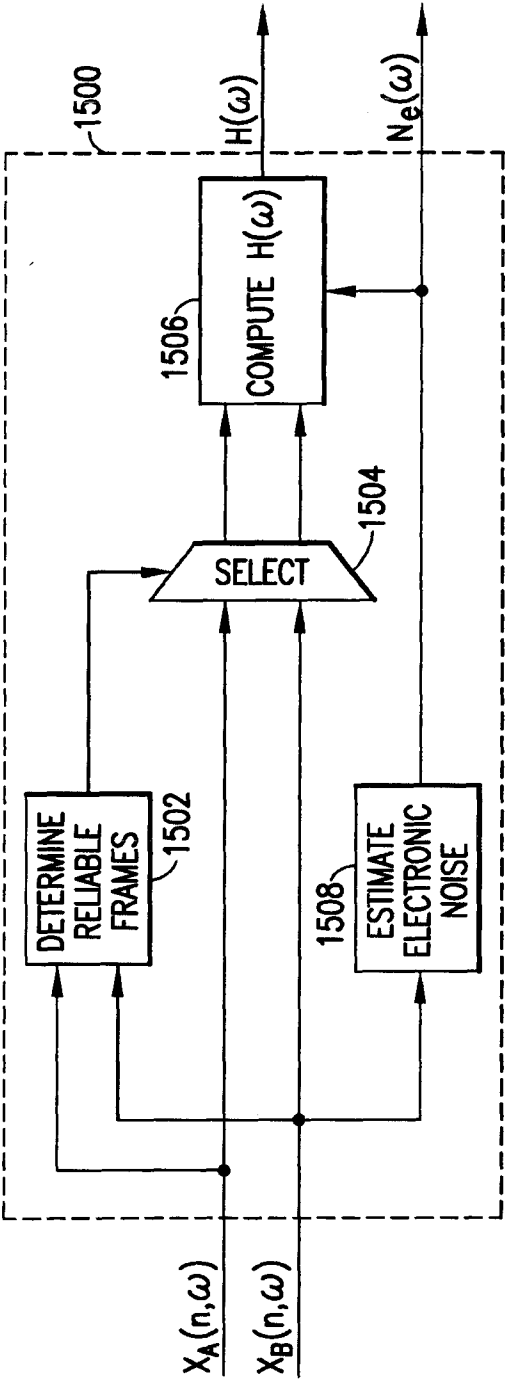


FIG. 15A

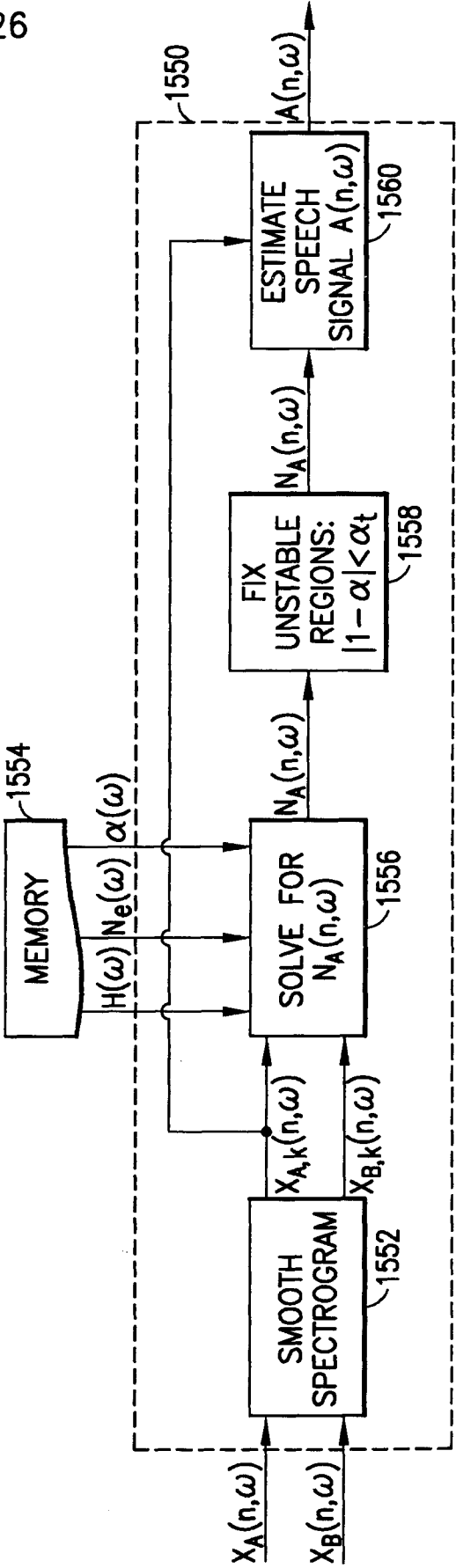


FIG. 15B

15/26

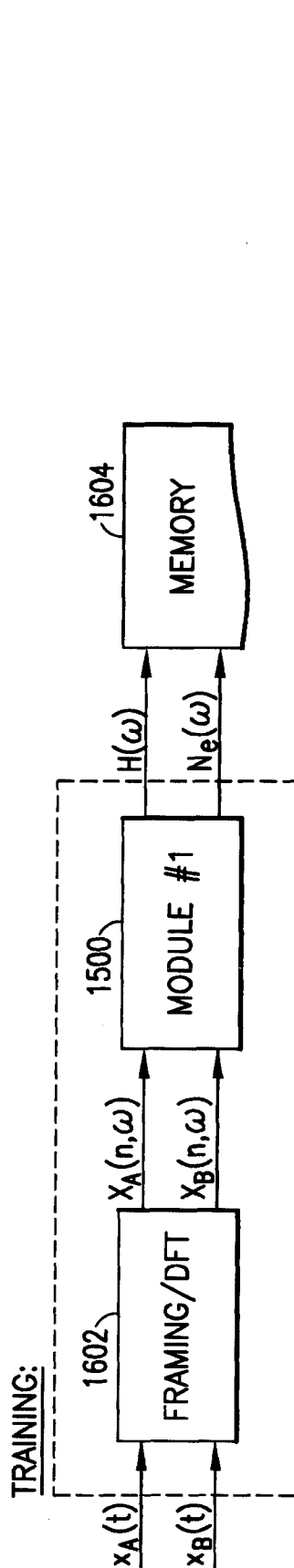


FIG. 16A

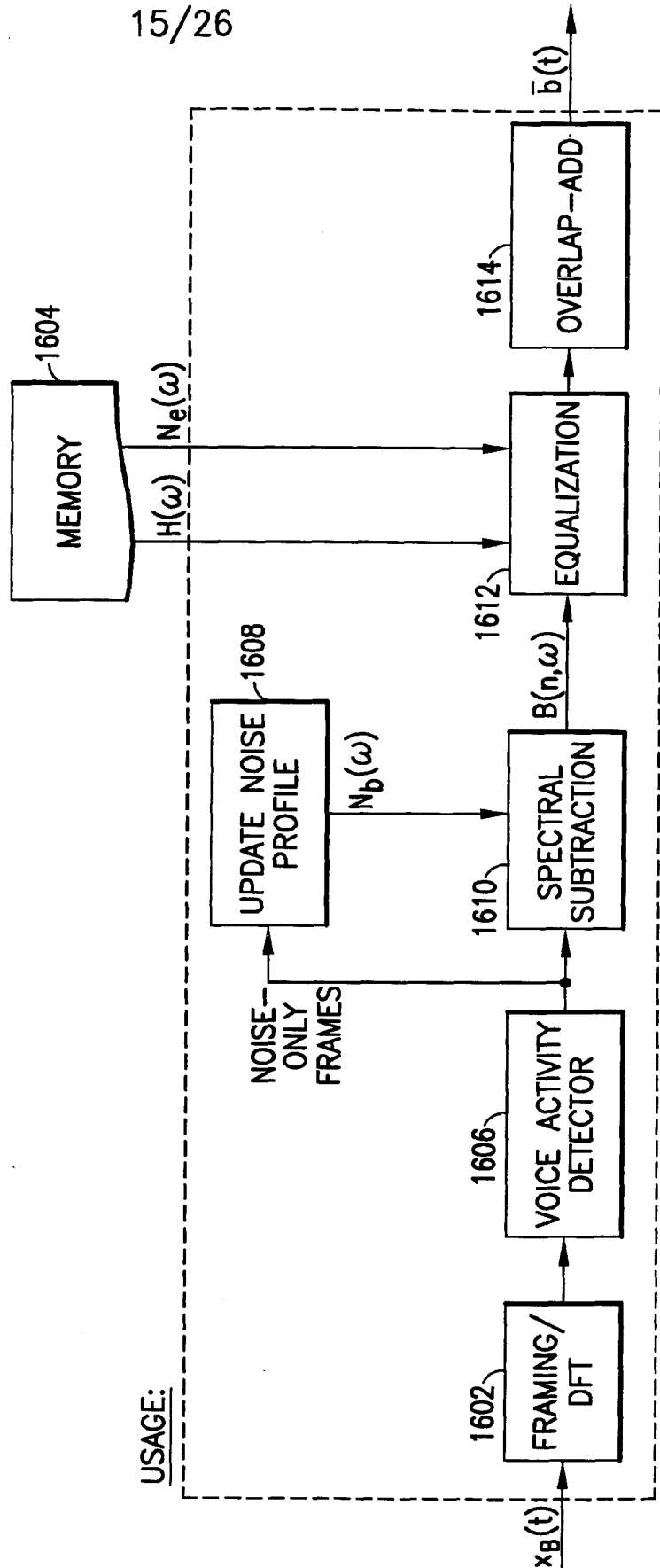


FIG. 16B

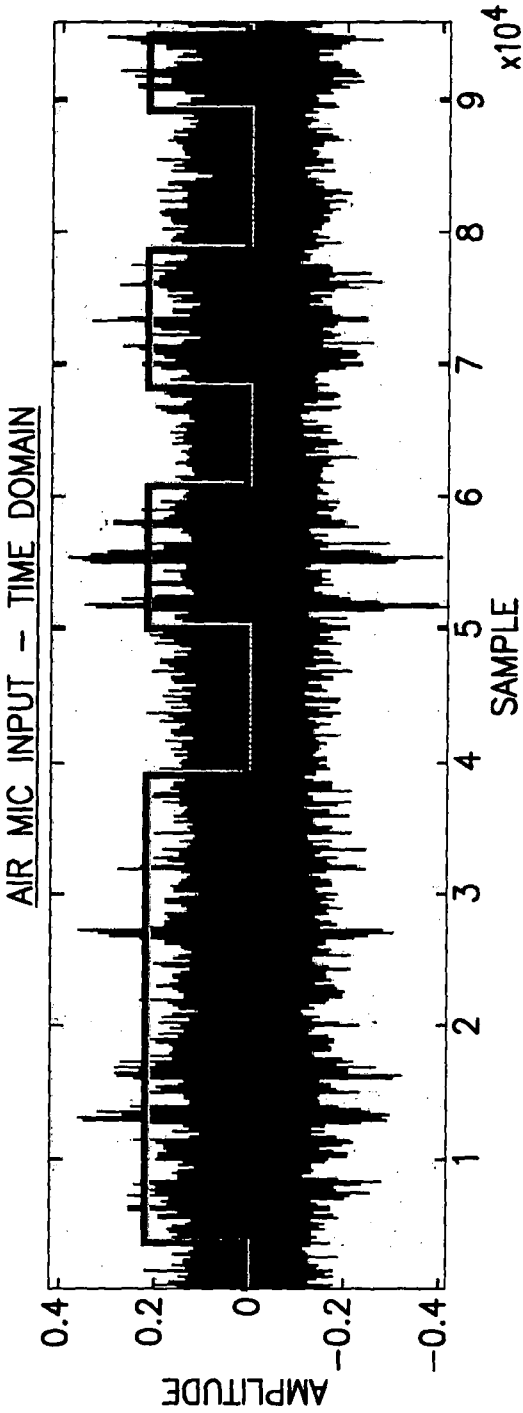


FIG. 17A

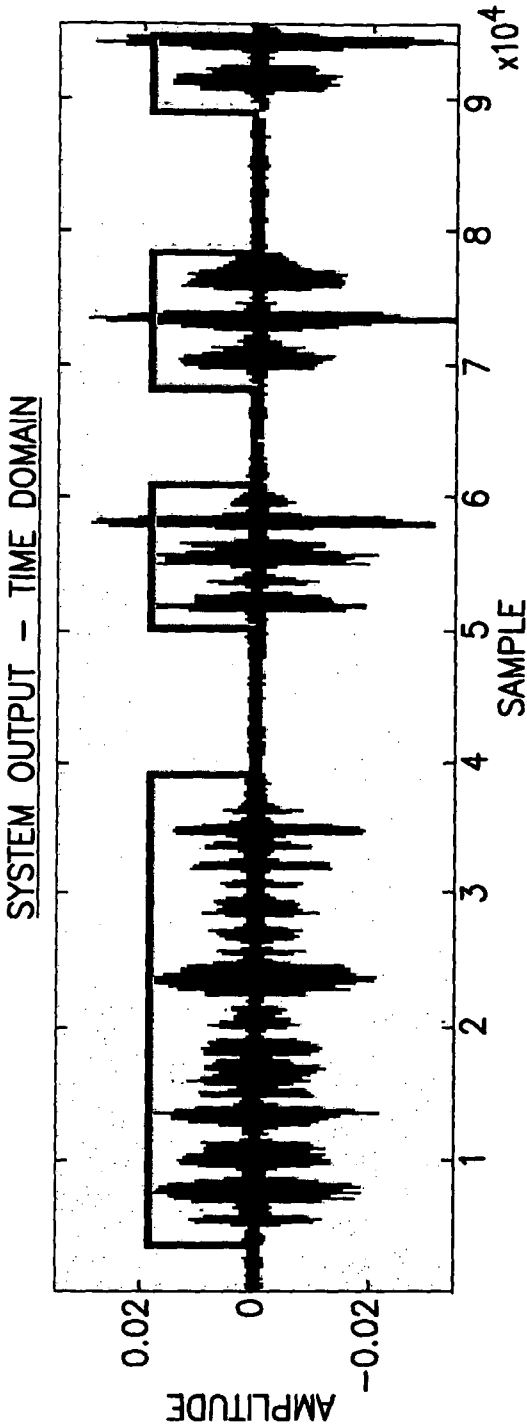


FIG. 17B

17/26

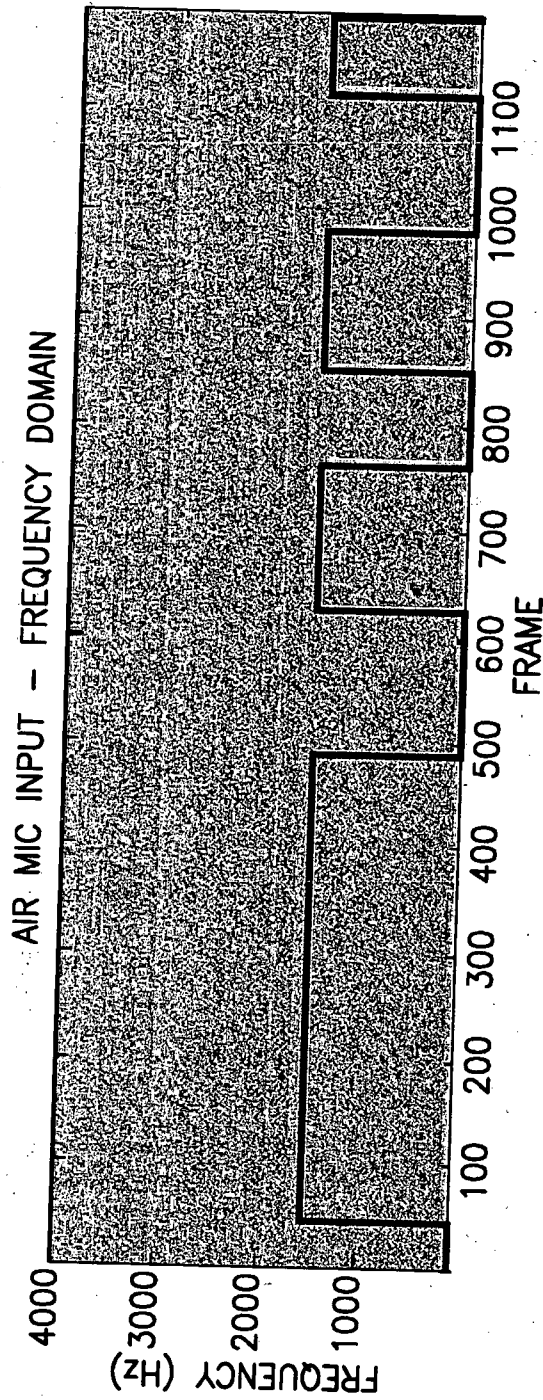


FIG.18A

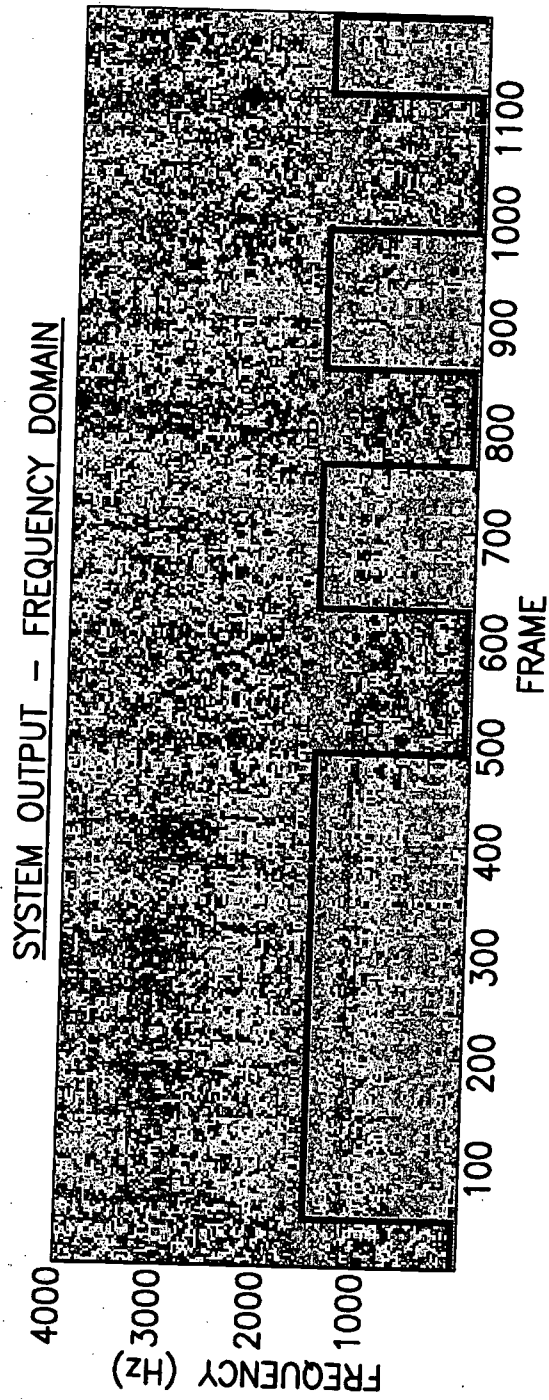


FIG.18B

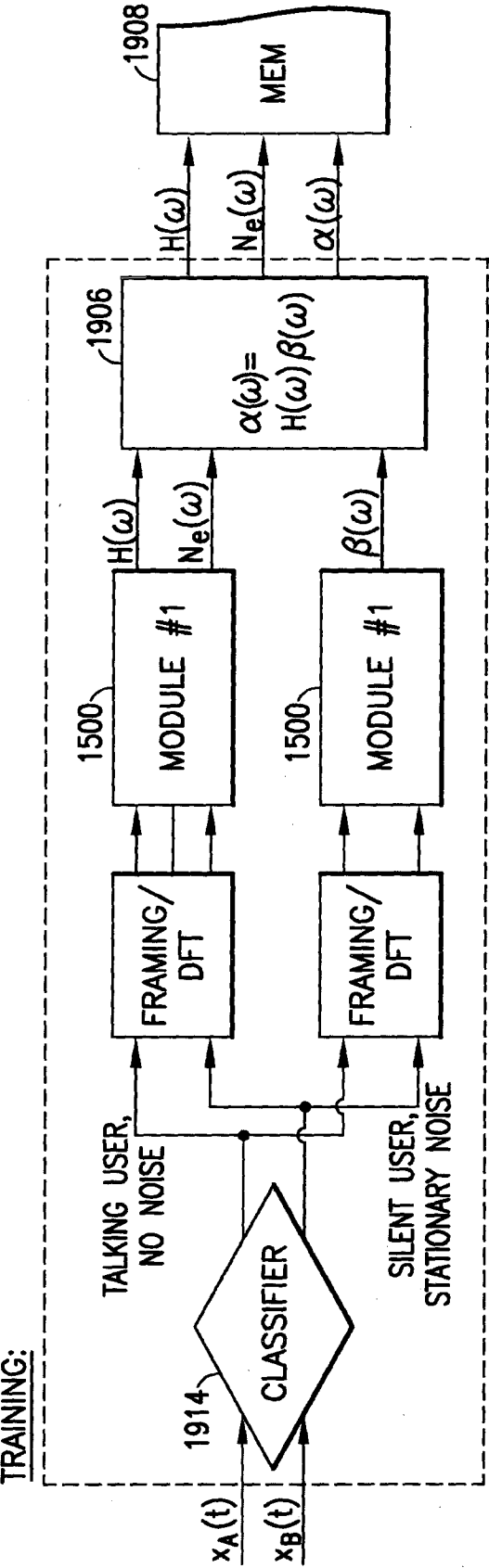


FIG. 19A

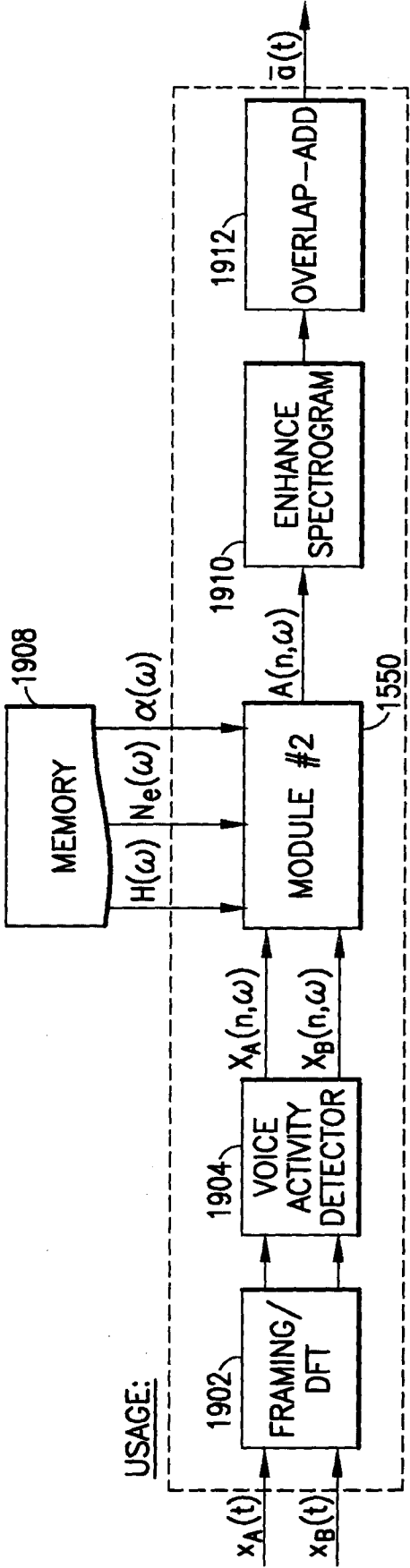


FIG. 19B

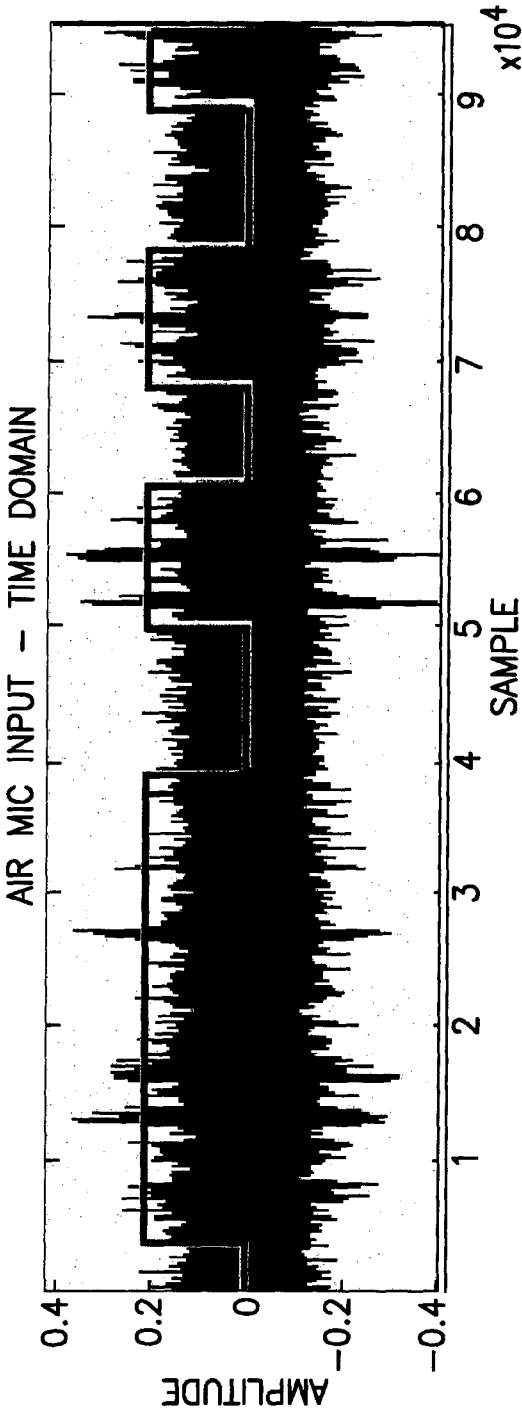


FIG. 20A

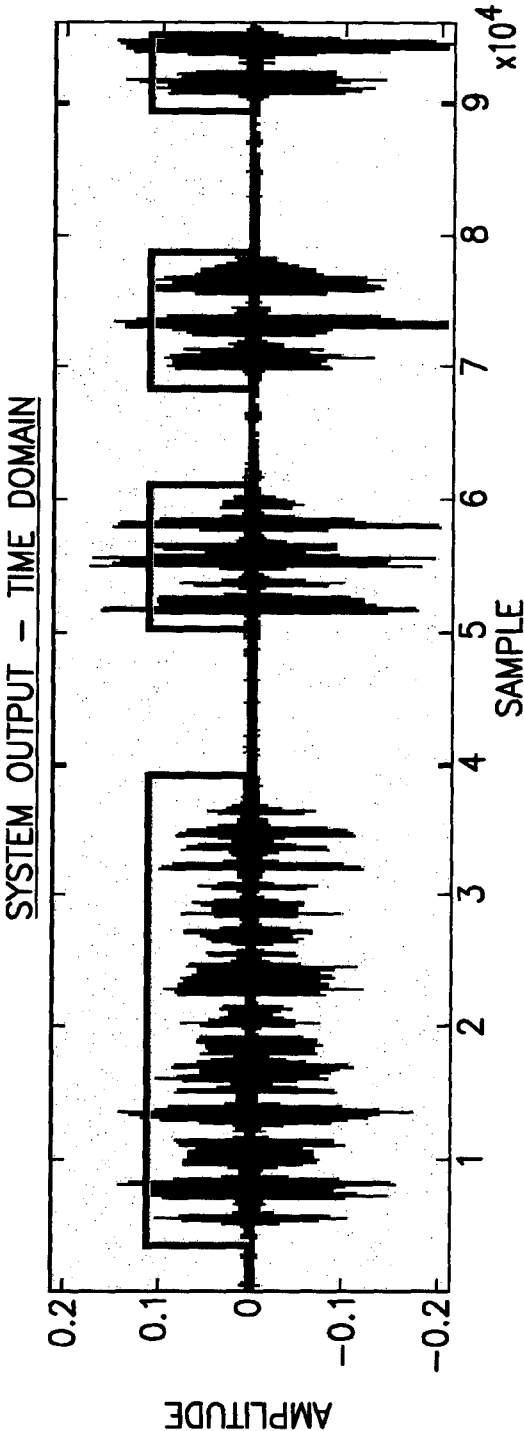


FIG. 20B

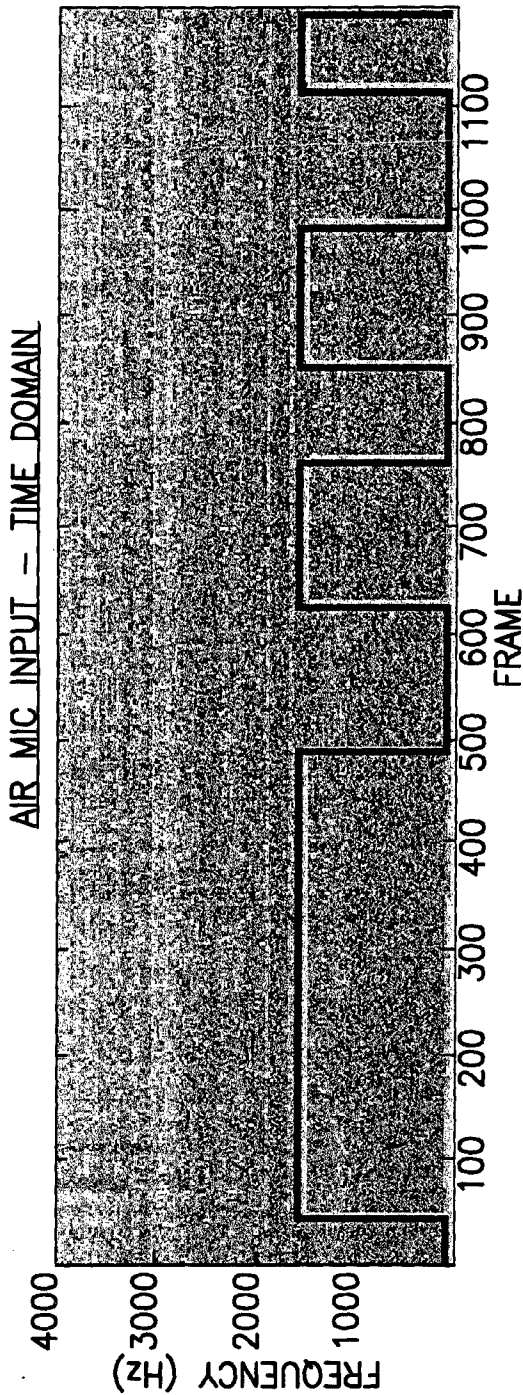


FIG.21A

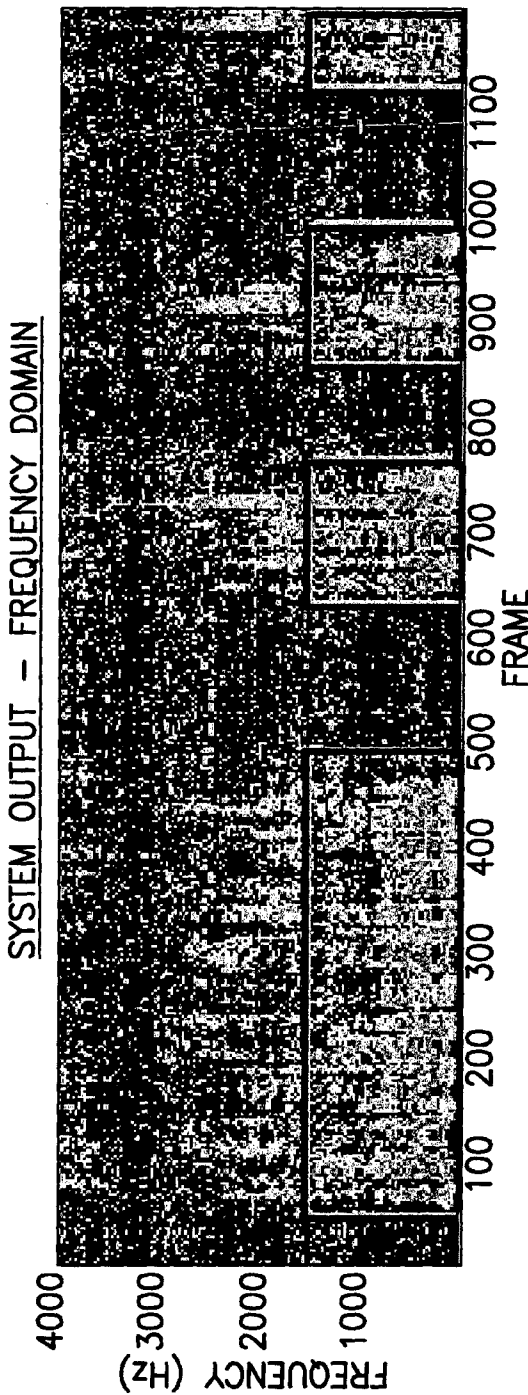


FIG.21B

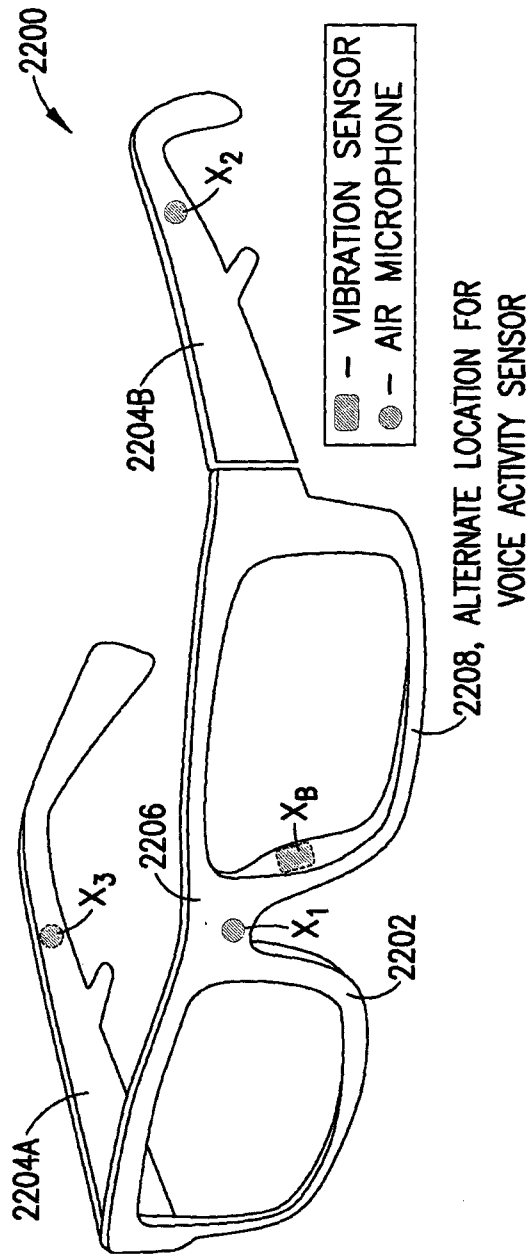


FIG. 22A

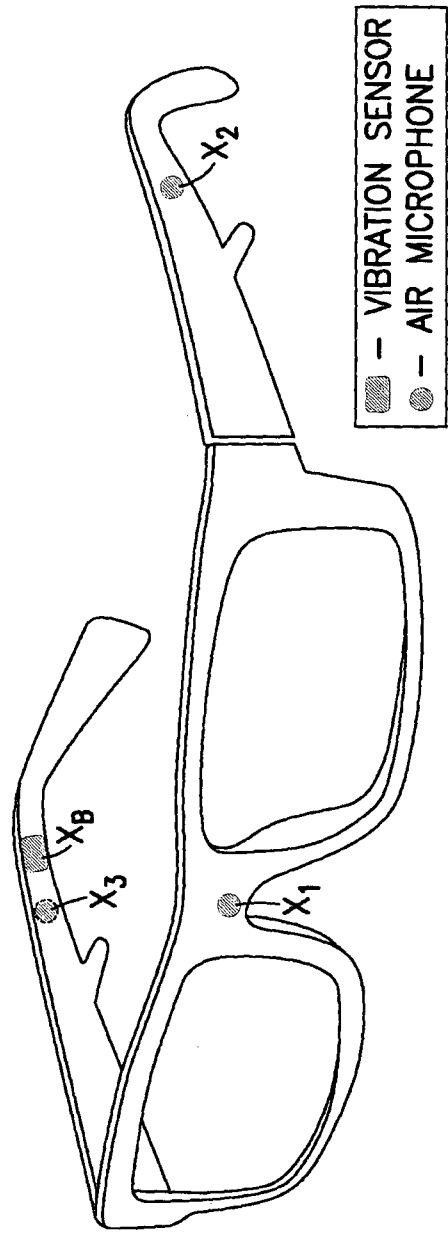


FIG. 22B

22/26

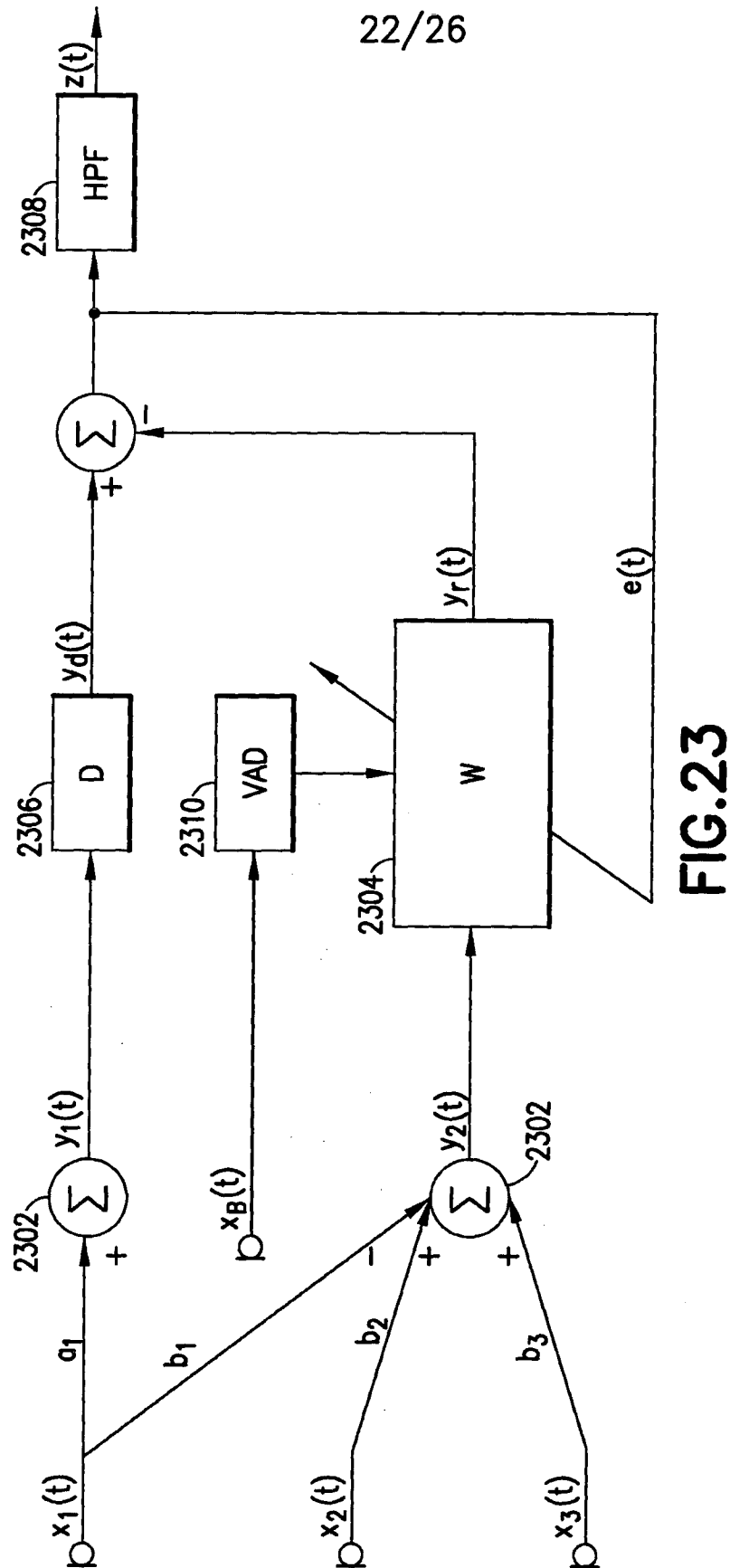


FIG. 23

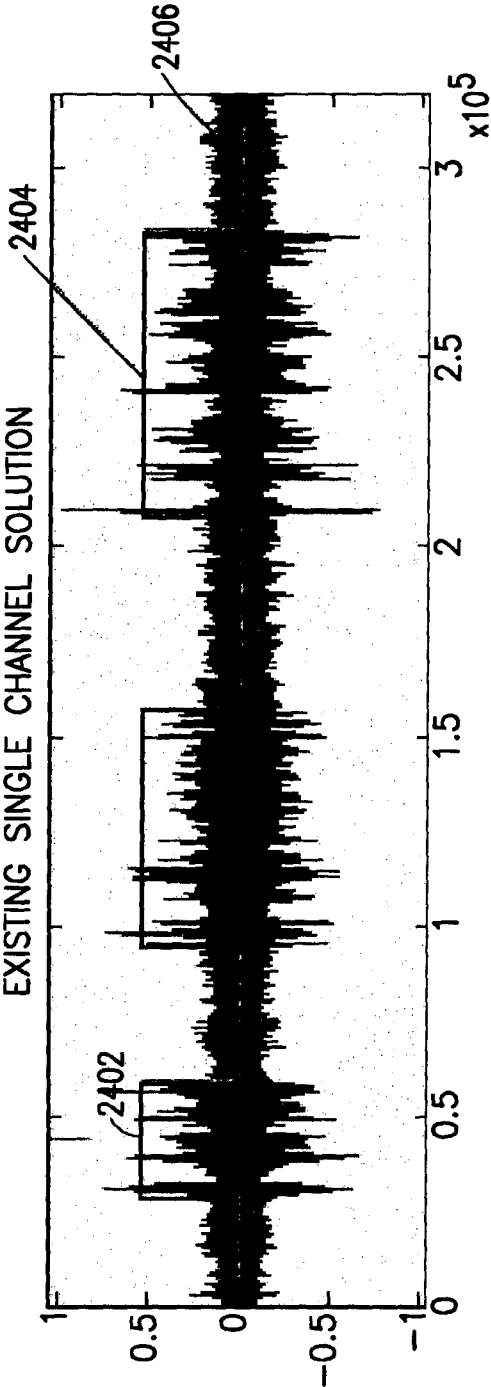


FIG. 24A

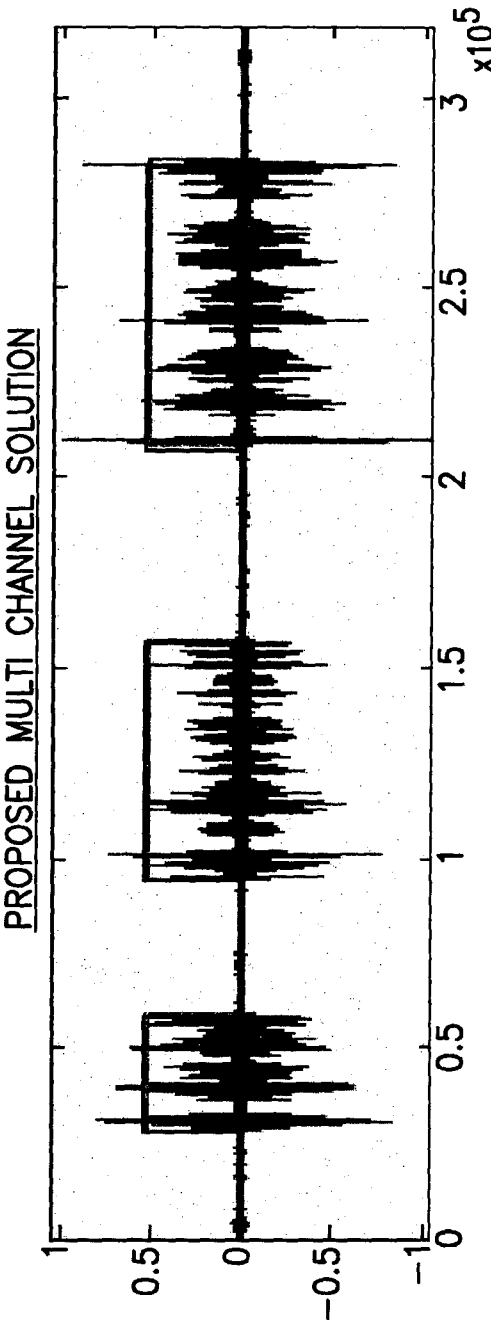


FIG. 24B

24/26

EXISTING SINGLE CHANNEL SOLUTION

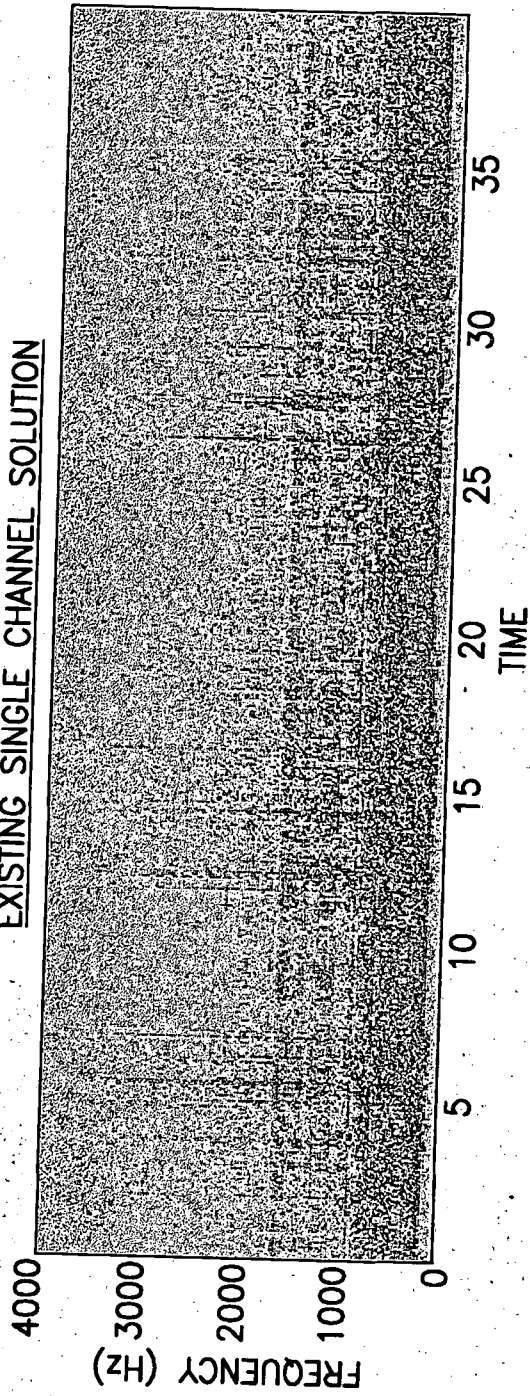


FIG. 25A

PROPOSED MULTI CHANNEL SOLUTION

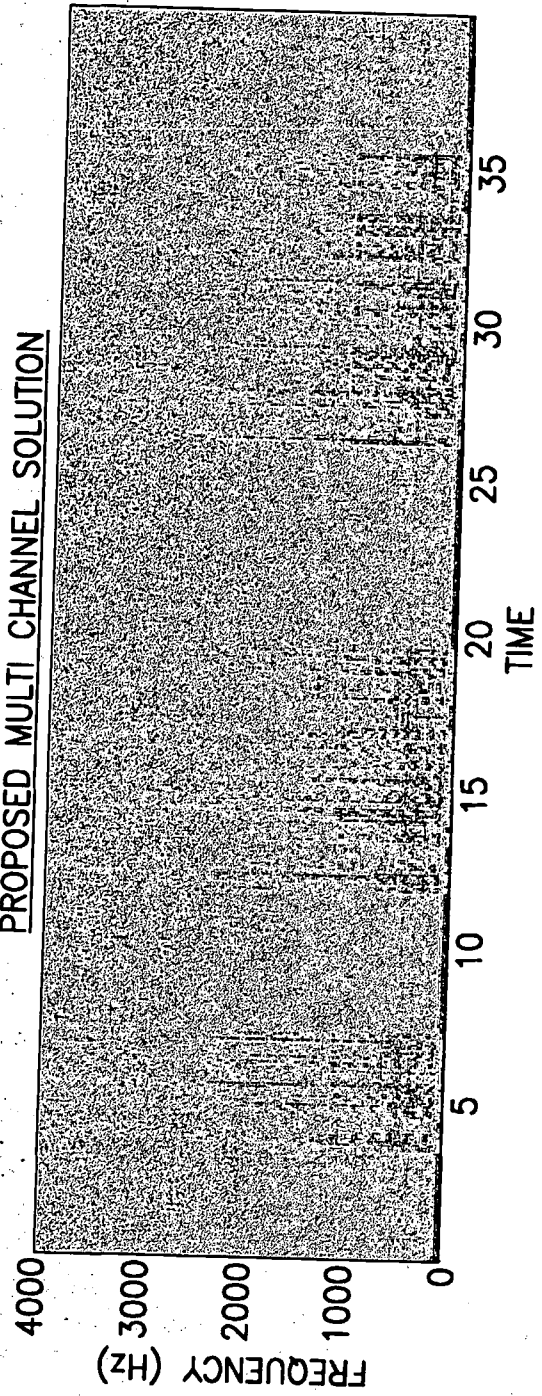


FIG. 25B

25/26

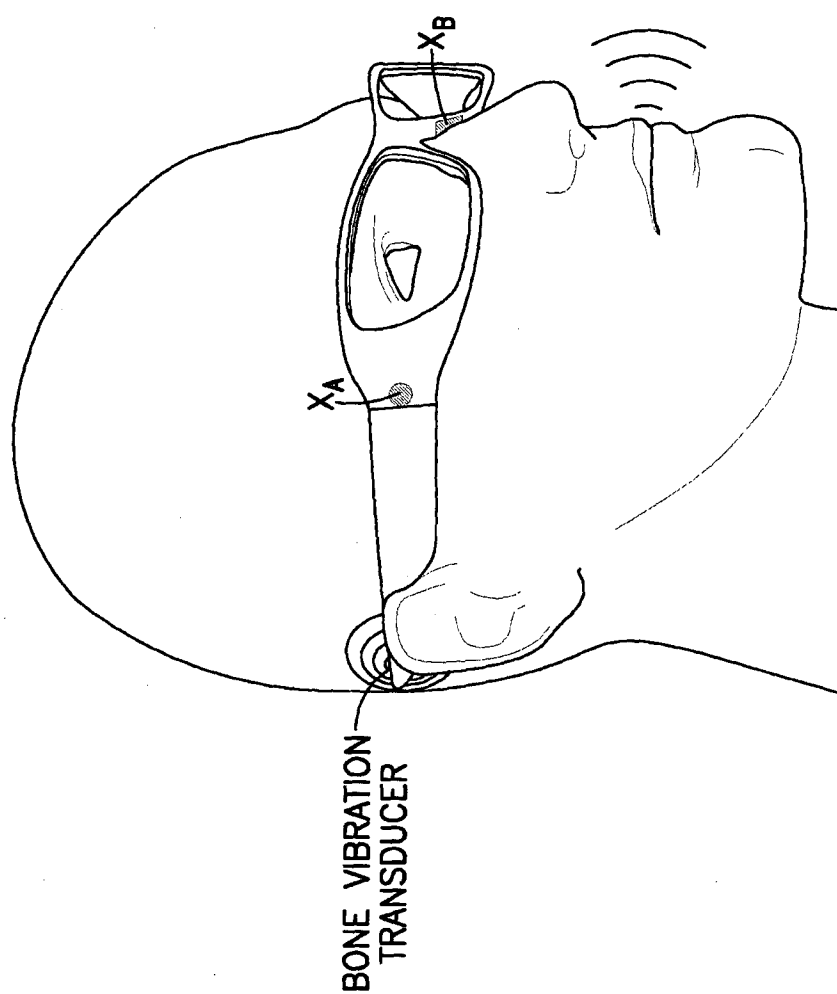
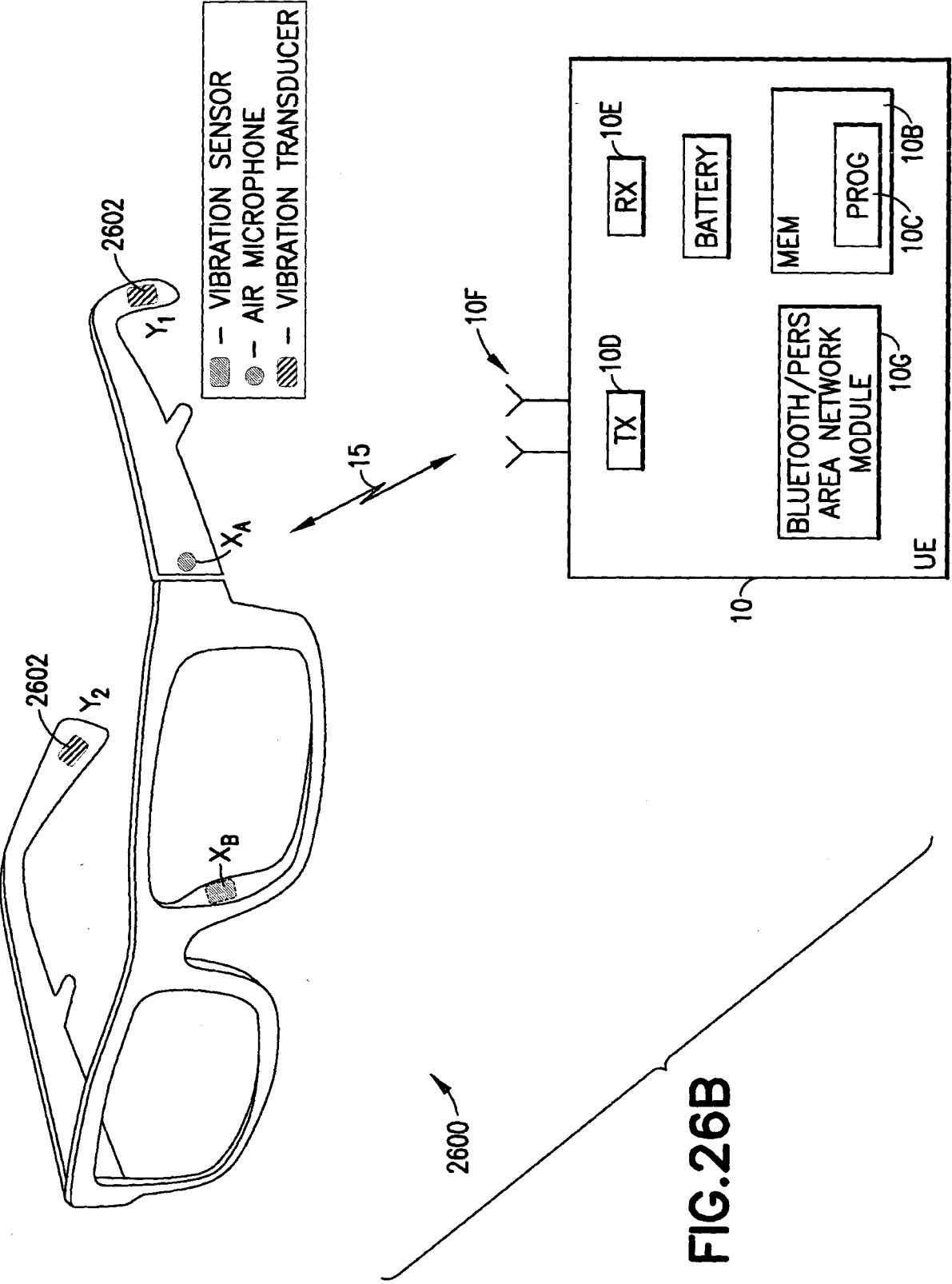


FIG. 26A



INTERNATIONAL SEARCH REPORT

International application No.

PCT/FI2013/050766

A. CLASSIFICATION OF SUBJECT MATTER

See extra sheet

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC: G10L, H04R, H04S, H04M, G02C

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
FI, SE, NO, DK

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, WPI, BIOSIS, COMPDX, EMBASE, INSPEC, MEDLINE, NPL, TDB, XP3GPP, XPAIP, XPESP, XPESP2, XPETSI, XPI3E, XPIEE, XPIETF, XPIOP, XPIPCOM, XPJPEG, XPMISC, XPOAC, XPRD, XPTK

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SUBRAMANYA, A. et al., 'Multisensory processing for speech enhancement and magnitude-normalized spectra for speech modeling'. Speech Communication, March 2008, Vol. 50, Issue 3, pp. 228 - 243, ISSN 0167-6393, DOI: 10.1016/j.specom.2007.09.002, XP 022436798 abstract; section 1, paragraphs 3 - 4, 6; section 2, paragraph 1, figure 1; section 4.1, paragraphs 1 - 2, equation (3); section 4.2; section 5, paragraphs 1 - 2, 5 - 6; section 7	1, 3 - 6, 9 - 10, 12 - 15, 18 - 19
Y	section 1, paragraph 6	7 - 8, 16 - 17
X	GB 2461315 A (WOLFSON MICROELECTRONICS PLC [GB]) 30 December 2009 (30.12.2009) abstract; figures 3, 8 - 10; page 4, line 21 - page 5, line 2; page 7, line 31 - page 8, line 3; page 12, line 16 - page 14, line 10; page 15, lines 12 - 31	1, 3 - 4, 7 - 10, 12 - 13, 16 - 19

☒ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance
 "E" earlier application or patent but published on or after the international filing date
 "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
 "O" document referring to an oral disclosure, use, exhibition or other means
 "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

01 November 2013 (01.11.2013)

Date of mailing of the international search report

05 November 2013 (05.11.2013)

Name and mailing address of the ISA/FI
 National Board of Patents and Registration of Finland
 P.O. Box 1160, FI-00101 HELSINKI, Finland

Facsimile No. +358 9 6939 5328

Authorized officer

Antti Salmela

Telephone No. +358 9 6939 500

INTERNATIONAL SEARCH REPORT

International application No.

PCT/FI2013/050766

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

See extra sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☒ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/FI2013/050766

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2011161487 A1 (NOKIA CORPORATION [FI] et al.) 29 December 2011 (29.12.2011) abstract; page 12, line 30 - page 13, line 26; page 14, lines 8 - 19; page 19, line 27 - page 20, line 23; figures 8 - 9	1, 3 - 4, 9 - 10, 12 - 13, 18 - 25
Y	ibid.	7 - 8, 16 - 17
X	US 2005027515 A1 (HUANG XUEDONG D [US] et al.) 03 February 2005 (03.02.2005) abstract; figures 3 - 4, 7 - 8, 10; paragraphs [0040] - [0047], [0052] - [0057], [0060]	1, 4, 9 - 10, 13, 18 - 19
Y	ibid.	2 - 3, 7 - 8, 11 - 12, 16 - 17
X	EP 1083769 A1 (YUGEN KAISHA GM & M [JP]) 14 March 2001 (14.03.2001) paragraphs [0015] - [0030], [0081], [0086], [0099], [0121], [0129], [0131], [0155] - [0156], [0164]; figures 1 - 2	1 - 2, 4, 9 - 11, 13, 18 - 19
Y	ibid.	3, 7 - 8, 12, 16 - 17
X	WO 2009101622 A2 (BONE TONE COMMUNICATIONS LTD [GB] et al.) 20 August 2009 (20.08.2009) paragraphs [0047] - [0049], [0051], [00125] - [00143]; figures 2A, 2B	1, 3 - 4, 9 - 10, 12 - 13, 18 - 19
Y	ibid.	2, 7 - 8, 11, 16 - 17
X	US 7346504 B2 (LIU ZICHENG [US] et al.) 18 March 2008 (18.03.2008) cited in the application, abstract; figures 3, 5 - 6; column 5, lines 13 - 56; column 6, lines 24 - 55; column 7, line 49 - column 8, line 21; column 9, line 30 - column 12, line 3	1, 3 - 6, 9 - 10, 12 - 15, 18 - 19
X	US 2011288860 A1 (SCHEVCIW ANDRE GUSTAVO PUCCI [US] et al.) 24 November 2011 (24.11.2011) abstract; figures 1A, 2 - 3, 8, 10, 13B, 14D, 15 - 16, 20 - 21; paragraphs [0062], [0071] - [0077], [0082] - [0083], [0100], [0108] - [0109], [0112], [0117] - [0118], [0130] - [0131], [0137] - [0140], [0154] - [0157], [0174]	19 - 25
Y	ibid.	3, 7 - 8, 12, 16 - 17
X	WO 2004016037 A1 (NANYANG TECHNOLOGICAL UNIVERSITY [SG] et al.) 19 February 2004 (19.02.2004) abstract; page 9, lines 4 - 7; page 13, lines 4 - 7; page 14, lines 1 - 21; page 15, lines 21 - 28; page 17, lines 13 - 27; page 23, line 31 - page 24, line 13; page 27, lines 7 - 27; page 28, lines 4 - 9; page 32, line 31 - page 33, line 10; figures 1 - 2	19 - 25
Y	ibid.	3, 7 - 8, 12, 16 - 17
Y	US 2011293109 A1 (NYSTRÖM MARTIN [SE] et al.) 01 December 2011 (01.12.2011) abstract; figures 3 - 4; paragraphs [0002], [0025] - [0031]	2, 11

INTERNATIONAL SEARCH REPORT

International application No.

PCT/FI2013/050766

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	JP 2011209548 A (NIPPON LOGICS KK [JP] et al.) 20 October 2011 (20.10.2011) & PAJ abstract; JPO-INPIT machine translation, paragraphs [0022] - [0024], [0027] - [0036], [0041] - [0046], [0051]; figure 1	2, 11
X	US 2004076301 A1 (ALGAZI V RALPH [US] et al.) 22 April 2004 (22.04.2004) abstract; figure 1; paragraphs [0053] - [0057], [0090] - [0091], [0141] - [0142]	20 - 25
X	US 7684571 B2 (GROSVENOR DAVID ARTHUR [GB] et al.) 23 March 2010 (23.03.2010) cited in the application, figures 2 - 3, 10; column 3, lines 29 - 41; column 4, lines 7 - 51; column 6, lines 33 - 60; column 7, line 66 - column 8, line 49	20 - 25
X	US 4819270 A (LOMBARDO LEONARD [US]) 04 April 1989 (04.04.1989) abstract; figures 1 - 2, 4 - 7; column 6, line 10 - column 7, line 2; column 7, line 34 - column 8, line 11; column 8, line 64 - column 9, line 17	20 - 25

INTERNATIONAL SEARCH REPORT

International application No.
PCT/FI2013/050766

CLASSIFICATION OF SUBJECT MATTER

Int.Cl.

H04R 1/46 (2006.01)

H04R 1/08 (2006.01)

H04R 3/00 (2006.01)

H04R 5/027 (2006.01)

H04M 1/05 (2006.01)

G10L 21/02 (2013.01)

G10L 25/93 (2013.01)

G10L 25/84 (2013.01)

H04M 1/60 (2006.01)

G02C 11/00 (2006.01)

The application contains two different inventions:

Invention I: Claims 1 – 19 directed to

- a platform configured to be removably placed symmetrically on or about a user's head
- at least a first transducer configured to capture vibration of the user's skull or facial movement generated by the user's voice activity
- processing the captured vibration or facial movement as voiced and/or unvoiced frames.

Invention II: Claims 20 – 25 directed to

- a platform configured to be removably placed symmetrically on or about a user's head
- resolving a spatial signal from the output of at least two transducers disposed about opposite sides of the platform.

Symmetrical platforms that can be placed removably on or about a user's head are part of common knowledge, e.g., eyeglasses, sunglasses, helmets, headbands, headphones, hats. It is also well known to mount microphones on such platforms (see, e.g., any of documents D1 – D6, D8 – D9, D12 – D14). Therefore, the first feature of inventions I and II defines no contribution over the prior art. The remaining special technical features of invention I address the problem of detecting the user's speaking activity based on the user's skull vibrations or facial movements. The special technical features of invention II, on the other hand, address the problem of recording spatial sound with transducers disposed on the symmetrical platform. Since these special technical features of inventions I and II are different and address different technical problems, no technical relationship as required by PCT Rule 13.2 exists between inventions I and II. Accordingly, the application does not relate to one invention or to a single general inventive concept.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/FI2013/050766

Patent document cited in search report	Publication date	Patent family members(s)	Publication date
GB 2461315 A	30/12/2009	CN 102099852 A	15/06/2011
		GB 0811863 D0	30/07/2008
		GB 2461315 B	14/09/2011
		US 2011130176 A1	02/06/2011
		WO 2009156756 A2	30/12/2009
.....			
WO 2011161487 A1	29/12/2011	US 2013114821 A1	09/05/2013
.....			
US 2005027515 A1	03/02/2005	AT 471554 T	15/07/2010
		AU 2004203357 A1	17/02/2005
		AU 2004203357 B2	28/01/2010
		BR PI0403027 A	31/05/2005
		CA 2473195 A1	29/01/2005
		CN 1591568 A	09/03/2005
		CN 100573664 C	23/12/2009
		DE 602004027687 D1	29/07/2010
		EP 1503368 A1	02/02/2005
		EP 1503368 B1	16/06/2010
		HK 1073010 A1	02/02/2011
		JP 2005049876 A	24/02/2005
		JP 4703142 B2	15/06/2011
		KR 20050013969 A	05/02/2005
		MX PA04007313 A	26/07/2005
		MY 138807 A	31/07/2009
		RU 2004123352 A	27/01/2006
		RU 2363994 C2	10/08/2009
		US 7383181 B2	03/06/2008
		US 2005033571 A1	10/02/2005
		ZA 200405491 A	26/07/2006
.....			
EP 1083769 A1	14/03/2001	AT 471039 T	15/06/2010
		AU 2571900 A	04/09/2000
		CA 2328953 A1	24/08/2000
		DE 60044521 D1	22/07/2010
		EP 1083769 B1	09/06/2010
		JP 2000308198 A	02/11/2000
		JP 3670180 B2	13/07/2005
		JP 4439740 B2	24/03/2010
		US 7676372 B1	09/03/2010
		WO 0049834 A1	24/08/2000
.....			
WO 2009101622 A2	20/08/2009	EP 2250822 A2	17/11/2010
		JP 2011512745 A	21/04/2011

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/FI2013/050766

Patent document cited in search report	Publication date	Patent family members(s)	Publication date
		US 2011301729 A1	08/12/2011
US 7346504 B2	18/03/2008	AT 476734 T	15/08/2010
		AU 2006262706 A1	04/01/2007
		AU 2006262706 B2	25/11/2010
		BR PI0611649 A2	28/09/2010
		CA 2607981 A1	04/01/2007
		CN 101199006 A	11/06/2008
		CN 101199006 B	24/08/2011
		DE 602006015954 D1	16/09/2010
		EP 1891627 A2	27/02/2008
		EP 1891627 B1	04/08/2010
		JP 2008544328 A	04/12/2008
		JP 4975025 B2	11/07/2012
		KR 20080018163 A	27/02/2008
		MX 2007014562 A	16/01/2008
		NO 20075732 A	17/03/2008
		RU 2007147463 A	27/06/2009
		RU 2407074 C2	20/12/2010
		US 2006287852 A1	21/12/2006
		WO 2007001768 A2	04/01/2007
US 2011288860 A1	24/11/2011	CN 102893331 A	23/01/2013
		EP 2572353 A1	27/03/2013
		JP 2013531419 A	01/08/2013
		KR 20130042495 A	26/04/2013
		WO 2011146903 A1	24/11/2011
WO 2004016037 A1	19/02/2004	AU 2002329160 A1	25/02/2004
US 2011293109 A1	01/12/2011	CN 103038824 A	10/04/2013
		EP 2577988 A1	10/04/2013
		WO 2011148271 A1	01/12/2011
JP 2011209548 A	20/10/2011	None	
US 2004076301 A1	22/04/2004	AU 2003273363 A1	13/05/2004
		CA 2502585 A1	06/05/2004
		EP 1554910 A1	20/07/2005
		JP 2006503526 A	26/01/2006

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/FI2013/050766

Patent document cited in search report	Publication date	Patent family members(s)	Publication date
		KR 20050056241 A	14/06/2005
		MX PA05004091 A	08/06/2005
		US 7333622 B2	19/02/2008
		US 2007009120 A1	11/01/2007
		US 2008056517 A1	06/03/2008
		WO 2004039123 A1	06/05/2004
.....			
US 7684571 B2	23/03/2010	GB 0414364 D0	28/07/2004
		GB 2415584 A	28/12/2005
		GB 2415584 B	26/09/2007
		US 2005286728 A1	29/12/2005
.....			
US 4819270 A	04/04/1989	None	
.....			