(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
2 February 2012 (02.02.2012)

PCT

(10) International Publication Number
WO 2012/015569 A1

(51) International Patent Classification:
G10L 21/02 (2006.01) *G10K 11/16* (2006.01)

(21) International Application Number:
PCT/US2011/043115

(22) International Filing Date:
7 July 2011 (07.07.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
12/844,954 28 July 2010 (28.07.2010) US

(71) Applicant *(for all designated States except US)*: **MOTOROLA SOLUTIONS, INC.** [US/US]; 1303 East Algonquin Road, Schaumburg, Illinois 60196 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **KALE, Kaustubh** [IN/US]; 8995 Chambers Street, Tamarac, Florida 33321 (US). **WANG, Yong** [CA/US]; 26 Idaho Court, Coram, New York 11727 (US).

(74) Agents: **GIANNETTA, Michael J.** et al.; 1303 East Algonquin Road, Schaumburg, Illinois 60196 (US).

(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, 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, KM, 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, PE, PG, PH, PL, PT, RO, RS, RU, 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, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, 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, ML, MR, NE, SN, TD, TG).

Published:

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

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

(54) **Title**: FORMANT AIDED NOISE CANCELLATION USING MULTIPLE MICROPHONES

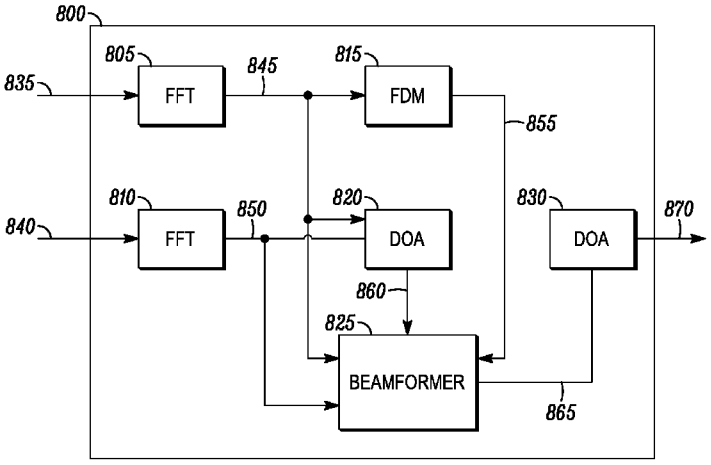


FIG. 8

(57) **Abstract**: A noise cancellation device includes a plurality of first computation modules, a formant detection module, a direction of arrival module and a beamformer. The plurality of first computation modules receives raw audio data and generates a respective transformed signal as a function of formants. A first transformed signal relates to speech data and a second transformed signal relates to noise data. The formant detection module receives the first transformed signal and generates a frequency range data signal. The direction of arrival module receives the first and second transformed signals, determines a cross-correlation between the first and second transformed signals, and generates a spatial orientation data signal. The beamformer receives the first and second transformed signals, the frequency range data signal, and the spatial orientation data signal and generates modification data at selected formant ranges to eliminate a maximum amount of the noise data.

WO 2012/015569 A1

Formant Aided Noise Cancellation Using Multiple Microphones

Background

[0001] An electronic device may include an audio input device such as a microphone to receive audio inputs from a user. The microphone is configured to receive any sound and convert the raw audio data into an audio signal for transmission. However, during the course of the microphone receiving the sound, ambient noise is also captured and incorporated into the audio signal.

[0002] Conventional technologies have created ways of reducing the ambient noise captured by microphones. For example, a single microphone noise suppressor attempts to capture ambient noise during silence periods and use this estimate to cancel noise. In another example, sophisticated algorithms attempt to reduce the noise floor during speech or are able to reduce non-stationary noise as it moves around. In multiple microphone noise cancellation systems, a beam is directed in space toward the desired talker and attempts to cancel maximum noise from all other directions. However, in all conventional approaches, the attempt to capture clean speech relates to spatial distribution.

Summary of the Invention

[0003] The exemplary embodiments describe a noise cancellation device comprising a plurality of first computation modules, a formant detection module, a direction of arrival module and a beamformer. The plurality of first computation modules receives raw audio data and generates a respective transformed signal as a function of formants. A first transformed signal relates

to speech data and a second transformed signal relates to noise data. The formant detection module receives the first transformed signal and generates a frequency range data signal. The direction of arrival module receives the first and second transformed signals, determines a cross-correlation between the first and second transformed signals, and generates a spatial orientation data signal. The beamformer receives the first and second transformed signals, the frequency range data signal, and the spatial orientation data signal and generates modification data at selected formant ranges to eliminate a maximum amount of the noise data.

Description of the Drawings

[0004] Fig. 1a shows a first formant for a first sound.

[0005] Fig. 1b shows a second formant for a second sound.

[0006] Fig. 2a shows a third formant for a third sound.

[0007] Fig. 2b shows a fourth formant for the third sound.

[0008] Fig. 3 shows a beam pattern for a microphone.

[0009] Fig. 4 shows a top view of a beam pattern for a multi-microphone noise cancellation system.

[0010] Fig. 5 shows a formant energy distribution of speech for a duration of time.

[0011] Fig. 6 shows a spectrogram of speech.

[0012] Fig. 7 shows beam patterns with two microphones at a set distance.

[0013] Fig. 8 shows a formant based noise cancellation device according to an exemplary embodiment.

[0014] Fig. 9 shows a method for a formant based noise cancellation according to an exemplary embodiment.

Detailed Description

[0015] The exemplary embodiments may be further understood with reference to the following description and the appended drawings, wherein like elements are referred to with the same reference numerals. The exemplary embodiments describe a device and method for noise cancellation using multiple microphones that is formant aided. Specifically, psychoacoustics is considered in reducing noise speech captured through a microphone. The microphones, the noise cancellation, the formants, the psychoacoustics, and a related method will be discussed in further detail below.

[0016] Those skilled in the art will understand that knowing the psychoacoustics of speech, the energy for a speech signal may be given by formants. Fig. 1a shows a first formant for a first sound. Specifically, Fig. 1a shows the formant for a typical "AH" sound. As shown, the energy distribution fluctuates throughout the sound. Fig. 1b shows a second formant for a second sound. Specifically, Fig. 1b shows the formant for a typical

"EE" sound. As shown, the energy distribution also fluctuates throughout the sound.

[0017] Furthermore, in view of the formants shown in Figs. 1a and 1b, the energy distribution changes drastically during conversational speech. For example, if there were noise with a frequency of 1.5kHz, the noise is more disruptive to the first formant of Fig. 1a (*i.e.*, "AH" sound) because the first formant has sufficient audible energy at 1.5kHz. In contrast, the second formant of Fig. 1b (*i.e.*, "EE" sound) is not affected by the noise at 1.5kHz because, perceptively, no sound is heard in the 1.5kHz range. Consequently, with noise energy at 1.5kHz, the "EE" sound is heard with almost no noise affect but the "AH" sound is more difficult to understand. This principle of noise energy at varying frequencies is incorporated in the formant based noise cancellation according to the exemplary embodiments.

[0018] Those skilled in the art will also understand that formant energies may differ from one speaker to another. Fig. 2a shows a third formant for a third sound (*i.e.*, "A" sound). Fig. 2b shows a fourth formant also for the third sound. It should be noted that Figs. 2a and 2b relating to different speakers is only exemplary. The formants of Figs. 2a and 2b may also represent an energy distribution from a different speaker for the same sound.

[0019] In view of the formants shown in Figs. 2a and 2b, the energy distribution differs from one speaker to another speaker although a common sound is being uttered. Again, using a noise with frequency of 1.5kHz, the noise

is more disruptive for the speaker in Fig. 2a while not as disruptive for the speaker in Fig. 2b. Consequently, with noise energy at 1.5kHz, the first sound coming from the first speaker is more difficult to understand while the first sound coming from the second speaker is more easily understood. This principle of noise energy at varying frequencies is also incorporated in the formant based noise cancellation according to the exemplary embodiments.

[0020] With conventional single or double microphone noise cancellation systems, speech is attempted to be captured as noise free as possible from a single direction by achieving predetermined spatial patterns. With multiple microphone noise cancellation systems, multiple directions may be used to capture the speech. Fig. 3 shows a beam pattern for a microphone. As illustrated in Fig. 3, the source of the speech may be directly in front of the microphone at 90 degrees. Fig. 4 shows a top view of a beam pattern for a multi-microphone noise cancellation system.

[0021] Despite spatial orientations of beams of microphones being capable of at least partially reducing noise, it does not account for the psychoacoustics fact that the spatial intensity direction and frequency intensity direction for noise is not always connected. For example, a first noise located at 45 degrees in front of a microphone may be the loudest but may have a maximum intensity at 1.5kHz. A second noise located at 135 degrees in front of a user might have a lower maximum intensity but may have more intensity than the first noise at a different frequency such as 700Hz. However, a

conventional beamformer will cancel the first noise and not the second noise. Thus, the first noise at 1.5kHz that does not cause much degradation gets cancelled whereas the noise at 700Hz that can cause degradation is not cancelled, resulting in a bad audio output signal. Therefore, canceling noise as a function of formant shaping and prioritizing cancellation of noise at frequencies that are more sensitive over noise at frequencies that are less sensitive to noise is desired, thereby leading to significantly improved audio performance. The exemplary embodiments further incorporate this aspect for the formant aided noise cancellation.

[0022] Fig. 5 shows a formant energy distribution of speech for a duration of time. The distribution illustrates the time domain speech signal of the speaker on the top graph with the corresponding frequency domain signal with formants highlighted on the bottom graph. If noise along the blotted lines 500 are cancelled, the audio quality of speech becomes superior over conventional noise cancellation methods that do not use psychoacoustics knowledge and merely attempts to cancel noise spatially.

[0023] The exemplary embodiments estimates formant position and/or maximum speech energy regions in real time using formant tracking algorithms such as Linear Predictive Coding (LPC), Hidden Markov Model (HMM), etc. The formant frequency range data generated is used at a beamforming algorithm that uses the dual microphone input to cancel noise in these frequency ranges.

[0024] Fig. 6 shows a spectrogram of speech for an interfering talker and pink noise coming from a single location in space. As illustrated, the intensity is different at different frequencies and changes with time. For example, between 0.2-0.3 seconds, the maximum intensity is around 500Hz while between 0.4-0.5 seconds, the intensity is around 500Hz as well as 2000Hz and 3000 Hz.

[0025] Fig. 7 shows beam patterns with two microphones at a set distance. Specifically, Fig. 7 illustrates beam patterns of beamformers. The pattern changes with distance between the at least two microphones. Furthermore, for the same direction, the pattern is different at various frequencies. For example, assuming the speaker is at 0 degrees in front of the microphone, speech is captured perfectly. However, if there is a 7000Hz noise at 75 degrees, the noise will be captured just as loudly as the speech.

[0026] Although there are other beamforming techniques that will, for example, attempt to place a null at 75 degrees to cancel the noise source or attempt to place a null at the speaker and use the rest of the signal as a noise estimate, these techniques succumb to the aforementioned problem in which the location is irrelevant when relating to noise capture. In contrast, the exemplary embodiments consider the location of the frequency of the speech's energy.

[0027] Fig. 8 shows a formant based noise cancellation device 800 according to an exemplary embodiment. The device 800 may be incorporated with any electronic device

that includes an audio receiving device such as a microphone. According to the exemplary embodiment of Fig. 8, the electronic device includes a multiple microphone system comprising two microphones. Furthermore, the exemplary embodiment is based on frames of 20ms of data. Thus, as will be described in further detail below, two frames of 20ms data will be used while 20ms of processed output is returned. It should be noted that the use of 20ms frames of data is only exemplary and the rate is configurable based on the acoustic needs of the platform. It should also be noted that the use of a two microphone system is only exemplary and a system including any number of microphones may be adapted using the exemplary embodiments. The device 800 may include a first Fast Fourier Transform Module (FFT) 805, a second FFT 810, a Formant Detection Module (FDM) 815, a Direction of Arrival module (DOA) 820, a beamformer 825, and an Inverse FFT (IFFT) 830.

[0028] The FFT 805 may receive a first microphone speech data 835 while the FFT 810 may receive a second microphone speech data 840. With reference to the exemplary rate of 20ms, speech samples from the first and second microphones in 20ms frames are computed by the FFTs 805, 810, respectively. According to the exemplary embodiments, the FFTs 805, 810 may compute a 128, 256, and/or 512 point FFT of a 8kHz signal, thereby breaking into 64, 128, and/or 256 frequency bins. Again, it should be noted that the computations of the FFTs 805, 810 is only exemplary and the computations may be changed as a function on the resolution desired and the platform capabilities to handle the FFTs' processing. For

example, if a 128 point FFT is selected, 64 frequency bins from 0-4000Hz are generated.

[0029] The FFT 805 generates a first speech FFT signal 845 which is received by the FDM 815. The FDM 815 may compute the first, second, and third formant frequency ranges in a particular speech block and generates a formant frequency signal 855 that is received by the beamformer 825.

[0030] The FFT 810 also generates a second speech FFT signal 850. Both the first speech FFT signal 845 and the second speech FFT signal 850 are received by the DOA 820. The DOA 820 may compute a cross-correlation between the two signals 845, 850. The resulting two peak signals 845, 850 are assumed to be speech and noise, respectively. If the DOA 820 determines that the second peak of the second signal 850 is not prominent, a null value is provided. This indicates that the noise is wideband and not concentrated around a narrow-band frequency. In general, the output of the DOA 820 are two angles in degrees, the first being for a desired speech signal while the second is for noise.

[0031] It should be noted that the assumption for the first signal 845 being for desired speech while the second signal 850 being for noise is also configurable. For example, in a situation where noise is louder than desired speech, the options may be changed so that the first signal 845 represents noise while the second signal 850 represents speech. Consequently, the second signal 850 may be received by the FDM 815 for the respective computations.

[0032] According to the exemplary embodiment in which two microphones are present, only two sources are detected. Upon the computations of the FFTs 805, 810, the FDM 815, and the DOA 820, the beamformer 825 receives the first speech FFT signal 845, the second speech FFT signal 850, the formant frequencies signal 855, and a DOA data signal 860.

[0033] The beamformer 825 places a null at the noise frequency direction for the formant range of frequencies, thereby eliminating the maximum noise in the range. This process may be performed for all the formant frequency ranges provided. The beamformer 825 may assume that the bandwidth of the formant range is B = [TL, TU], where L is the lower frequency of the formant range and U is the upper frequency of the formant range. It should be noted that the placement of a null is only exemplary. The beamformer 825 may further be used for other purposes. For example, with the signals received by the beamformer 825, modified signal enhancement may also be performed. That is, the beamformer 825 may generate modification data to be used to modify an audio signal to isolate a speech therein or used to enhance a speech of an audio signal.

[0034] The DOA 825 may initially select the desired FFT bin frequencies in the bandwidth range. The steering vector is determined by the following:

$$S(\theta) = \left[1, e^{-jkd\sin\theta}, e^{-2jkd\sin\theta}, \ldots, e^{-j(N-1)kd\sin\theta}\right]^T$$

[0035] Where k = 2 Π f / c, for M number of sources.

[0036] For M narrowband sources, the input vector is determined by the following:

$$X(t) = \left[\sum_{i=1}^{M} m_i(t) S_i \right] e^{j\omega t}$$

[0037] With w = [w1, w2, … ,wN] t as the weight vector, the array output is determined by the following:

$$Y(t) = w^T X(t)$$

[0038] Assuming θN is the direction of noise and θS is the direction of sound and the requirement is to place a null at θN and unity at θS, the individual weights for the two microphones is determined by the following:

$$w_1 = \frac{e^{-jkd\sin\theta N}}{e^{-jkd\sin\theta N} - e^{-jkd\sin\theta S}} \qquad w_2 = \frac{-1}{e^{-jkd\sin\theta N} - e^{-jkd\sin\theta S}}$$

[0039] The DOA 825 multiplies these weights to all the FFT bin frequencies in the formant ranges. Once the weights are multiplied, the DOA 825 generates an output signal 865 including the 128 samples. The IFFT 830 receives the output signal 865 which performs the inverse FFT to generate a speech signal 870 that has noise cancelled for that formant frequency range. Thus, the beamformer 825 receiving the above described signals is capable of canceling noise directly where noise cancellation is required and important.

[0040] It should be noted that the exemplary embodiments further account for other scenarios. For example, if a particular speech frame for a formant structure is not detected, the beamformer 825 may use the bandwidth range from 0 to 4000Hz to allow similar noise

suppression when a regular formant structure is missing. Such a scenario may arise, for example, during non-voiced syllables or fricatives. In another example, when the noise is wideband and a distinct direction for noise is not provided (*e.g.*, a null pointer is returned), the beamformer 825 may use a default value of 90 degrees to the user to attempt to cancel the wideband noise affecting the formant structure.

[0041] Fig. 9 shows a method 900 for a formant based noise cancellation according to an exemplary embodiment. The method 900 may relate to the device 800 and the components thereof including the signals that are passed therein. Therefore, the method 900 will be discussed with reference to the device 800 of Fig. 8. However, those skilled in the art will understand that the exemplary method is not limited to being performed on the exemplary hardware described in Fig. 8. For example, the method 900 may also be applied to multiple microphone systems including more than two microphones.

[0042] In step 905, the device 800 receives the raw audio data. As discussed above with reference to the exemplary embodiment of the device 800, the electronic device may include two microphones. Each microphone may generate respective raw audio data 835, 840. In another exemplary embodiment, the raw audio data may be received from more than two microphones. Each microphone may generate a respective raw audio data signal.

[0043] In step 910, the speech signal is processed. An initial step may be to determine which of the raw audio data signals comprises the speech signal. As

discussed above, a microphone may be designated as the speech receiving microphone. Other factors may be considered such as common formants, formants with known patterns, etc. Upon determining which microphone received the speech signal, a first processing may be the FFT. As discussed above, the speech signal is received at the FFT 805 for the computation to generate the first microphone speech signal 845. Subsequently, a second processing may be performed at the FDM 815. Once the FDM 815 receives the speech signal, the FDM 815 performs the respective computation to generate the formant frequencies signal 855.

[0044] In step 915, the other signals are processed. Upon the above described initial step, the remaining signals may be determined to be noise related. In the above exemplary embodiment of the electronic device 800, the remaining signal is the raw audio data 840. However, in other exemplary embodiments including more than two microphones, the remaining signals may include further raw audio data. The remaining raw audio data may be received at the FFT 810 for the computation to generate the second microphone speech signal 845.

[0045] In step 920, a direction of arrival for the audio data is determined. For example, the first and second microphone speech signals 845 and 850 are sent to the DOA 820 to perform the respective computation to generate the DOA data signal 860.

[0046] In step 925, the noise cancellation is processed. For example, all resulting signals are sent to the beamformer 825. Thus, the beamformer 825 receives

the first microphone speech signal 845, the second microphone speech signal 850, the formant frequencies signal 855, and the DOA data signal 860. Using these signals, the beamformer 825 is configured to perform the above described computations according to the exemplary embodiment for a particular frequency. The computations may also be performed for other frequencies. For example, with reference to the above described embodiment, 128 samples are generated by the beamformer 825.

[0047] In step 930, a modified audio signal is generated. For example, once the beamformer 825 performs all necessary computations, all samples are sent to the IFFT 830 which performs the respective computation to generate the modified audio signal 870 having only the speech data and canceling the noise data.

[0048] The exemplary embodiments provide a different approach for canceling out noise from an audio stream. Specifically, the noise cancellation is performed as a function of formant data and knowledge of psychoacoustics. Using this further information, conventional issues are bypassed in which spatial orientations can only cancel some noise. Spatial orientations also include other issues when noise data is mistaken for speech data and the conversion results in a bad audio stream. The use of formant data and psychoacoustics avoid these issues altogether.

[0049] Furthermore, the exemplary embodiments do not rely on techniques like spectral subtraction or Cepstrum synthesis where degradation of speech is possible due to

incorrect estimation of speech boundaries or pitch information. The exemplary embodiments instead rely on weight multiplication to the original FFT signal and then continues with IFFT, thereby maintaining a true fidelity of the speech signal to the maximum extent possible.

[0050] It will be apparent to those skilled in the art that various modifications may be made in the present invention, without departing from the spirit or scope of the invention. Thus, it is intended that the present invention cover the modifications and variations of this invention provided they come within the scope of the appended claims and their equivalents.

What is claimed is:

Claim 1. A noise cancellation device, comprising:

a plurality of first computation modules receiving raw audio data and generating a respective transformed signal as a function of formants, a first transformed signal relating to speech data and a second transformed signal relating to noise data;

a formant detection module receiving the first transformed signal and generating a frequency range data signal;

a direction of arrival module receiving the first and second transformed signals, determining a cross-correlation between the first and second transformed signals, and generating a spatial orientation data signal; and

a beamformer receiving the first and second transformed signals, the frequency range data signal, and the spatial orientation data signal and generating modification data at selected formant ranges to eliminate a maximum amount of the noise data.

Claim 2. The device of claim 1, further comprising:

a second computation module receiving the modification data to generate a modified audio data signal that isolates the speech data.

Claim 3. The device of claim 2, wherein the first computation modules are Fast Fourier Transform (FFT) modules.

Claim 4. The device of claim 3, wherein the second computation module is an inverse FFT module.

Claim 5. The device of claim 1, wherein the modification data further enhances the speech data.

Claim 6. The device of claim 1, wherein the transformed signals are separated into a plurality of frequency bins.

Claim 7. The device of claim 1, wherein the frequency range data signal includes a plurality of ranges for a predetermined speech block.

Claim 8. The device of claim 1, wherein the spatial orientation signal includes at least two angles, a first angle relating to the speech data and a second angle relating to the noise data.

Claim 9. The device of claim 1, wherein the modification data is generated at least using weighted data as a function of a direction of the speech and noise signals.

Claim 10. The device of claim 9, wherein the weighted data is incorporated to bin frequencies in selected formant ranges.

Claim 11. A method, comprising:

receiving raw audio data by a plurality of first computation modules, the first computation modules generating a respective transformed signal as a function of formants, a first transformed signal relating to speech data and a second transformed signal relating to noise data;

generating a frequency range data signal as a function of the first transformed signal;

generating a spatial orientation signal as a function of a cross-correlation between the first and second transformed signals; and

generating modification data at selected formant ranges to eliminate a maximum amount of the noise data as a function of the first and second transformed signals, the frequency range data signal, and the spatial orientation data signal.

Claim 12. The method of claim 11, further comprising:

generating a modified audio data signal that isolates the speech data as a function of the modification data.

Claim 13. The method of claim 12, wherein the first computation modules are FFT modules.

Claim 14. The method of claim 13, wherein the second computation module is an inverse FFT module.

Claim 15. The method of claim 11, wherein the modification data further enhances the speech data.

Claim 16. The method of claim 11, wherein the transformed signals are separated into a plurality of frequency bins.

Claim 17. The method of claim 11, wherein the frequency range data signal includes a plurality of ranges for a predetermined speech block.

Claim 18. The method of claim 11, wherein the spatial orientation signal includes at least two angles, a first

angle relating to the speech data and a second angle relating to the noise data.

Claim 19. The method of claim 11, wherein the modification data is generated at least using weighted data as a function of a direction of the speech and noise signals.

Claim 20. A noise cancellation device, comprising:

a plurality of first computing means that receive raw audio data for generating a respective transformed signal as a function of formants, a first transformed signal relating to speech data and a second transformed signal relating to noise data;

a formant detecting means that receive the first transformed signal for generating a frequency range data signal;

a direction of arrival determining means that receive the first and second transformed signals for determining a cross-correlation between the first and second transformed signals and for generating a spatial orientation data signal; and

a beamforming means that receive the first and second transformed signals, the frequency range data signal, and the spatial orientation data signal for generating modification data at selected formant ranges to eliminate a maximum amount of the noise data.

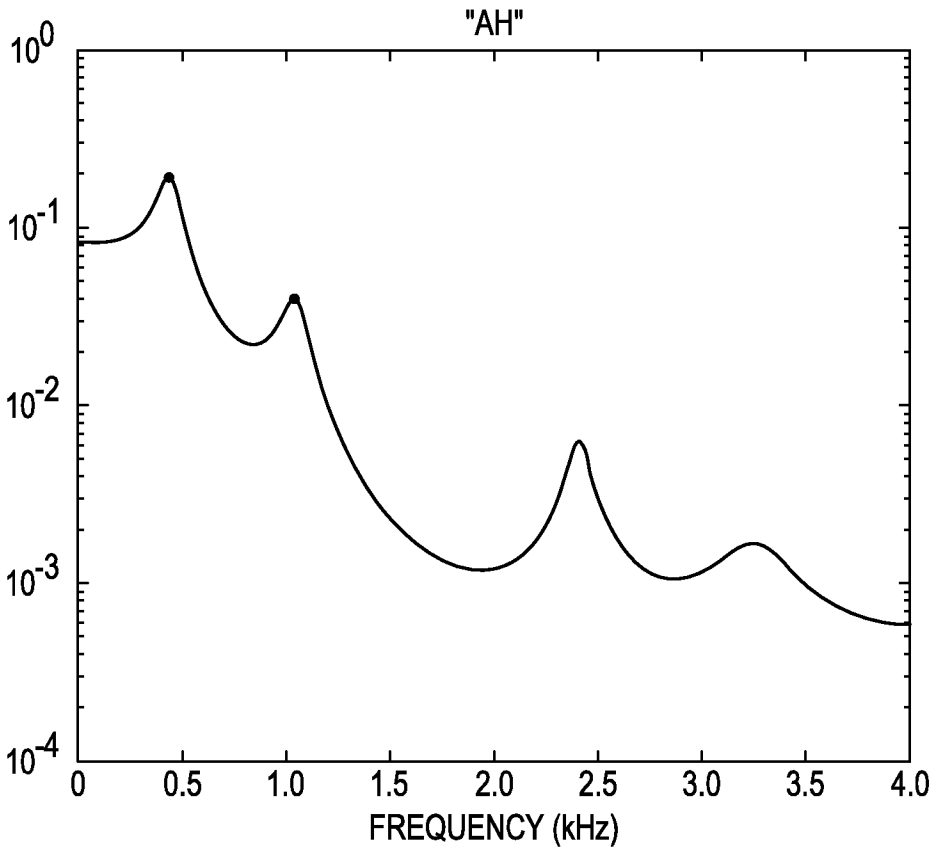

"AH"
10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
FREQUENCY (kHz)

FIG. 1A

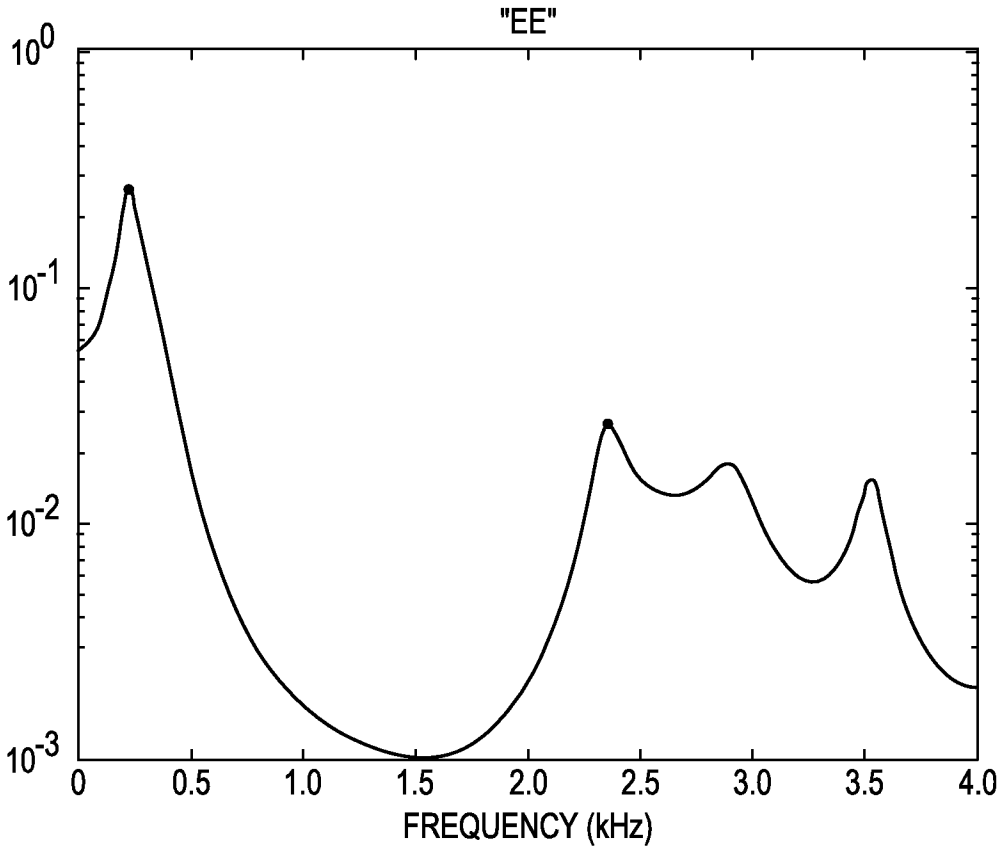

"EE"
10^0
10^{-1}
10^{-2}
10^{-3}
0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
FREQUENCY (kHz)

FIG. 1B

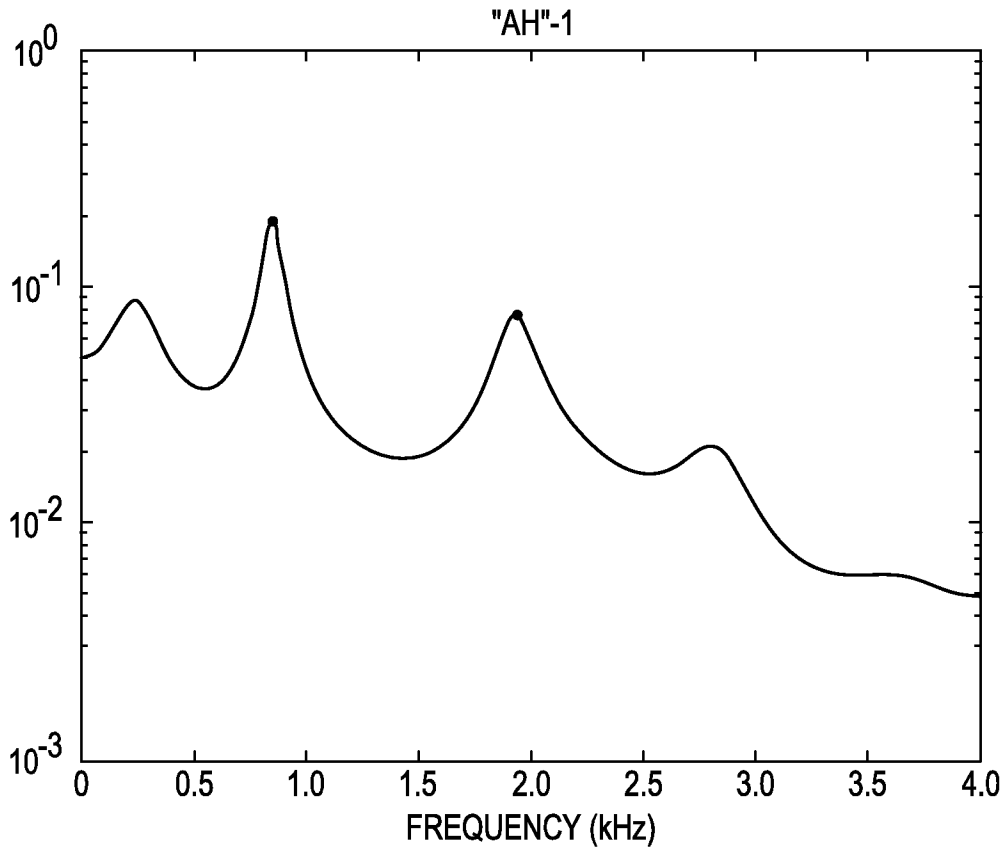

"AH"-1
10^0
10^{-1}
10^{-2}
10^{-3}
0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
FREQUENCY (kHz)

FIG. 2A

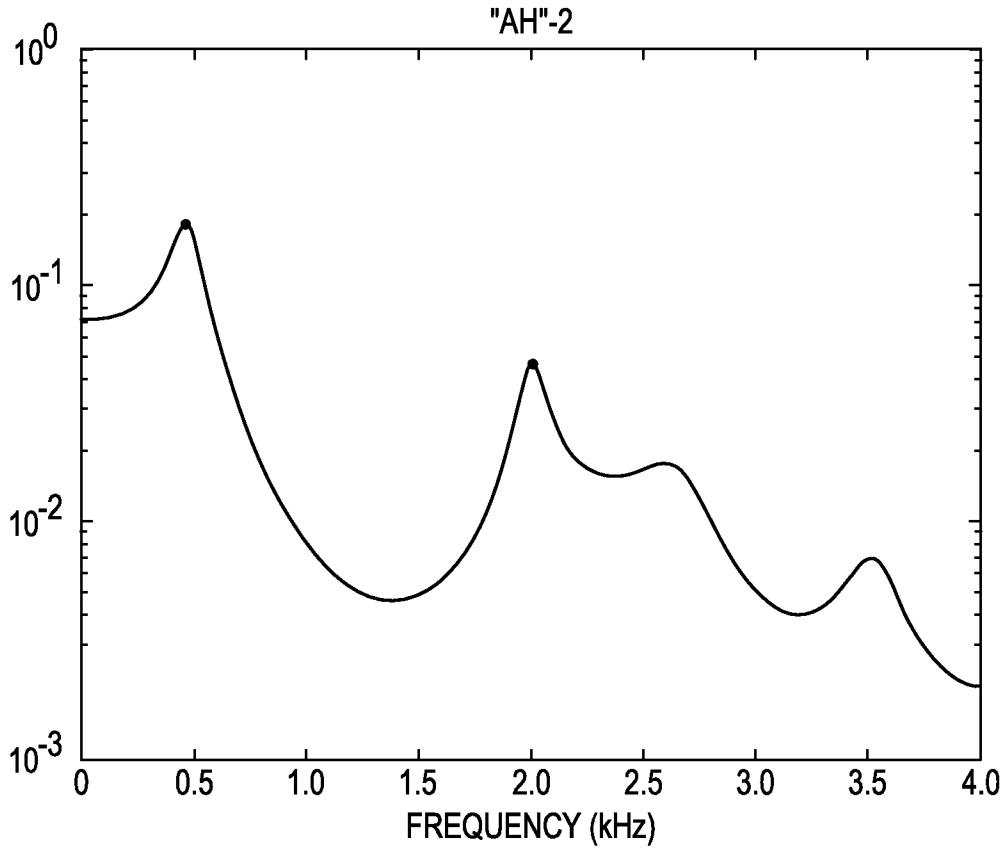

"AH"-2
10^0
10^{-1}
10^{-2}
10^{-3}
0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
FREQUENCY (kHz)

FIG. 2B

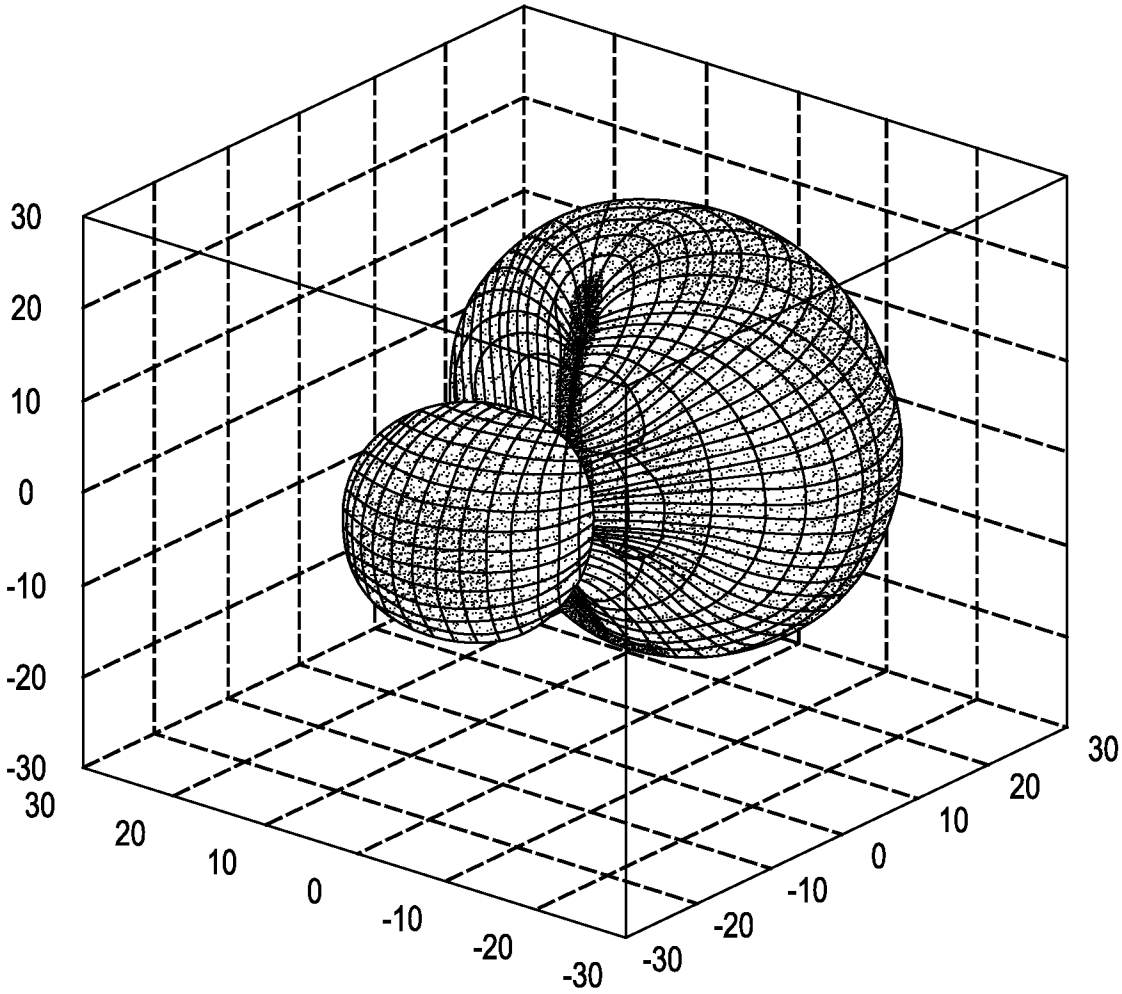


FIG. 3

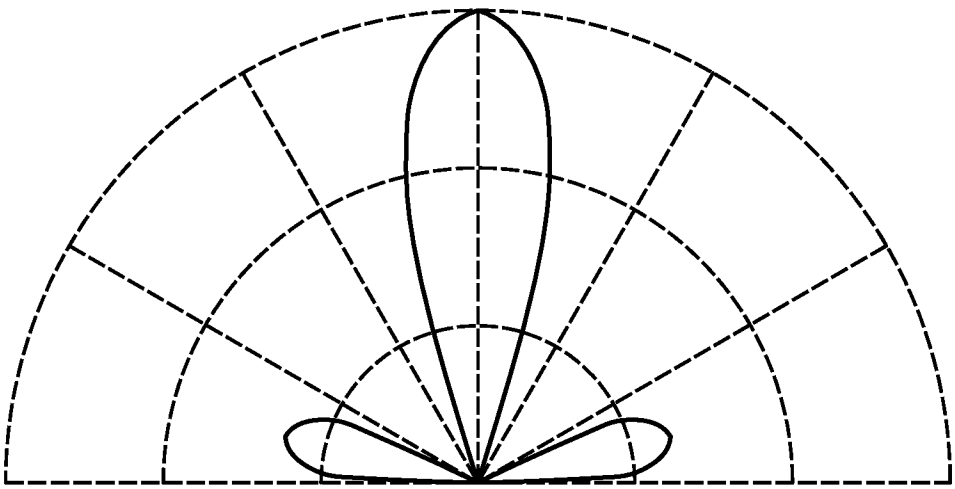


FIG. 4

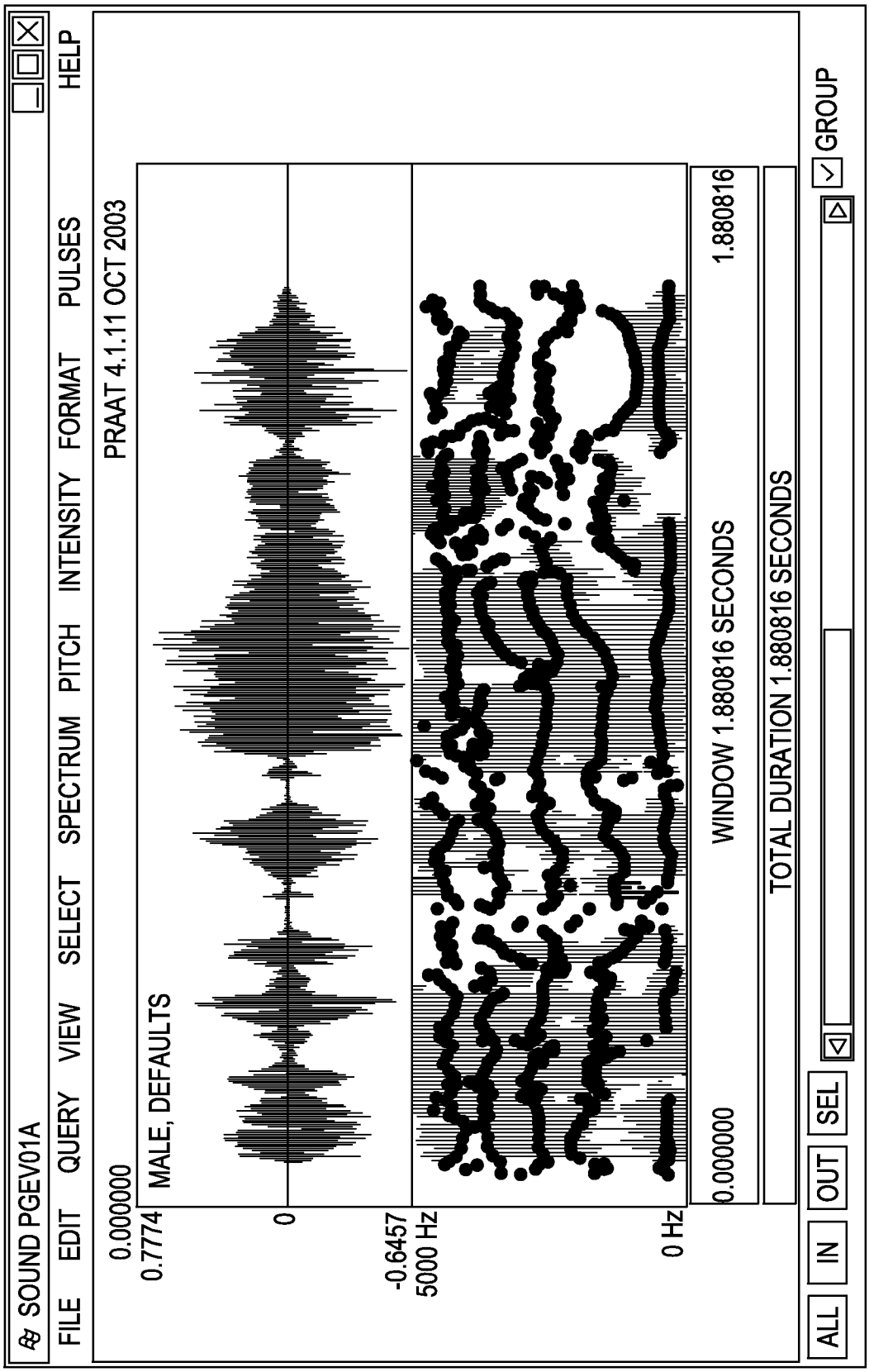

SOUND PGEV01A
FILE EDIT QUERY VIEW SELECT SPECTRUM PITCH INTENSITY FORMAT PULSES HELP
PRAAT 4.1.11 OCT 2003
MALE, DEFAULTS
0.000000
0.7774
0
-0.6457
5000 Hz
0 Hz
0.000000
WINDOW 1.880816 SECONDS
1.880816
TOTAL DURATION 1.880816 SECONDS
ALL IN OUT SEL
GROUP

FIG. 5

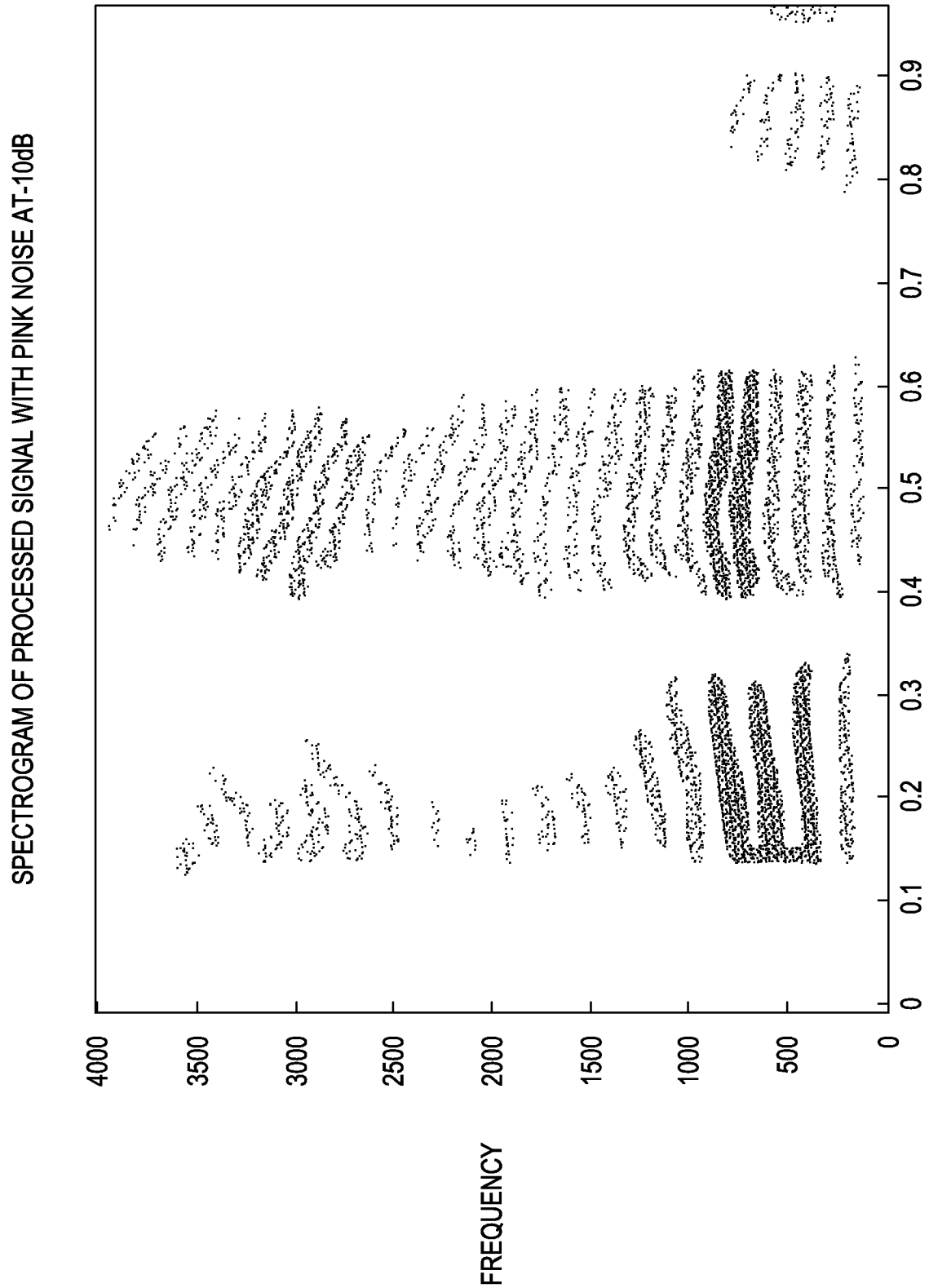

SPECTROGRAM OF PROCESSED SIGNAL WITH PINK NOISE AT-10dB
FREQUENCY
4000
3500
3000
2500
2000
1500
1000
500
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
TIME

FIG. 6

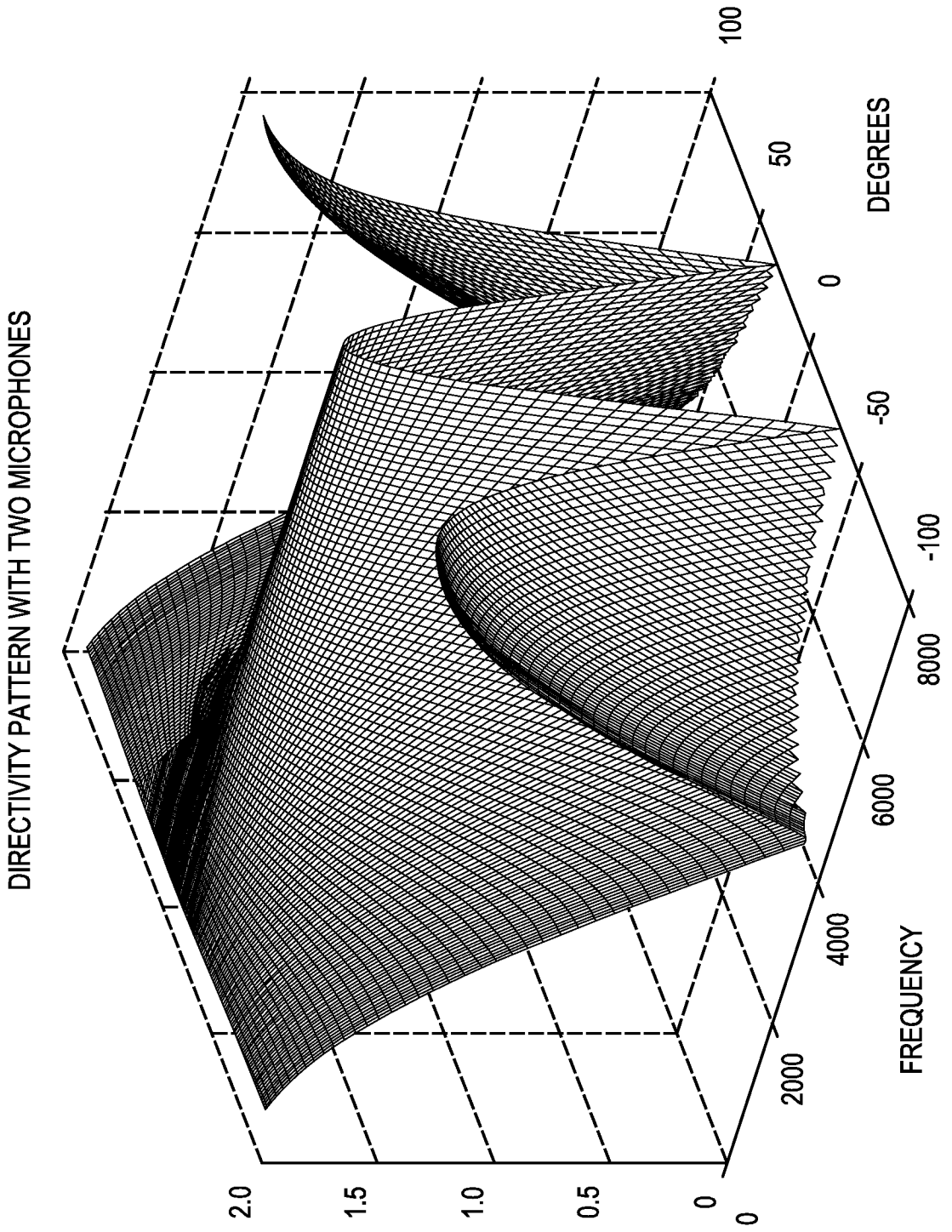

DIRECTIVITY PATTERN WITH TWO MICROPHONES
2.0
1.5
1.0
0.5
0
0
2000
4000
6000
8000
FREQUENCY
100
50
0
-50
-100
DEGREES

FIG. 7

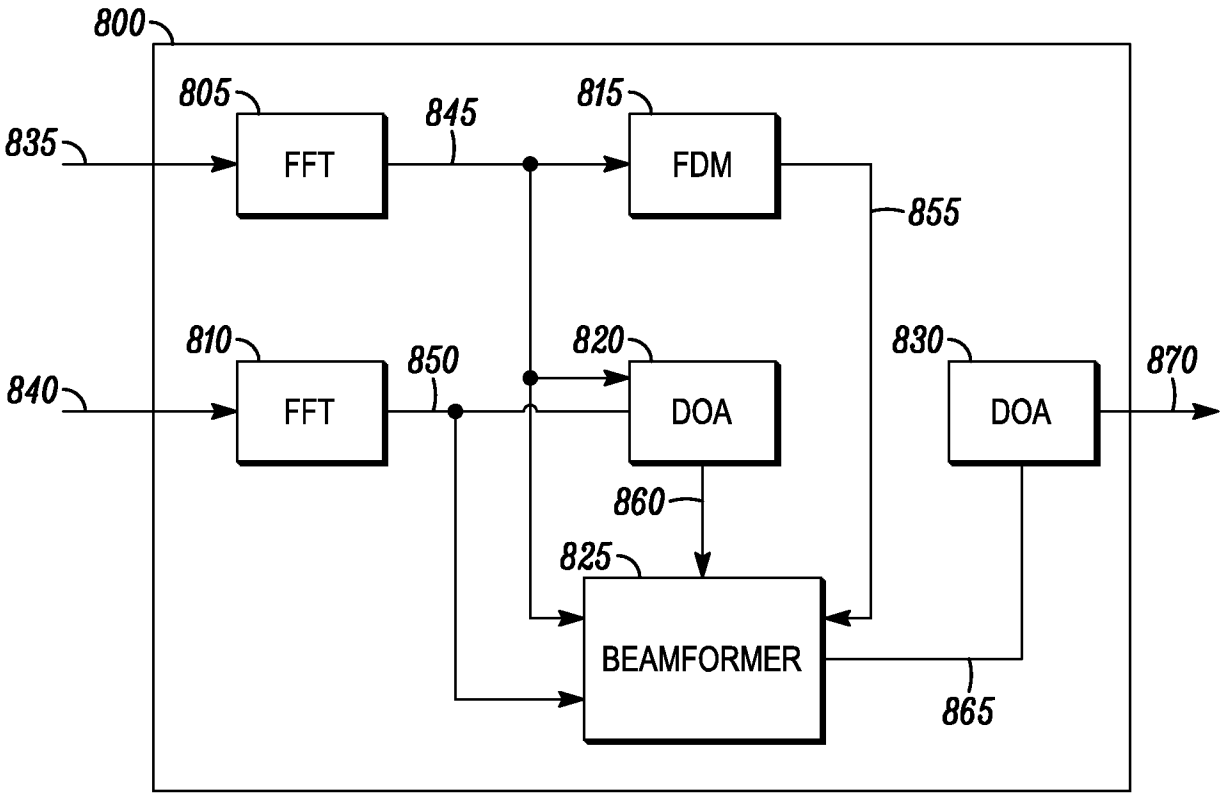

800
805
FFT
845
815
FDM
835
855
810
FFT
850
820
DOA
830
DOA
870
840
860
825
BEAMFORMER
865

FIG. 8

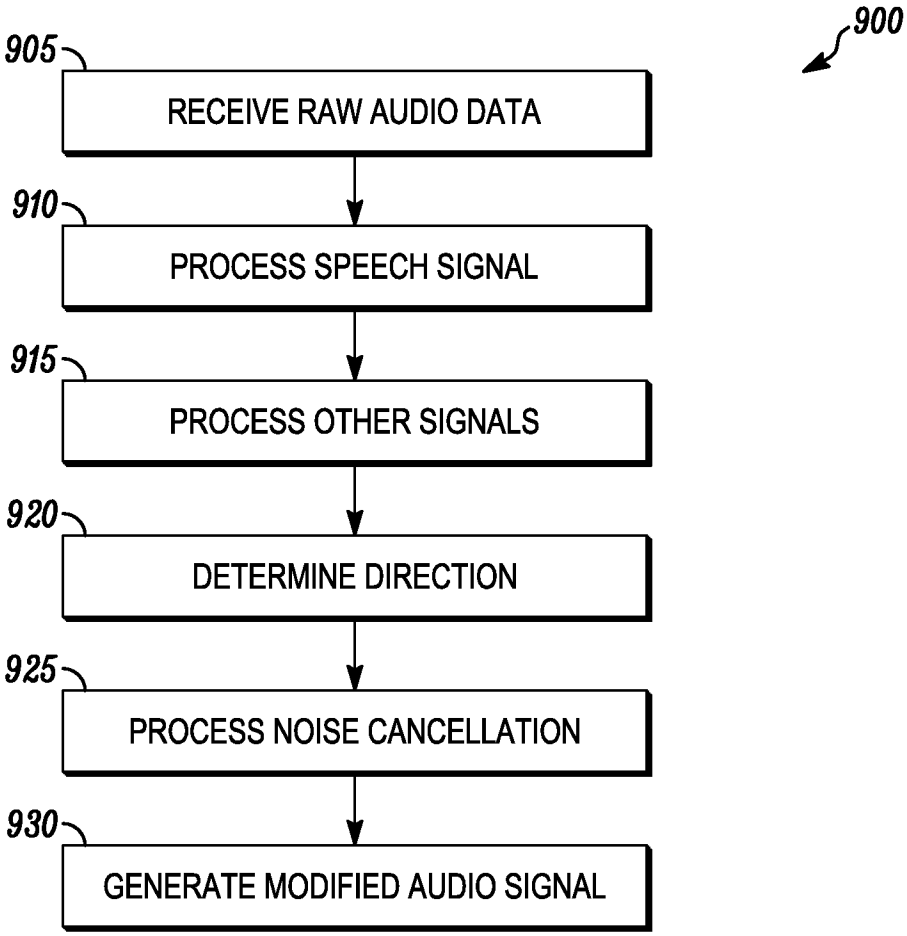

900
905
RECEIVE RAW AUDIO DATA
910
PROCESS SPEECH SIGNAL
915
PROCESS OTHER SIGNALS
920
DETERMINE DIRECTION
925
PROCESS NOISE CANCELLATION
930
GENERATE MODIFIED AUDIO SIGNAL

FIG. 9

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2011/043115

A. CLASSIFICATION OF SUBJECT MATTER

G10L 21/02(2006.01)i, G10K 11/16(2006.01)i

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)
G10L 21/02; H03H 21/00; H04R 3/00; G10L 15/02; H04R 1/40; H04R 1/20; G01S 3/802

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Korean utility models and applications for utility models
Japanese utility models and applications for utility models

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
eKOMPASS(KIPO internal) & Keywords:"formant", "beamf*", "noise", "speech"

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2001-100800 A (TOSHIBA CORP.) 13 April 2001 See the abstract; claim 1; and figure 1.	1-20
A	KR 10-2006-0091591 A (SAMSUNG ELECTRONICS CO., LTD.) 21 August 2006 See the abstract; claims 1-6; and figure 1.	1-20
A	JP 2007-535853 A (KONINKLIJKE PHILIPS ELECTRONICS N.V.) 06 December 2007 See the abstract; claim 1; and figure 1.	1-20
A	KR 10-2008-0087939 A (INDUSTRY-ACADEMIC COOPERATION FOUNDATION GYEONGSANG NATIONAL UNIVERSITY) 02 October 2008 See the abstract; claims 1-7; and figure 1.	1-20

☐ 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 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	Date of mailing of the international search report
04 JANUARY 2012 (04.01.2012)	**05 JANUARY 2012 (05.01.2012)**

Name and mailing address of the ISA/KR	Authorized officer
Korean Intellectual Property Office Government Complex-Daejeon, 189 Cheongsa-ro, Seo-gu, Daejeon 302-701, Republic of Korea Facsimile No. 82-42-472-7140	JUNG, Sung Yun Telephone No. 82-42-481-8483

Form PCT/ISA/210 (second sheet) (July 2009)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2011/043115

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
JP 2001-100800 A	13.04.2001	JP 3484112 B2	06.01.2004
KR 10-2006-0091591 A	21.08.2006	None	
JP 2007-535853 A	06.12.2007	CN 1947171 A	11.04.2007
		CN 1947171 B	04.05.2011
		EP 1743323 A1	17.01.2007
		KR 10-2007-0004893 A	09.01.2007
		US 2007-0273585 A1	29.11.2007
		US 7957542 B2	07.06.2011
		WO 2005-106841 A1	10.11.2005
KR 10-2008-0087939 A	02.10.2008	None	