US 20120076316A1

(54) **MICROPHONE ARRAY SYSTEM**

(76) Inventors: **Manli Zhu**, Pearl River, NY (US);
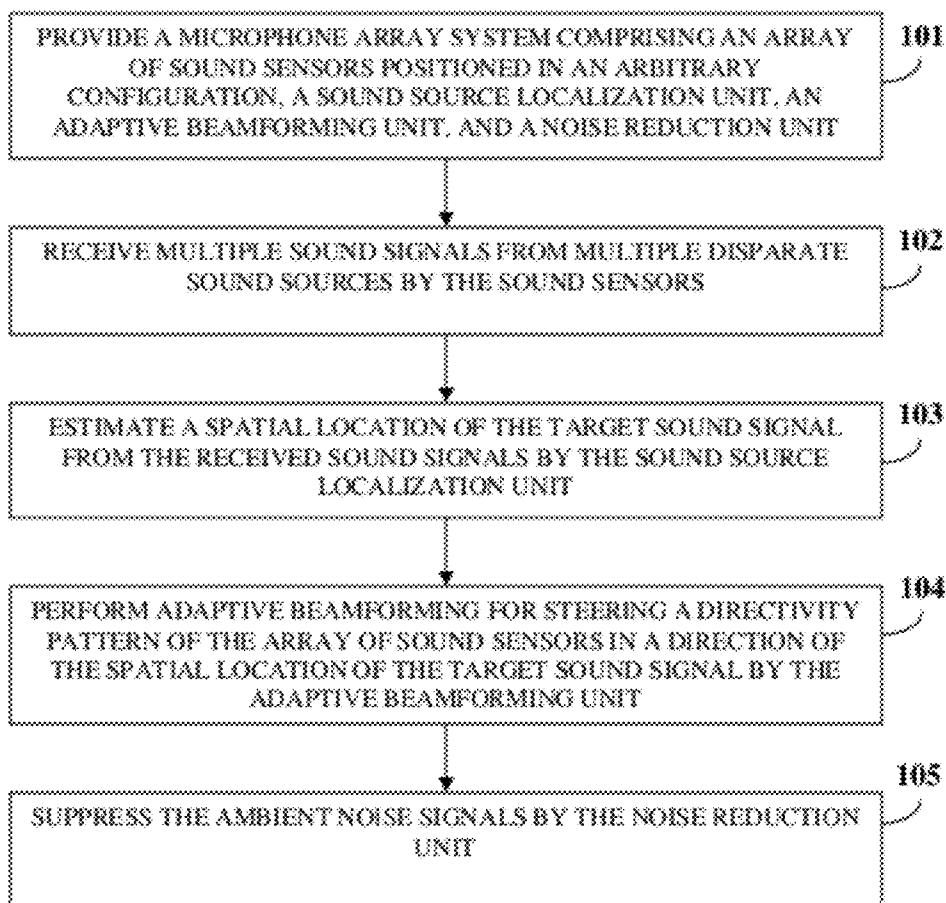**Qi Li**, New Providence, NJ (US)

(57) **ABSTRACT**

A method and system for enhancing a target sound signal from multiple sound signals is provided. An array of an arbitrary number of sound sensors positioned in an arbitrary configuration receives the sound signals from multiple disparate sources. The sound signals comprise the target sound signal from a target sound source, and ambient noise signals. A sound source localization unit, an adaptive beamforming unit, and a noise reduction unit are in operative communication with the array of sound sensors. The sound source localization unit estimates a spatial location of the target sound signal from the received sound signals. The adaptive beamforming unit performs adaptive beamforming by steering a directivity pattern of the array of sound sensors in a direction of the spatial location of the target sound signal, thereby enhancing the target sound signal and partially suppressing the ambient noise signals, which are further suppressed by the noise reduction unit.

PROVIDE A MICROPHONE ARRAY SYSTEM COMPRISING AN ARRAY OF SOUND SENSORS POSITIONED IN AN ARBITRARY CONFIGURATION, A SOUND SOURCE LOCALIZATION UNIT, AN ADAPTIVE BEAMFORMING UNIT, AND A NOISE REDUCTION UNIT — 101

RECEIVE MULTIPLE SOUND SIGNALS FROM MULTIPLE DISPARATE SOUND SOURCES BY THE SOUND SENSORS — 102

ESTIMATE A SPATIAL LOCATION OF THE TARGET SOUND SIGNAL FROM THE RECEIVED SOUND SIGNALS BY THE SOUND SOURCE LOCALIZATION UNIT — 103

PERFORM ADAPTIVE BEAMFORMING FOR STEERING A DIRECTIVITY PATTERN OF THE ARRAY OF SOUND SENSORS IN A DIRECTION OF THE SPATIAL LOCATION OF THE TARGET SOUND SIGNAL BY THE ADAPTIVE BEAMFORMING UNIT — 104

SUPPRESS THE AMBIENT NOISE SIGNALS BY THE NOISE REDUCTION UNIT — 105

PROVIDE A MICROPHONE ARRAY SYSTEM COMPRISING AN ARRAY
OF SOUND SENSORS POSITIONED IN AN ARBITRARY
CONFIGURATION, A SOUND SOURCE LOCALIZATION UNIT, AN
ADAPTIVE BEAMFORMING UNIT, AND A NOISE REDUCTION UNIT — 101

RECEIVE MULTIPLE SOUND SIGNALS FROM MULTIPLE DISPARATE
SOUND SOURCES BY THE SOUND SENSORS — 102

ESTIMATE A SPATIAL LOCATION OF THE TARGET SOUND SIGNAL
FROM THE RECEIVED SOUND SIGNALS BY THE SOUND SOURCE
LOCALIZATION UNIT — 103

PERFORM ADAPTIVE BEAMFORMING FOR STEERING A DIRECTIVITY
PATTERN OF THE ARRAY OF SOUND SENSORS IN A DIRECTION OF
THE SPATIAL LOCATION OF THE TARGET SOUND SIGNAL BY THE
ADAPTIVE BEAMFORMING UNIT — 104

SUPPRESS THE AMBIENT NOISE SIGNALS BY THE NOISE REDUCTION
UNIT — 105

FIG. 1

FIG. 2

FIG. 3

FIG. 4

FIG. 5

Sound Sensor	Distance (m)	Delay τ (number of samples)
M0	$d*\cos(\theta+\Phi_0)$	$d*\cos(\theta+\Phi_0)*f/c$
M1	$d*\cos(\theta-\Phi_1)$	$d*\cos(\theta-\Phi_1)*f/c$
M2	$-d*\cos(\theta+\Phi_2)$	$-d*\cos(\theta+\Phi_2)*f/c$
M3	$-d*\cos(\theta-\Phi_3)$	$-d*\cos(\theta-\Phi_3)*f/c$

FIG. 6A

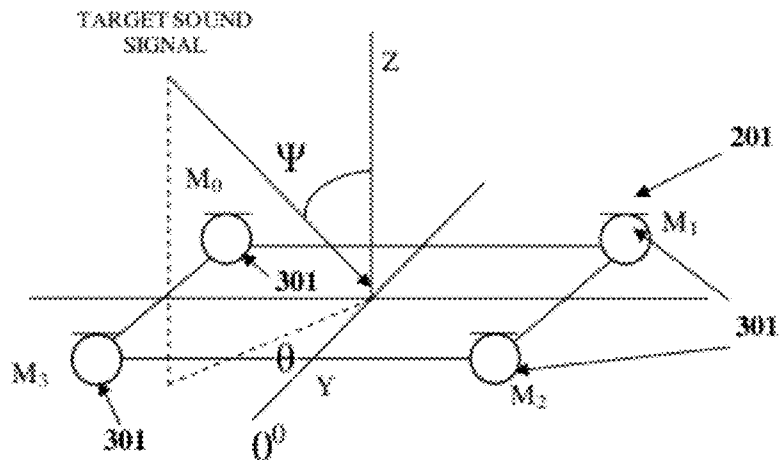| Sound Sensor Position	Distance (m)	Delay τ (number of samples)
0°	$-d*\cos(\theta)$	$-d*\cos(\theta)*f/c$
180°	$d*\cos(\theta)$	$d*\cos(\theta)*f/c$
90°	$-d*\sin(\theta)$	$-d*\sin(\theta)*f/c$
-90°	$d*\sin(\theta)$	$d*\sin(\theta)*f/c$
Φ clockwise away from 0° $(0 \leq \Phi \leq 90^\circ)$	$-d*\cos(\theta-\Phi)$	$-d*\cos(\theta-\Phi)*f/c$
Φ anticlockwise away from 0° $(0 \leq \Phi \leq 90^\circ)$	$-d*\cos(\theta+\Phi)$	$-d*\cos(\theta+\Phi)*f/c$
Φ clockwise away from 180° $(0 \leq \Phi \leq 90^\circ)$	$d*\cos(\theta-\Phi)$	$d*\cos(\theta-\Phi)*f/c$
Φ anticlockwise away from 180° $(0 \leq \Phi \leq 90^\circ)$	$d*\cos(\theta+\Phi)$	$d*\cos(\theta+\Phi)*f/c$

FIG. 6B

FIG. 7A

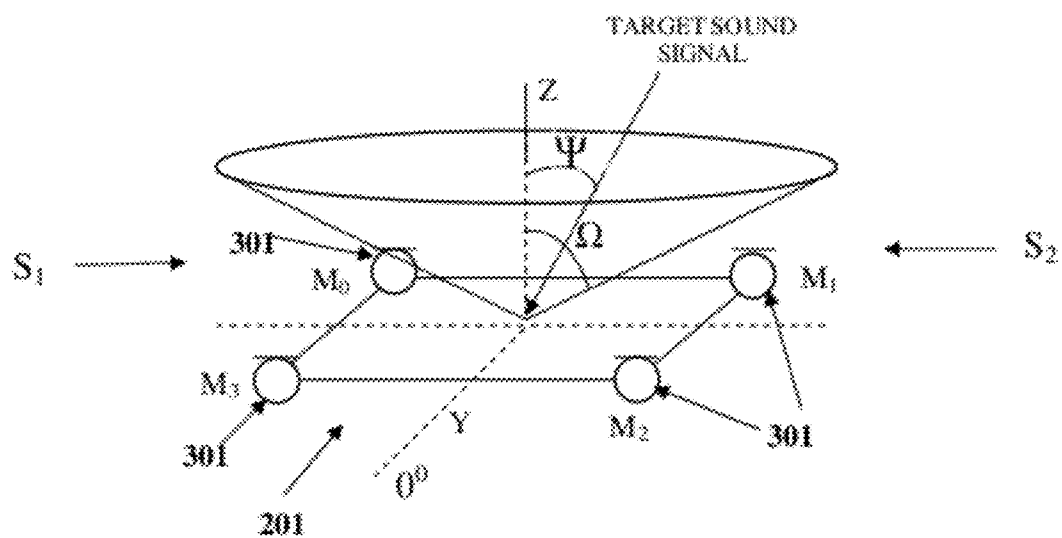| $\tau_0 =$ $d \cdot fs \cdot \cos(\theta + \Phi_0)\sin(\Psi)/c$	$\tau_1 =$ $d \cdot fs \cdot \cos(\theta - \Phi_1)\sin(\Psi)/c$	$\tau_2 =$ $-d \cdot fs \cdot \cos(\theta + \Phi_2)\sin(\Psi)/c$	$\tau_3 =$ $-d \cdot fs \cdot \cos(\theta - \Phi_3)\sin(\Psi)/c$

FIG. 7B

FIG. 7C

FOR DIRECTION i $(0 \le i \le 360)$, CALCULATE THE DELAY D_{it} BETWEEN THE tth PAIR OF THE SOUND SENSORS $(t=1$: ALL PAIRS) — 801

CALCULATE THE CORRELATION VALUE $corr(D_{it})$ BETWEEN THE tth PAIR OF THE SOUND SENSORS CORRESPONDING TO THE DELAY OF D_{it} — 802

FOR THE DIRECTION i $(0 \le i \le 360)$,

$$CORR_i = \sum_{t=1}^{ALLPAIR} corr(D_{it})$$

— 803

THE TARGET SOUND SIGNAL COMES FROM DIRECTION
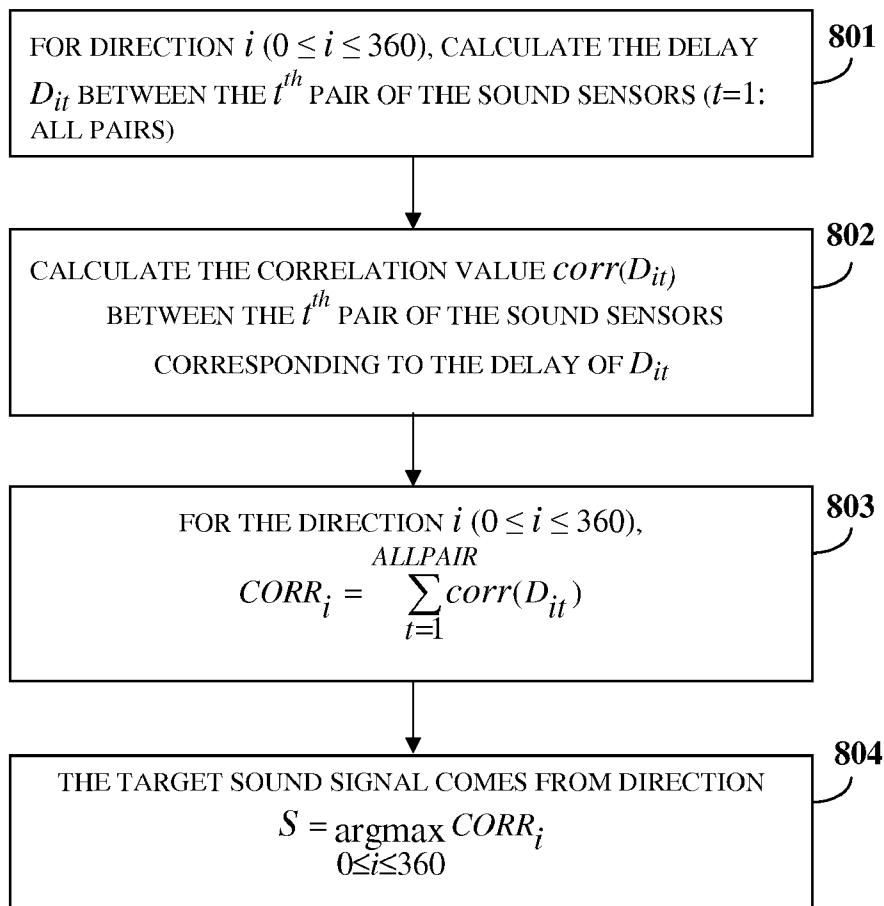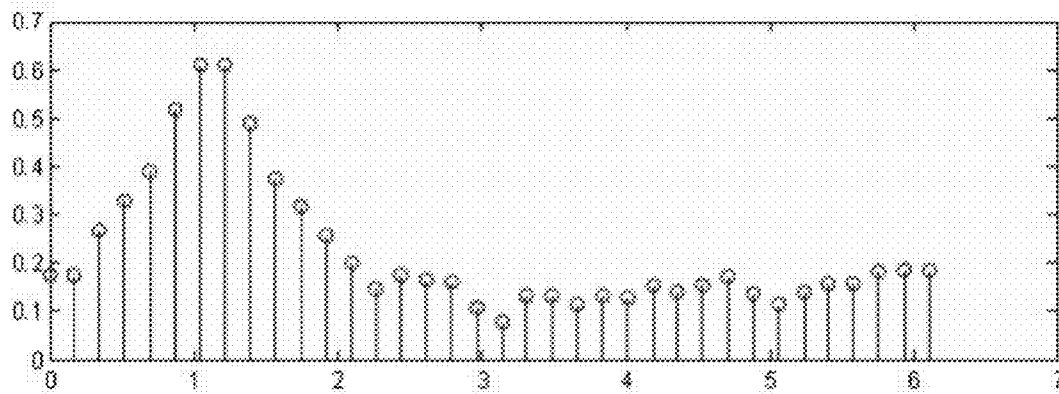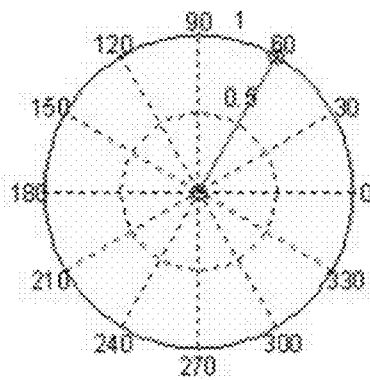
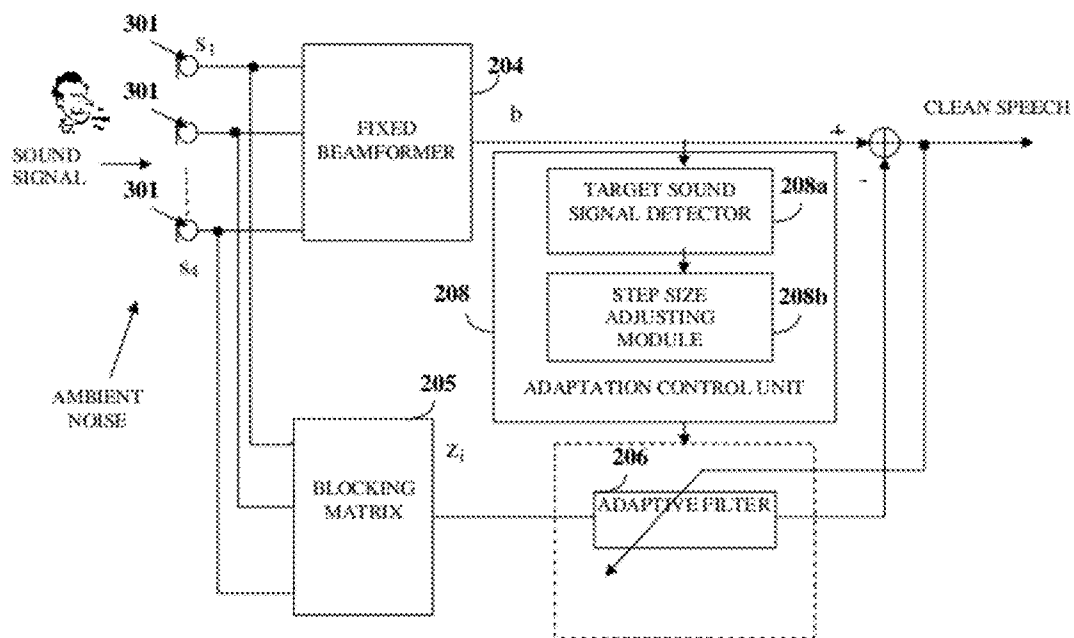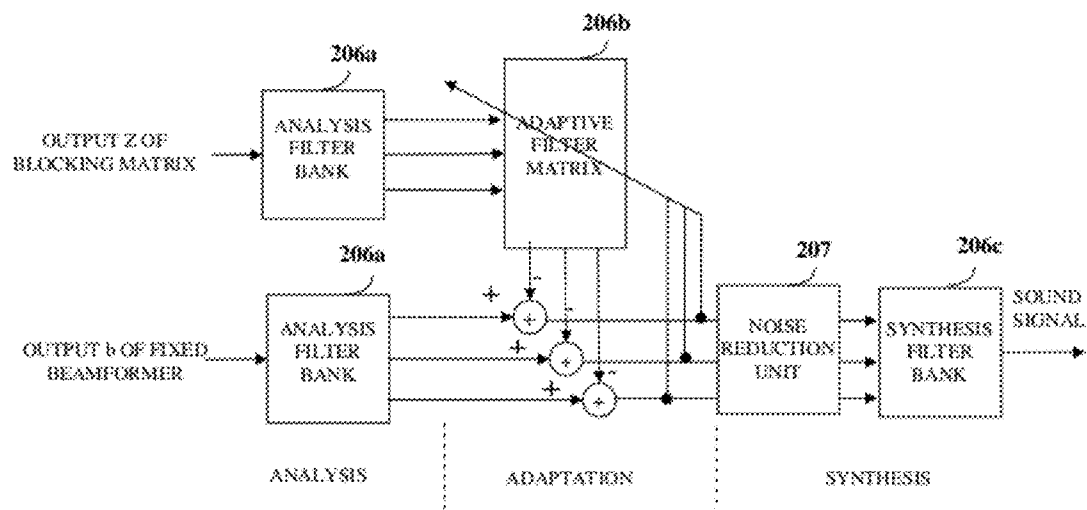$$S = \underset{0 \le i \le 360}{\arg\max} \, CORR_i$$

— 804

FIG. 8

FIG. 9A

FIG. 9B

FIG. 10

FIG. 11
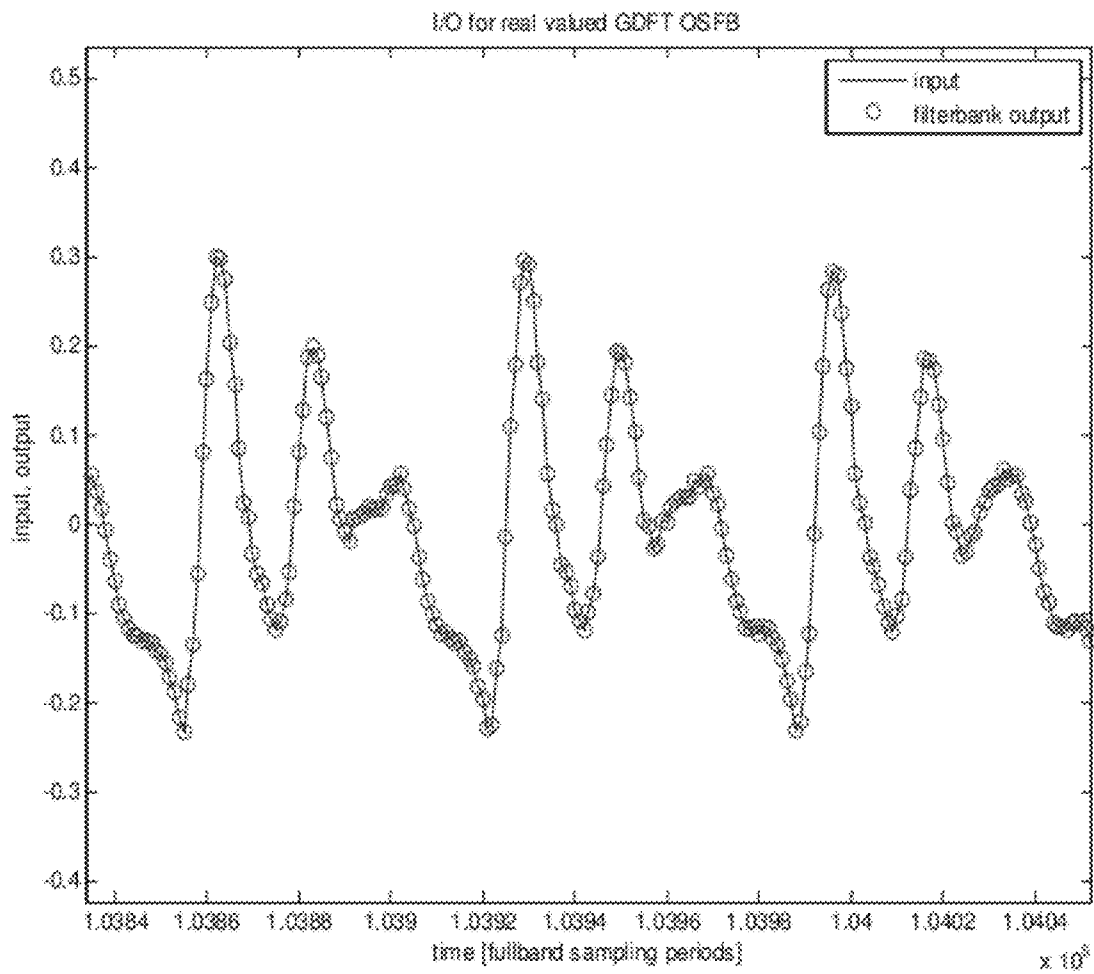
FIG. 12
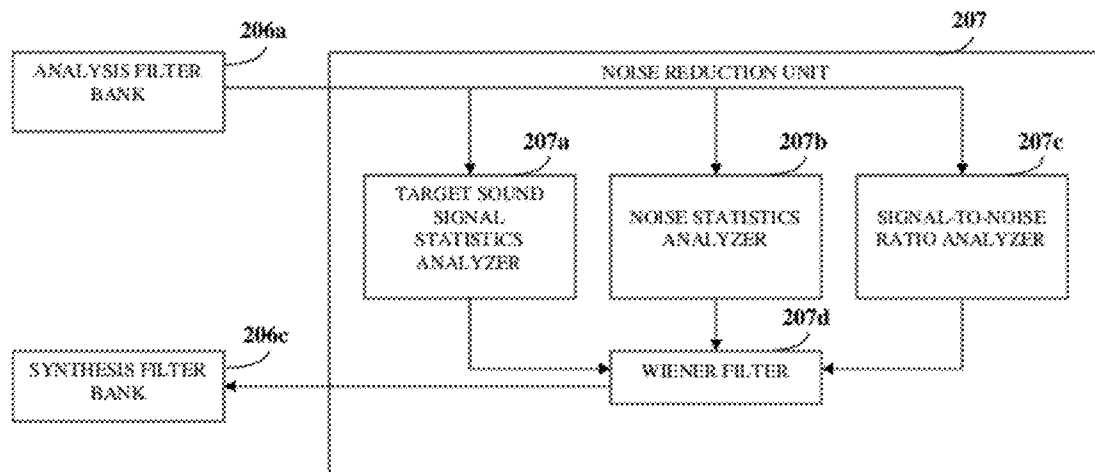
FIG. 13
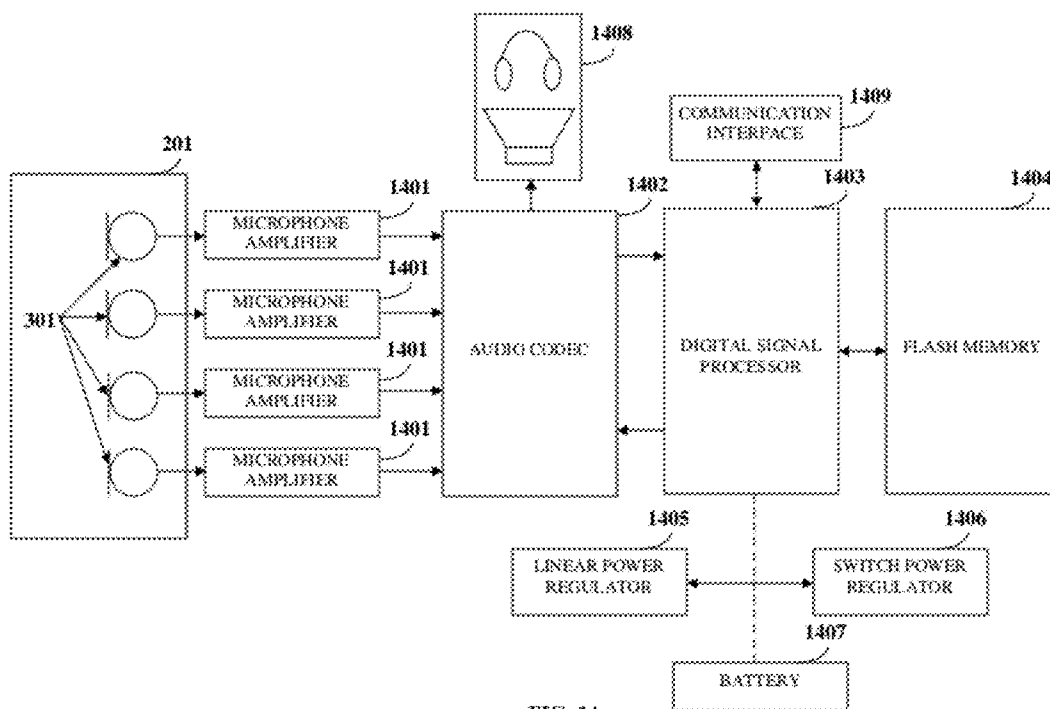
FIG. 14

FIG. 15A

FIG. 15B

1502
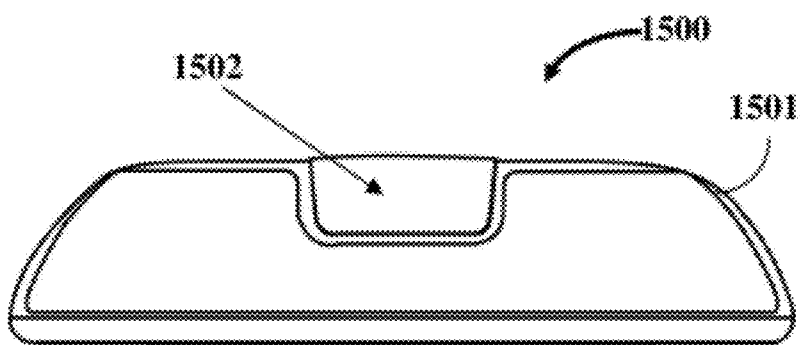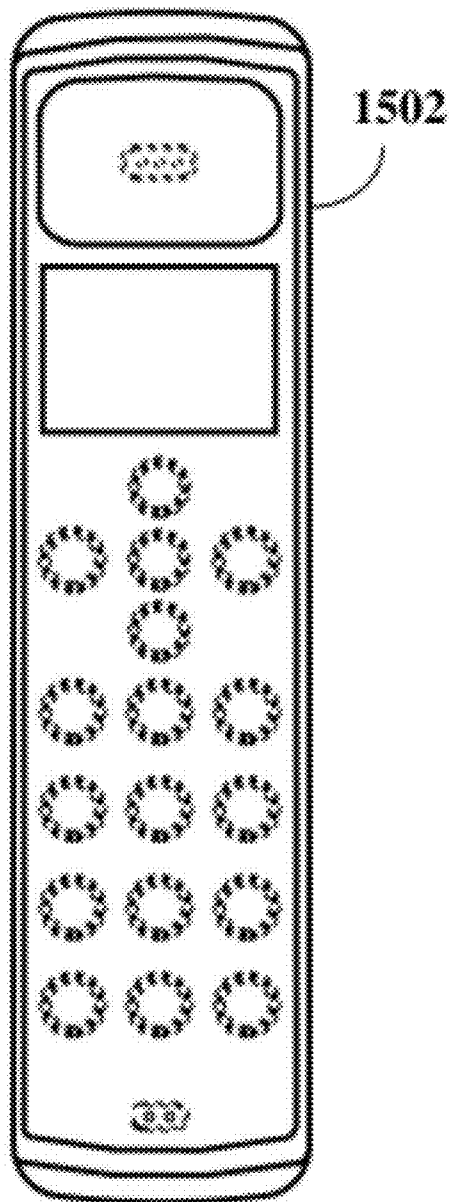
FIG. 15C
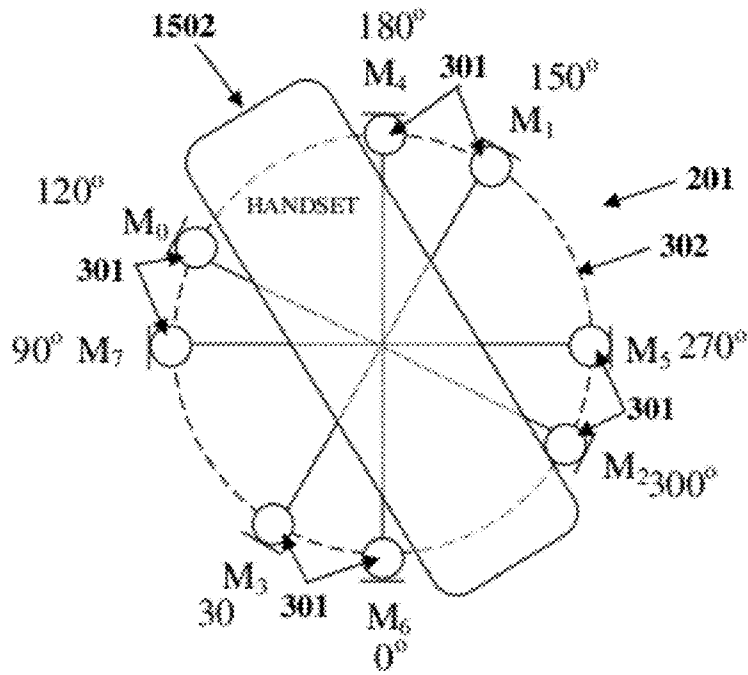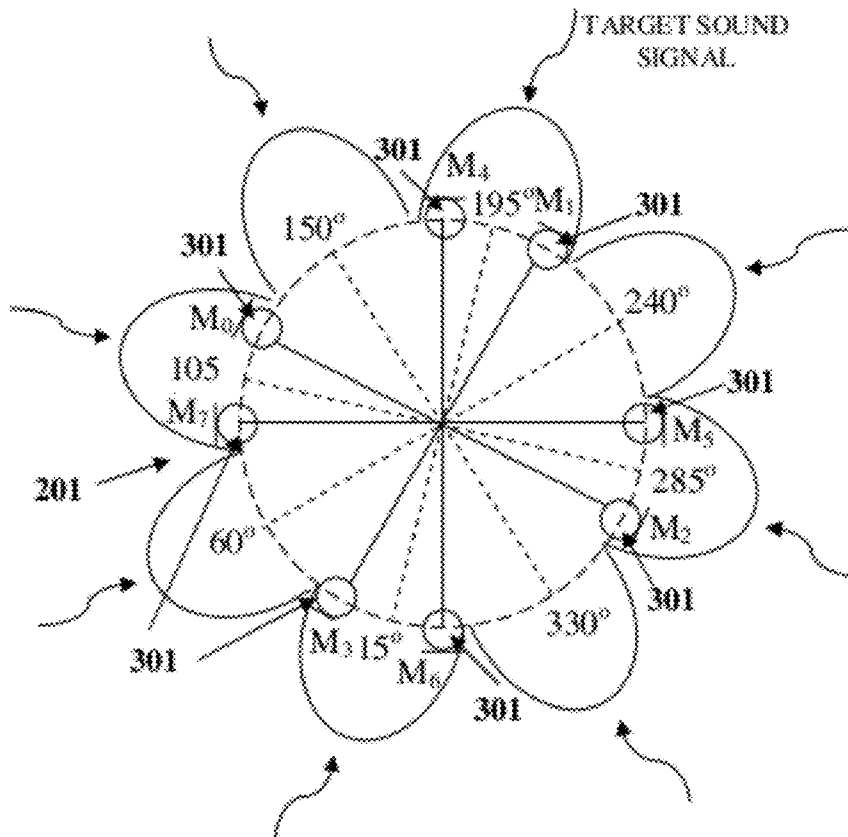
FIG. 16A

FIG. 16B
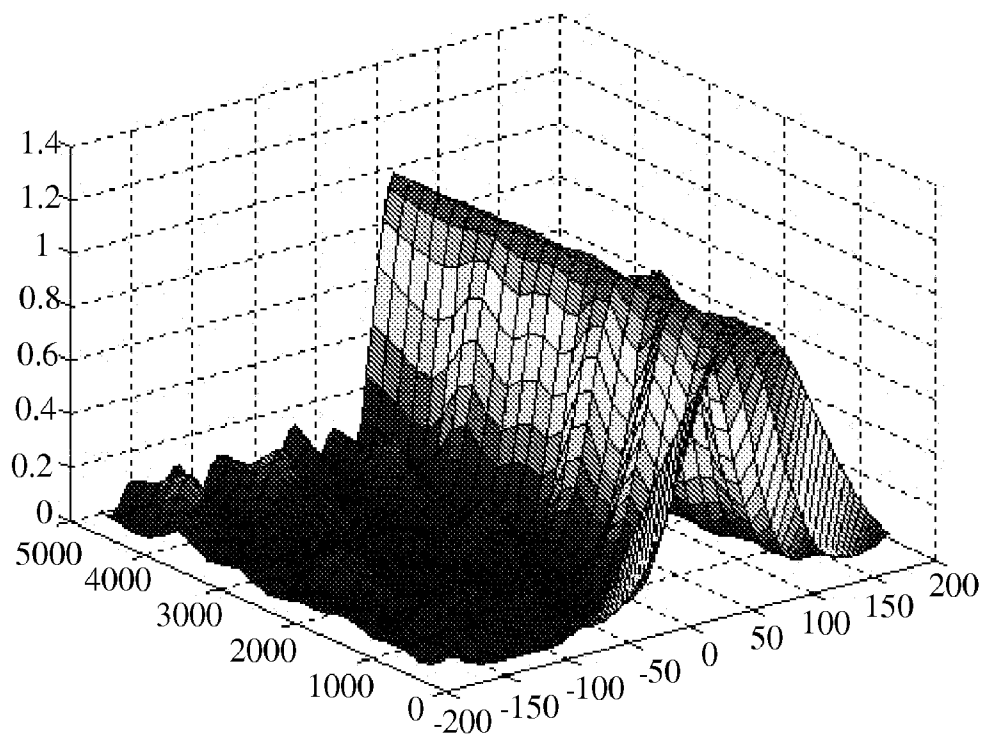
FIG. 16C
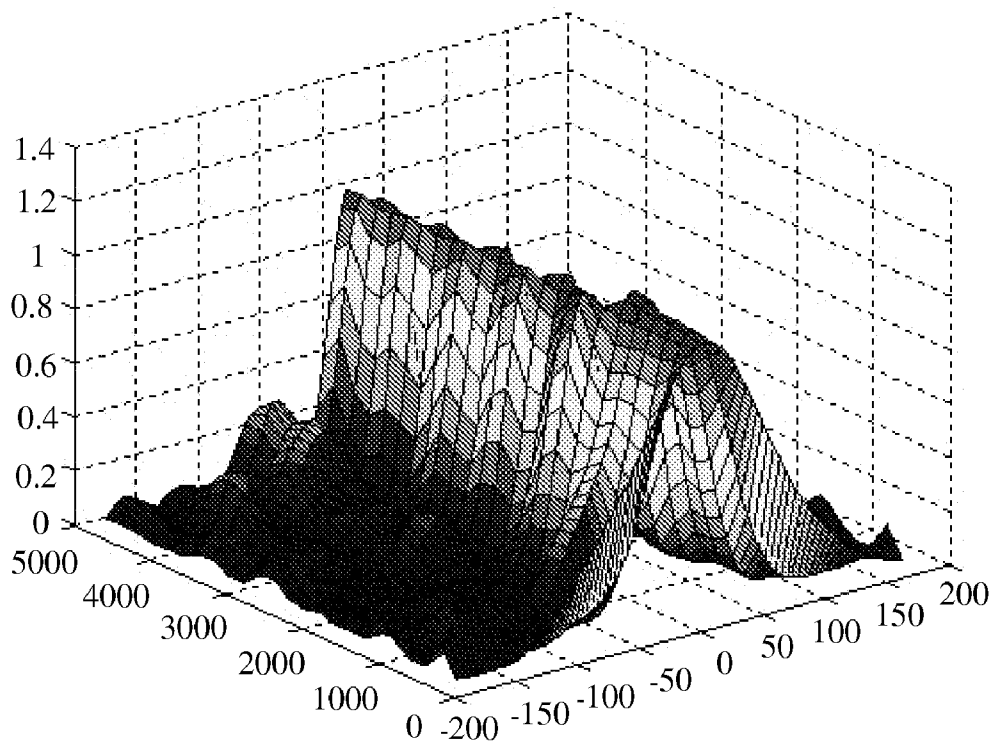
FIG. 16D
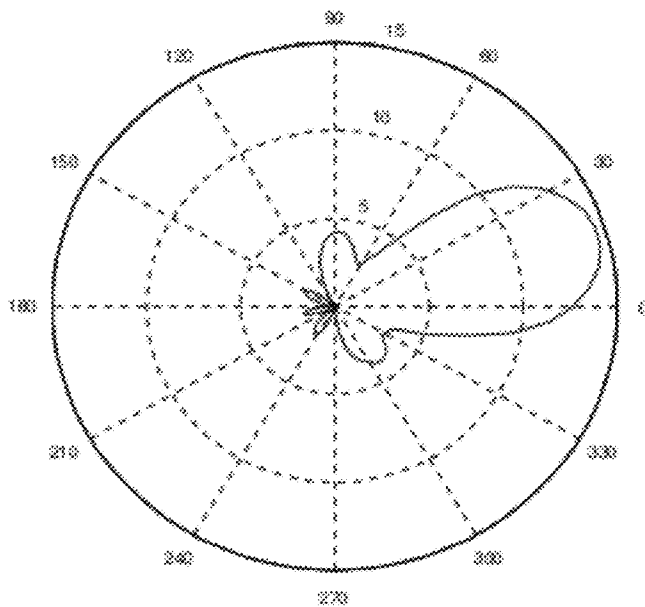
FIG. 16E
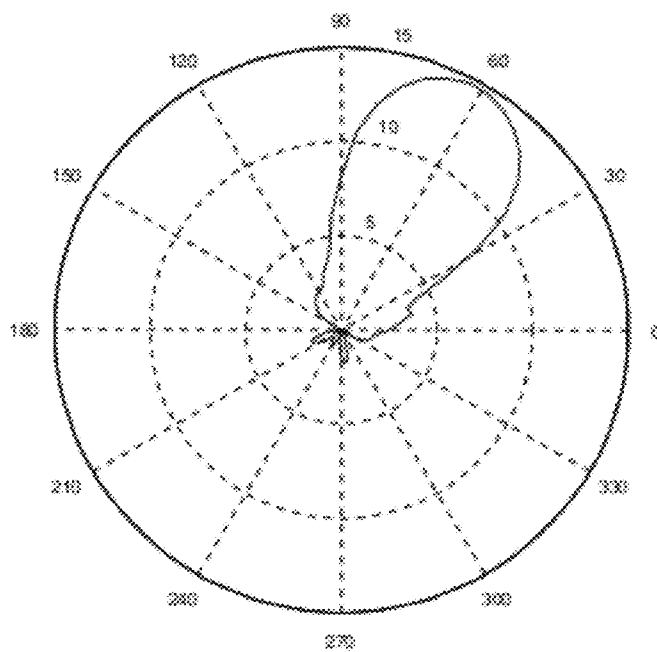
FIG. 16F
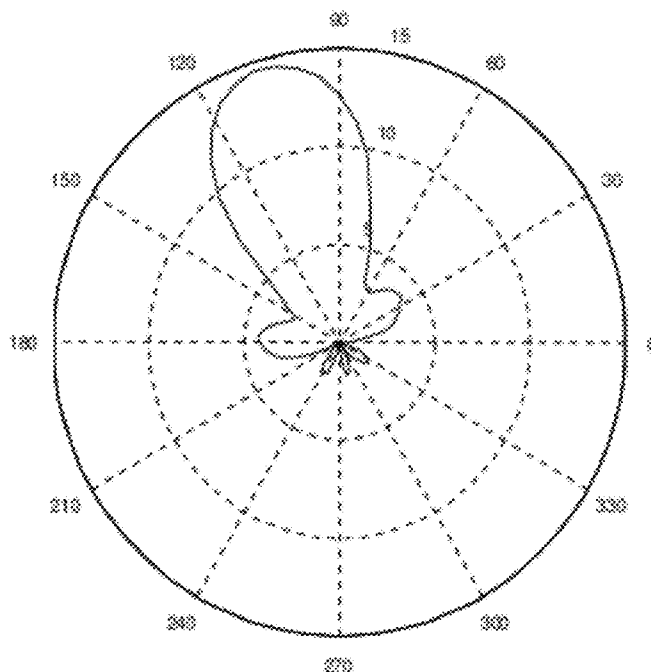
FIG. 16G
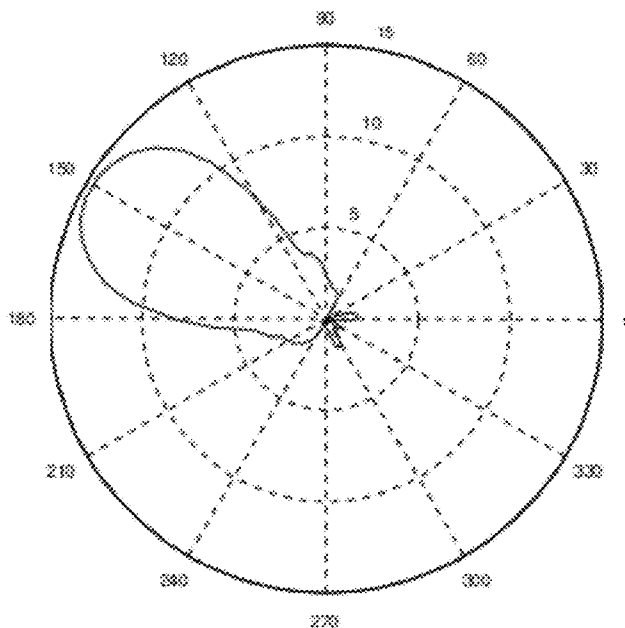
FIG. 16H
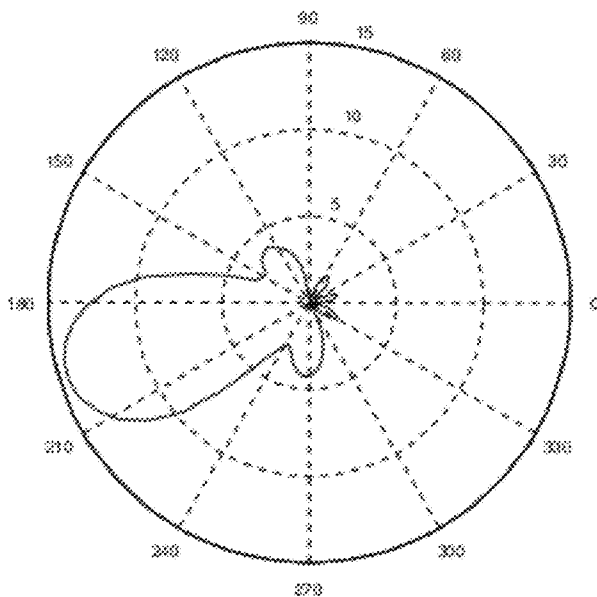
FIG. 16I
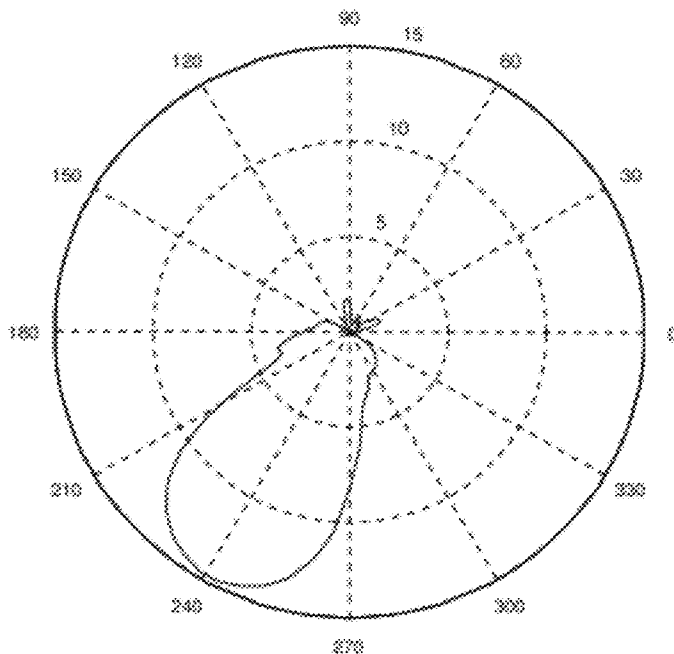
FIG. 16J
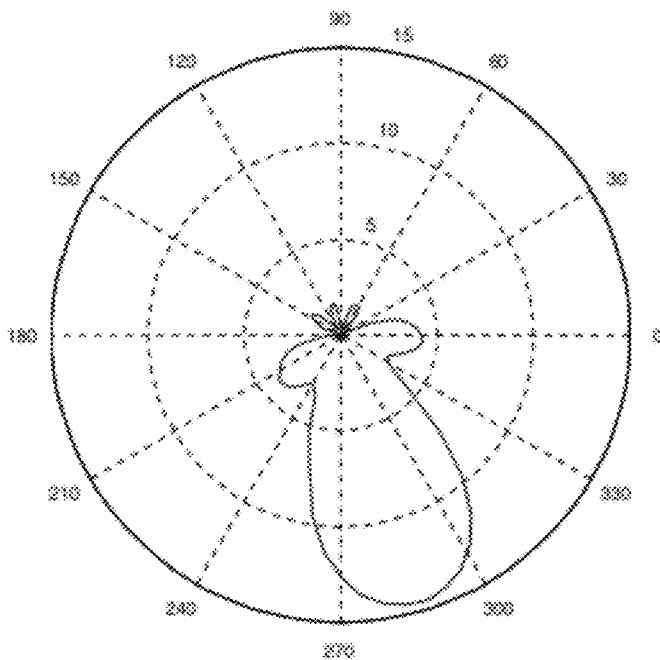
FIG. 16K
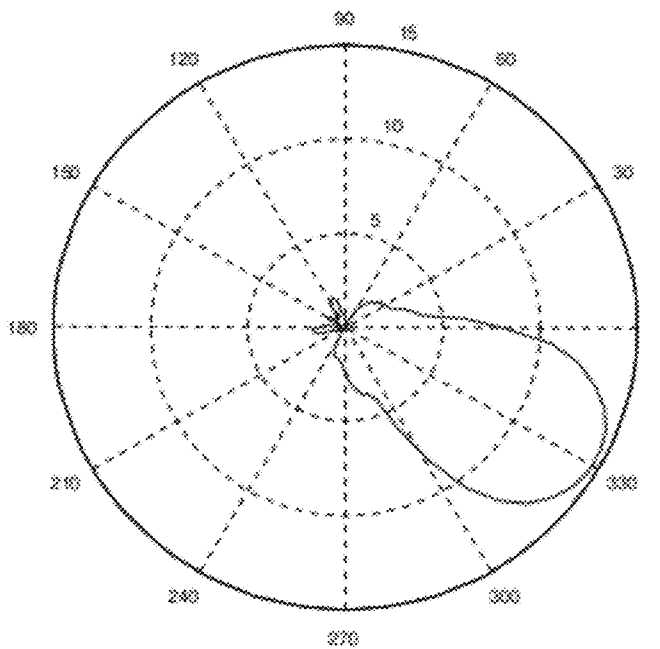
FIG. 16L
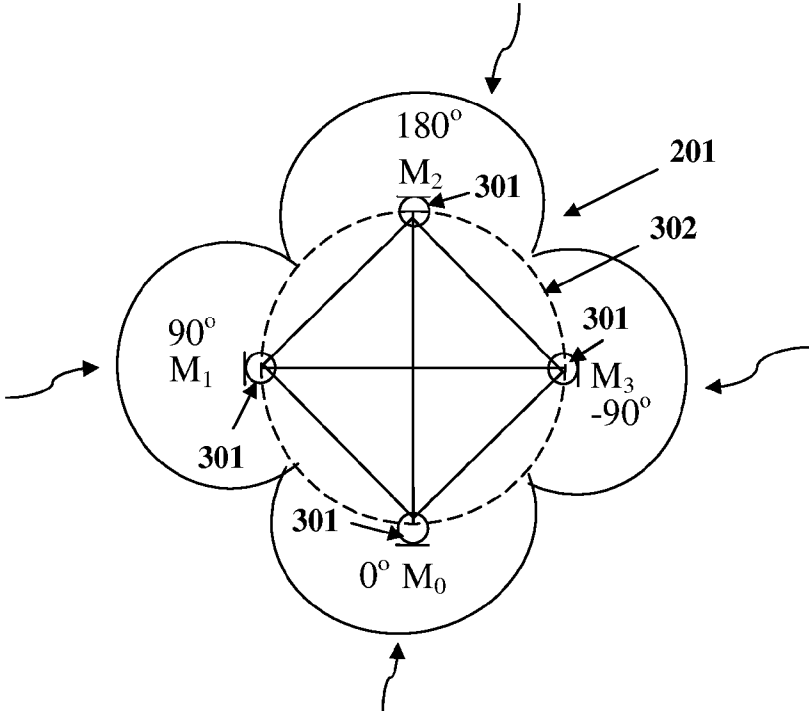
FIG. 17A
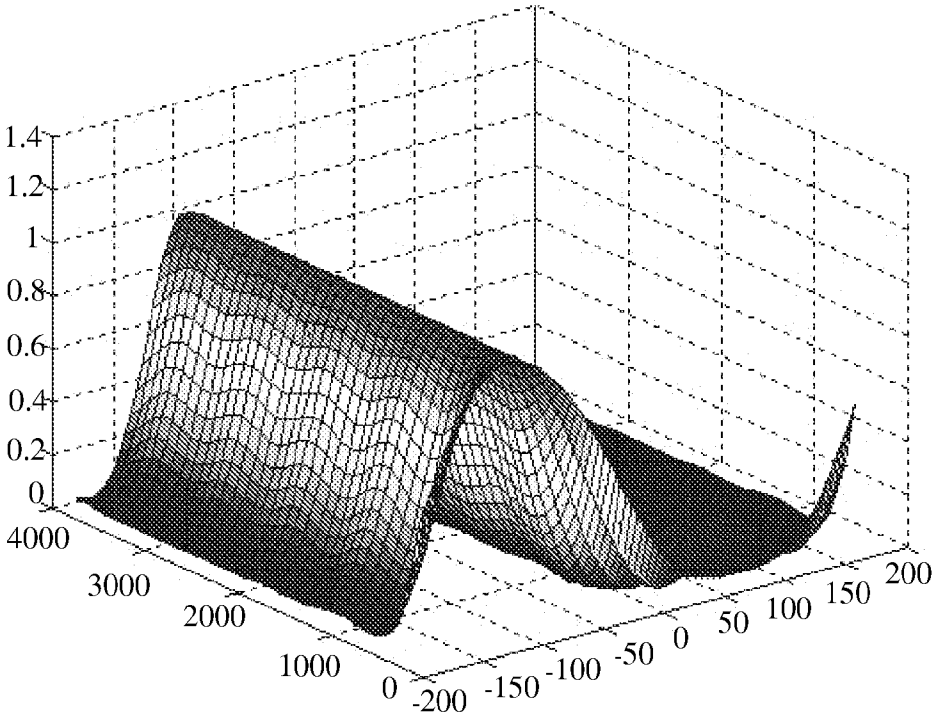
FIG. 17B
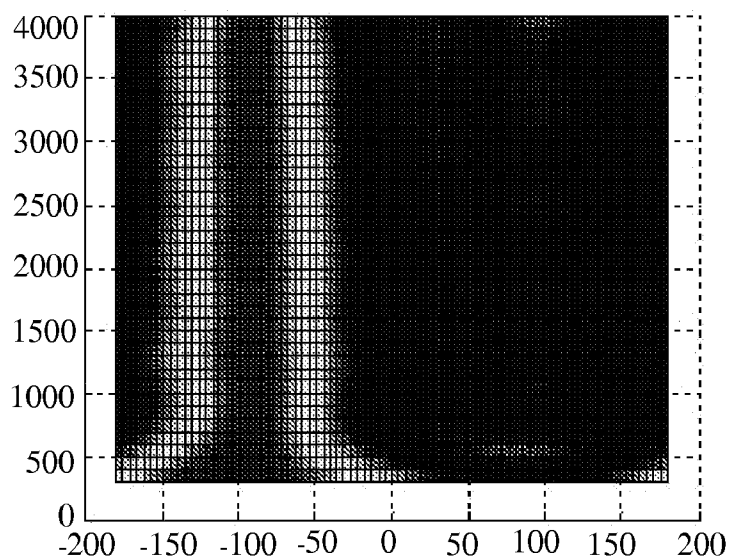
FIG. 17C
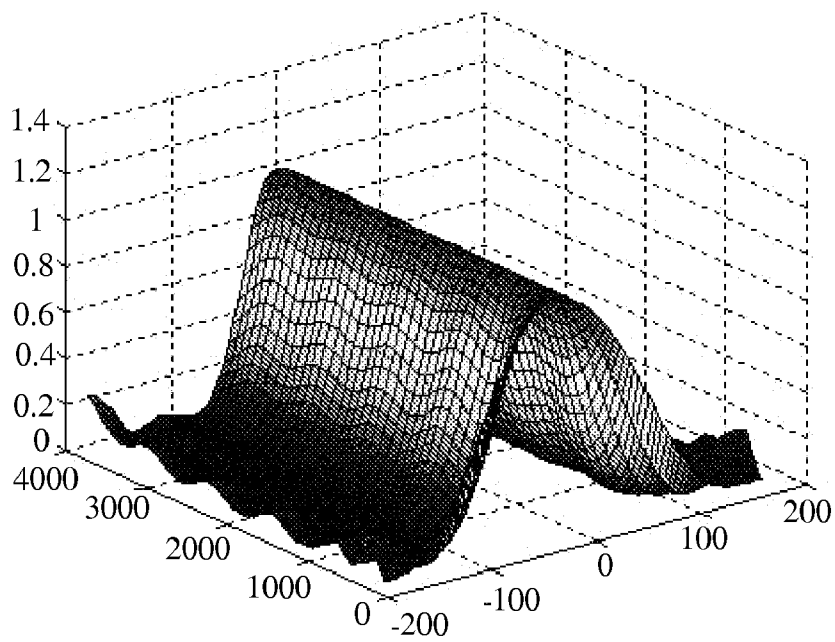
FIG. 17D
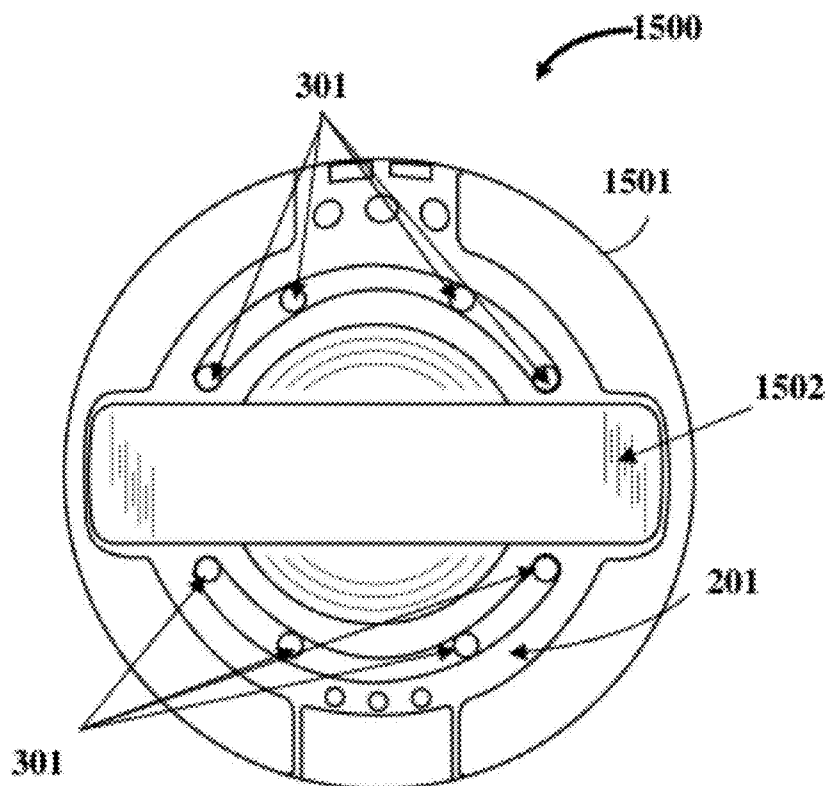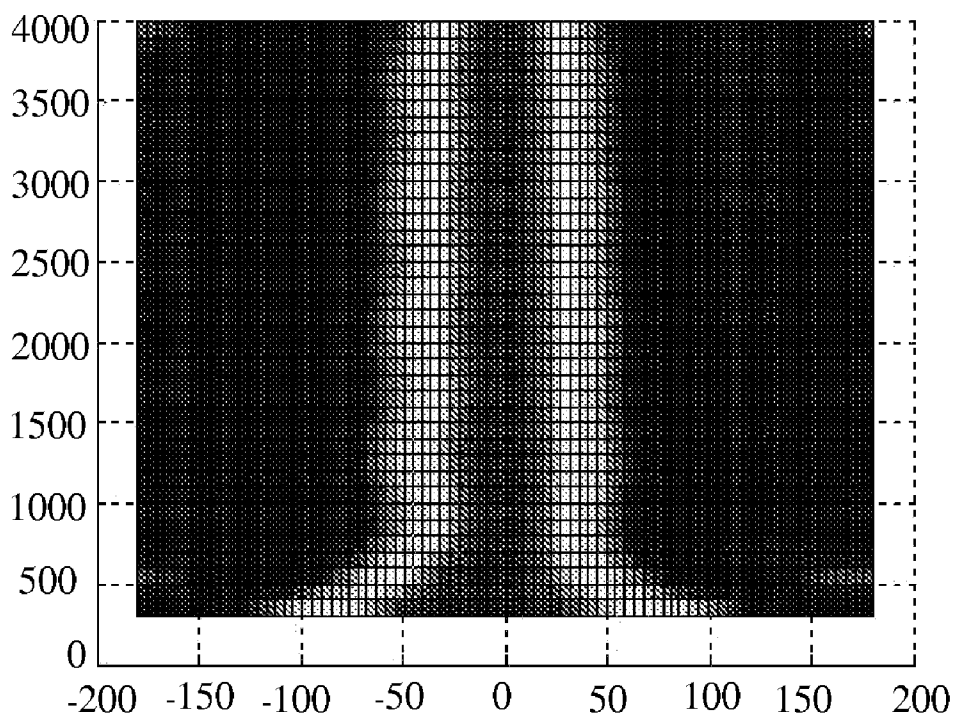
FIG. 17E
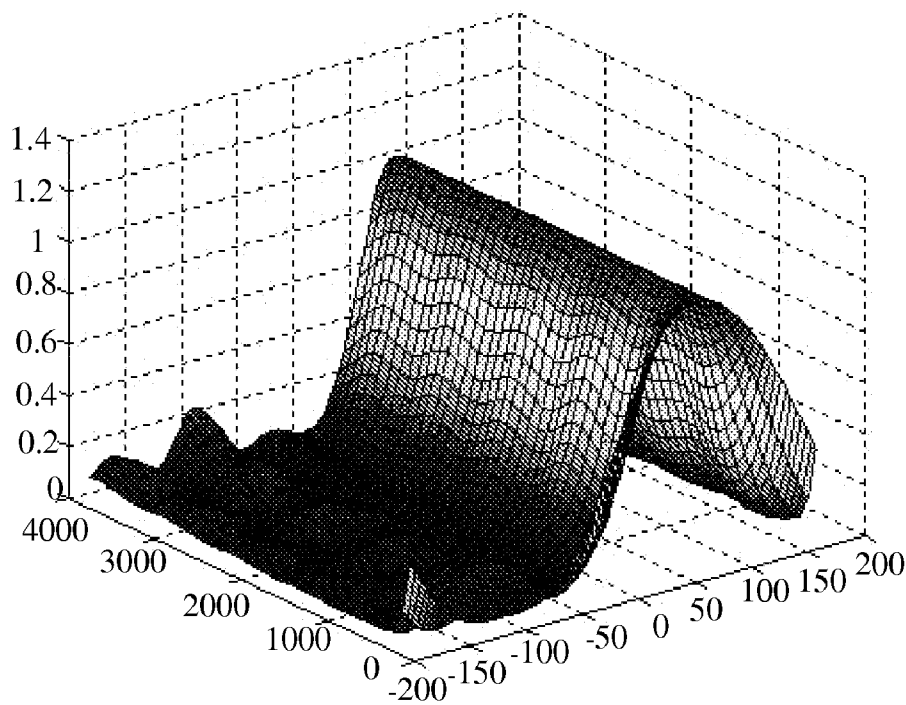
FIG. 17F
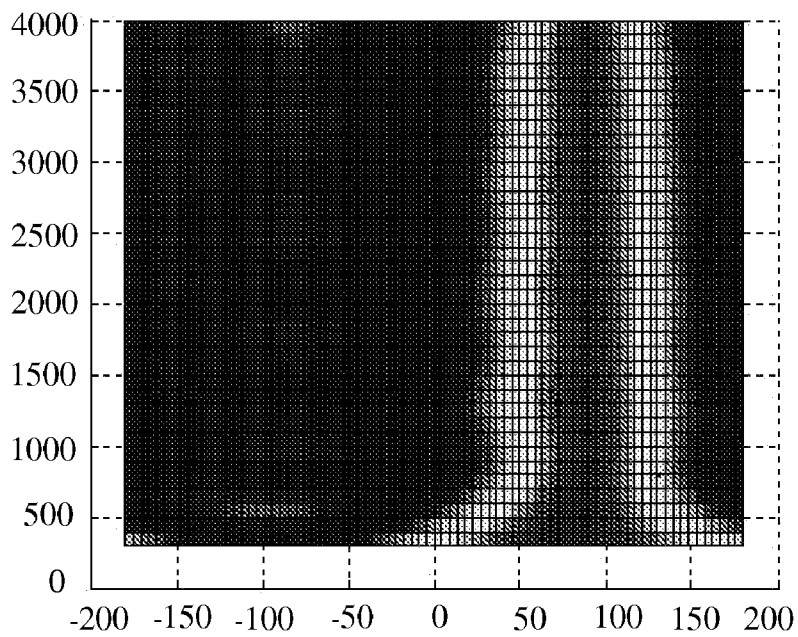
FIG. 17G
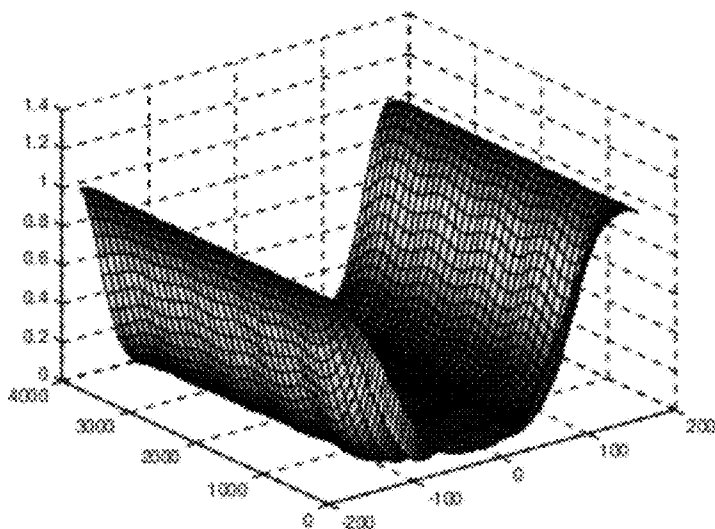
FIG. 17H
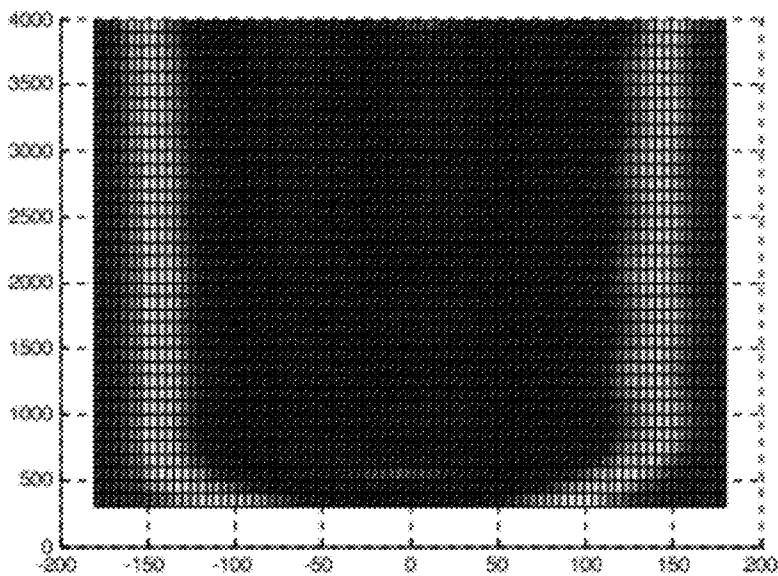
FIG. 17I

FIG. 18A

FIG. 18B

301

301

201

1801

1802
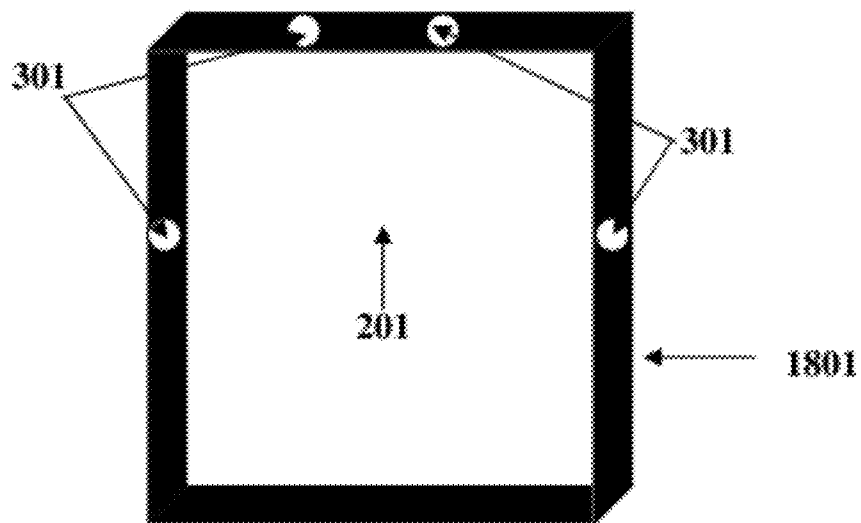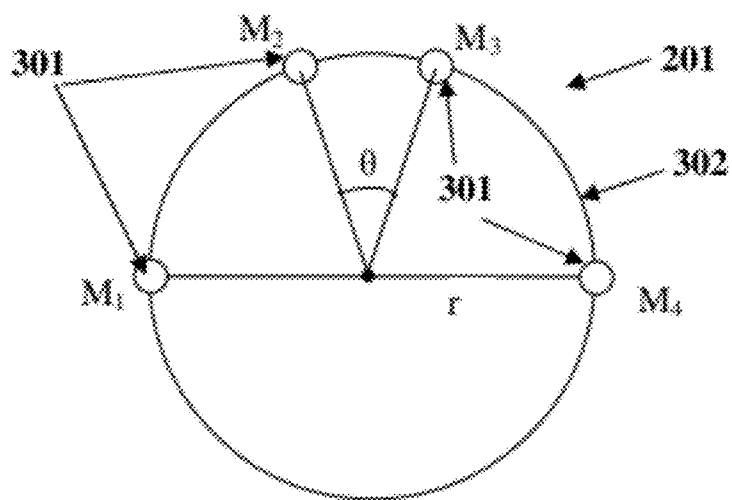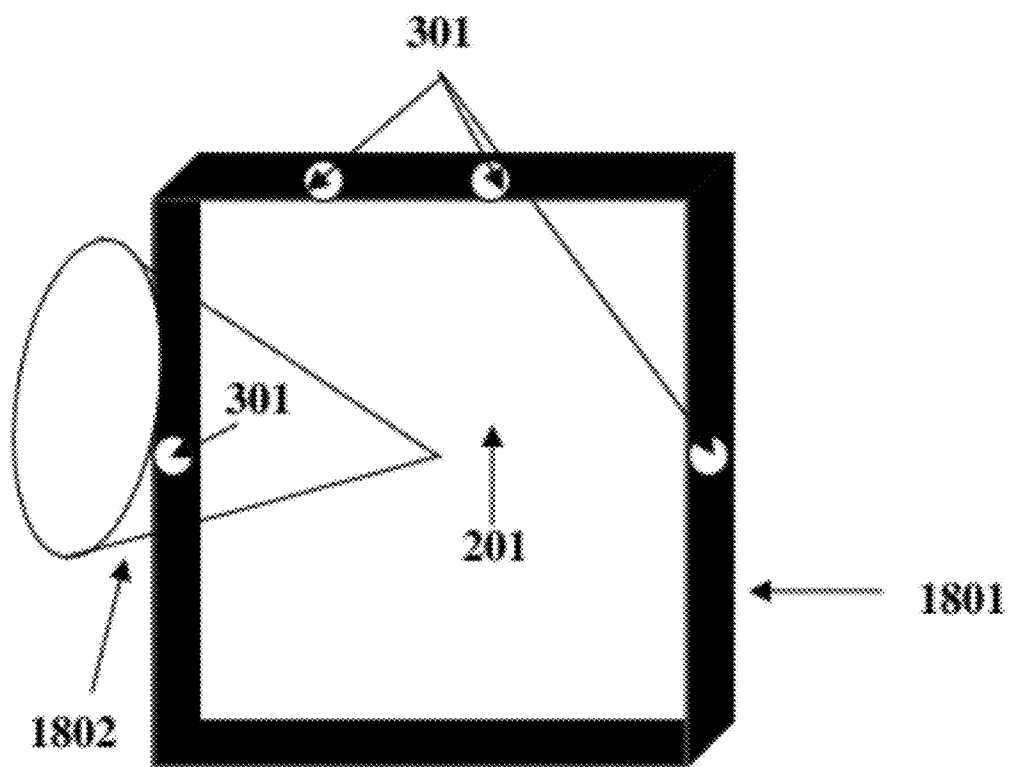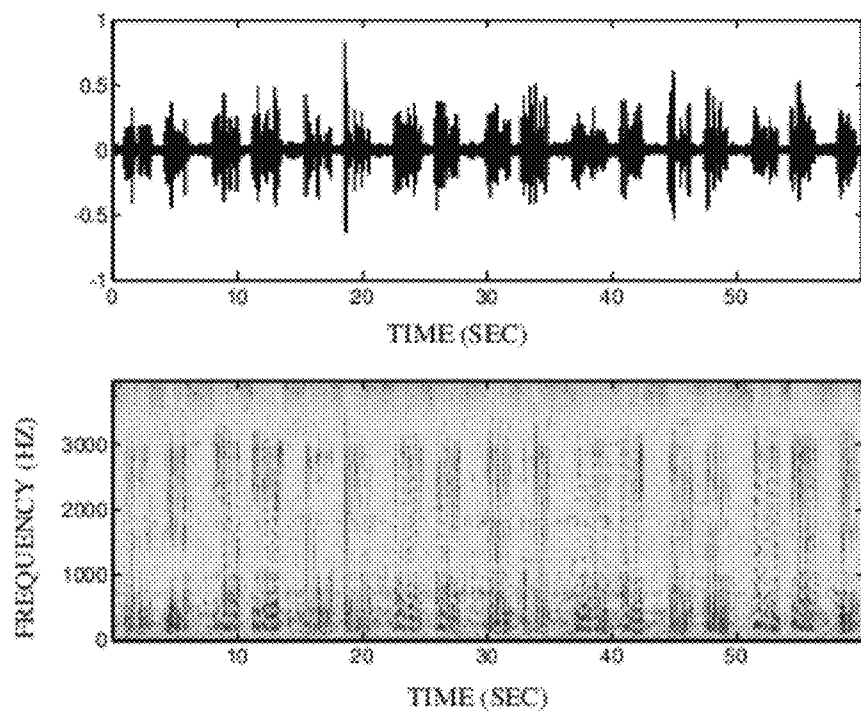
FIG. 18C

FIG. 18D

FIG. 18E

FIG. 18F

FIG. 18G

$$\tau_1 = D*\cos(\theta+\Phi_1)$$
$$\tau_2 = d*\cos(\theta+\Phi_2)$$
$$\tau_3 = d*\cos(\theta-\Phi_3)$$
$$\tau_4 = D*\cos(\theta-\Phi_4)$$
$$\tau_5 = -D*\cos(\theta-\Phi_5)$$
$$\tau_6 = -d*\cos(\theta-\Phi_6)$$
$$\tau_7 = -d*\cos(\theta+\Phi_7)$$
$$\tau_8 = -D*\cos(\theta+\Phi_8)$$

FIG. 19A

$$\tau_1 = D*\cos(\theta+\Phi_{12})$$
$$\tau_2 = d*\cos(\theta+\Phi_{12})$$
$$\tau_3 = d*\cos(\theta-\Phi_{34})$$
$$\tau_4 = D*\cos(\theta-\Phi_{34})$$
$$\tau_5 = -D*\cos(\theta-\Phi_{56})$$
$$\tau_6 = -d*\cos(\theta-\Phi_{56})$$
$$\tau_7 = -d*\cos(\theta+\Phi_{78})$$
$$\tau_8 = -D*\cos(\theta+\Phi_{78})$$

FIG. 19B

$\tau_1=-D*\cos(\theta)$

$\tau_2=-D*\sin(\theta)$

$\tau_3=D*\cos(\theta)$

$\tau_4=D*\sin(\theta)$

$\tau_5=-d*\cos(\theta-\Phi_5)$

$\tau_6=d*\cos(\theta+\Phi_6)$

$\tau_7=d*\cos(\theta-\Phi_7)$

$\tau_8=-d*\cos(\theta+\Phi_8)$

FIG. 19C

$\tau_1=-D*\cos(\theta)$

$\tau_2=-D*\sin(\theta)$

$\tau_3=D*\cos(\theta)$

$\tau_4=D*\sin(\theta)$

$\tau_5=-d*\cos(\theta-\Phi_5)$

$\tau_6=d*\cos(\theta+\Phi_6)$

$\tau_7=d*\cos(\theta-\Phi_7)$

$\tau_8=-d*\cos(\theta+\Phi_8)$

FIG. 19D

$\tau_1 = -d1 * \sin(\theta)$

$\tau_2 = -d2 * \sin(\theta)$

$\tau_3 = d3 * \cos(\theta)$

$\tau_4 = d4 * \cos(\theta)$

FIG. 19E

$\tau_1 = d * \cos(\theta + \Phi_1)$

$\tau_2 = d * \cos(\theta - \Phi_2)$

$\tau_3 = -d * \cos(\theta - \Phi_3)$

$\tau_4 = -d * \cos(\theta + \Phi_4)$
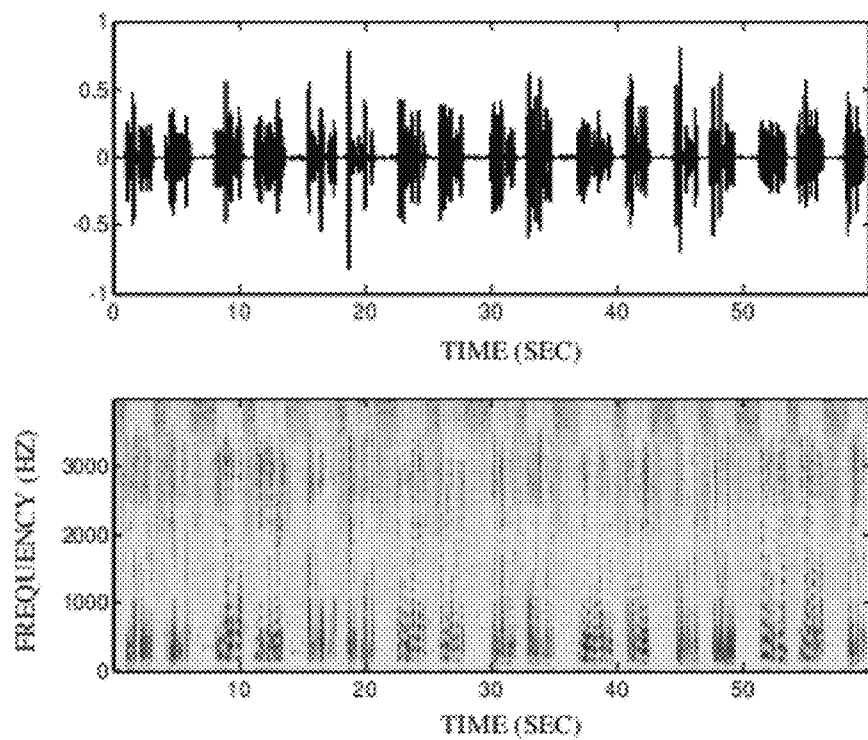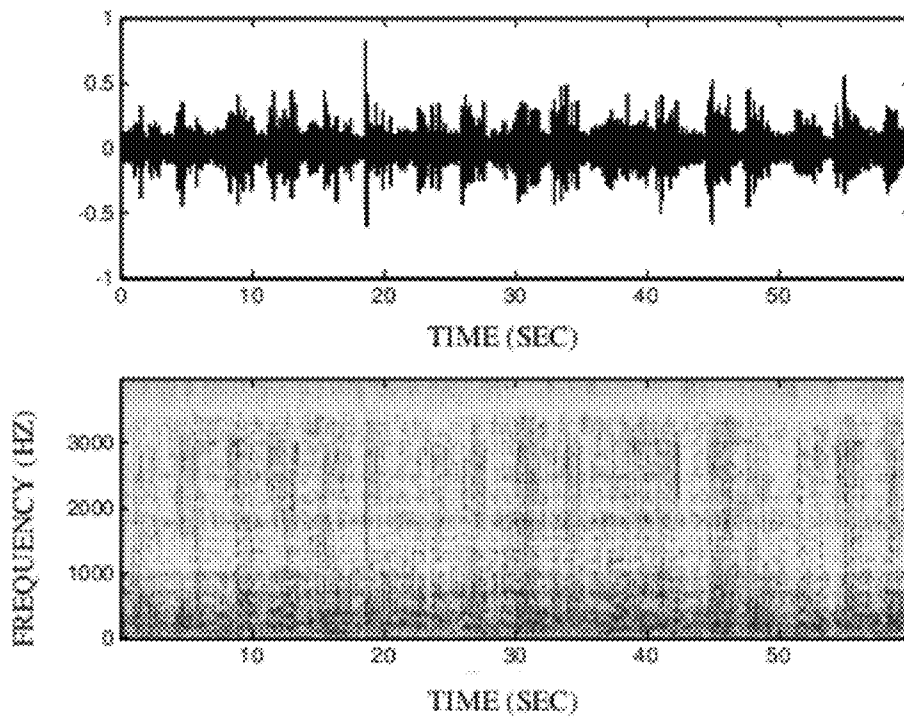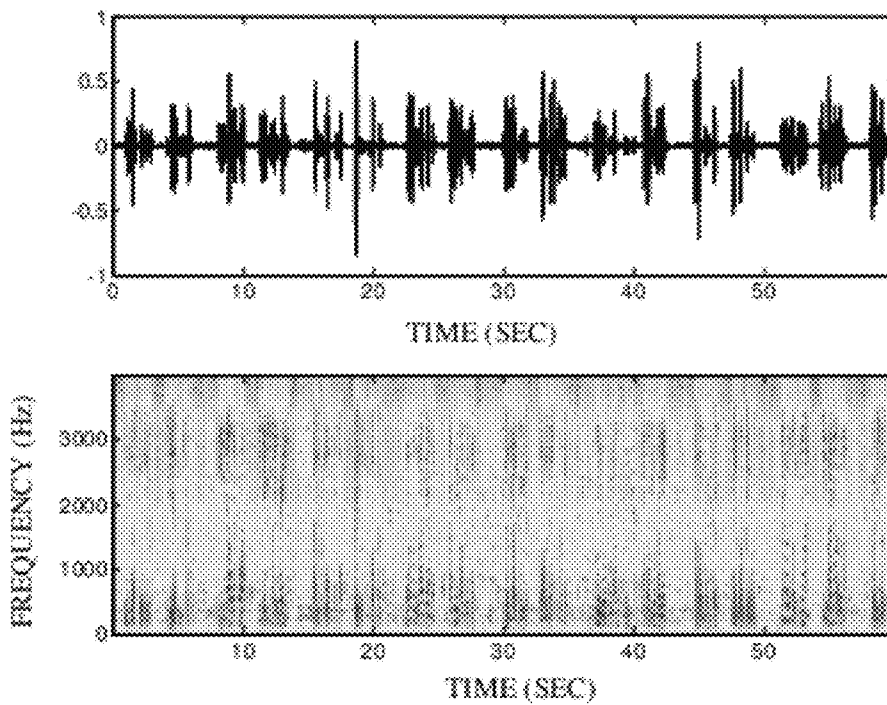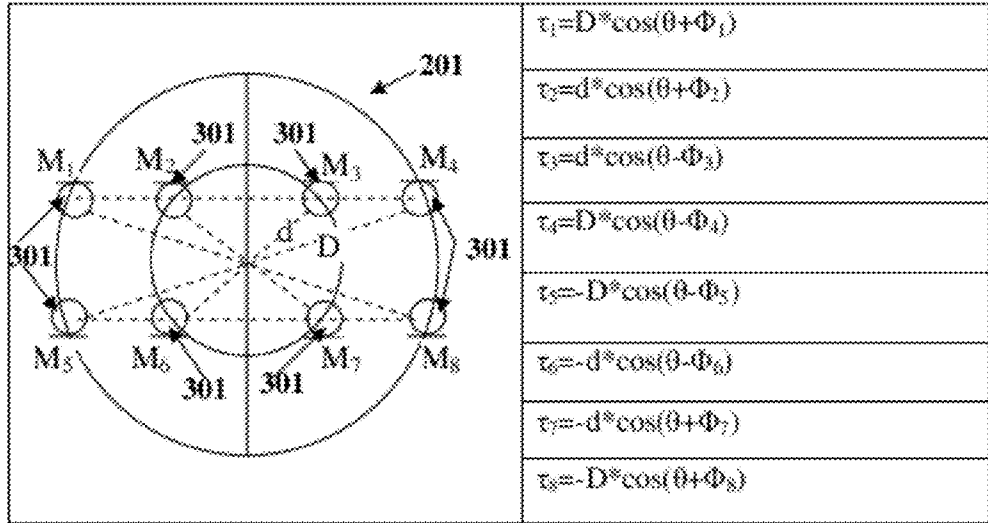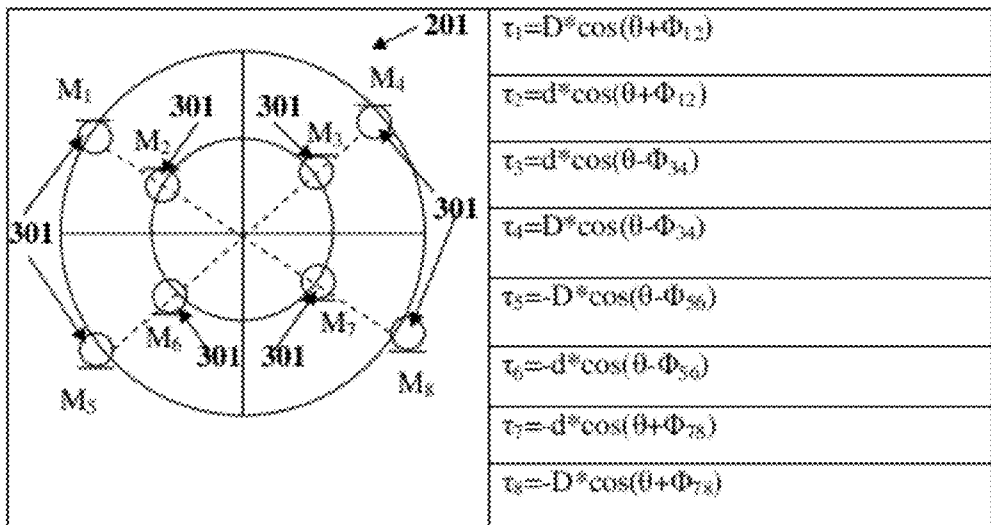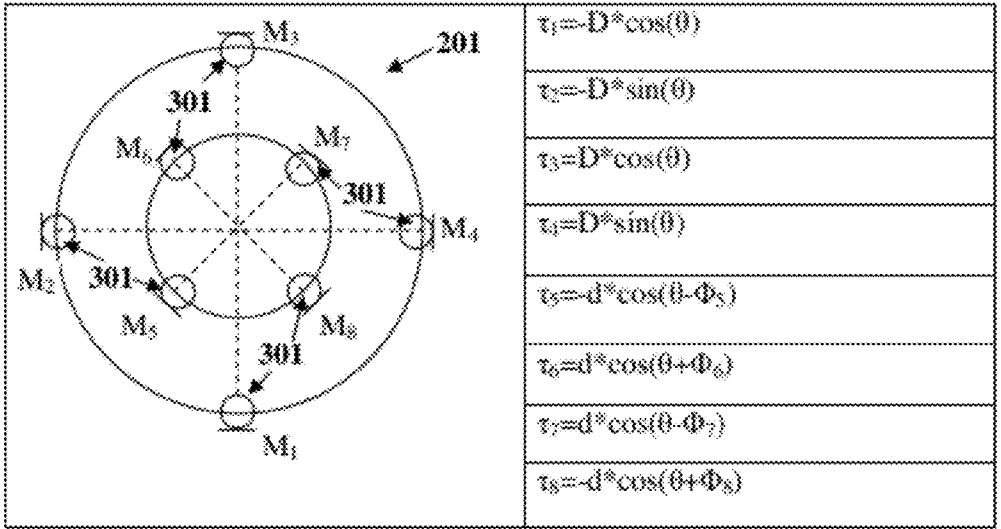
FIG. 19F

MICROPHONE ARRAY SYSTEM

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of provisional patent application No. 61/403,952 titled "Microphone array design and implementation for telecommunications and handheld devices", filed on Sep. 24, 2010 in the United States Patent and Trademark Office.

[0002] The specification of the above referenced patent application is incorporated herein by reference in its entirety.

BACKGROUND

[0003] Microphones constitute an important element in today's speech acquisition devices. Currently, most of the hands-free speech acquisition devices, for example, mobile devices, lapels, headsets, etc., convert sound into electrical signals by using a microphone embedded within the speech acquisition device. However, the paradigm of a single microphone often does not work effectively because the microphone picks up many ambient noise signals in addition to the desired sound, specifically when the distance between a user and the microphone is more than a few inches. Therefore, there is a need for a microphone system that operates under a variety of different ambient noise conditions and that places fewer constraints on the user with respect to the microphone, thereby eliminating the need to wear the microphone or be in close proximity to the microphone.

[0004] To mitigate the drawbacks of the single microphone system, there is a need for a microphone array that achieves directional gain in a preferred spatial direction while suppressing ambient noise from other directions. Conventional microphone arrays include arrays that are typically developed for applications such as radar and sonar, but are generally not suitable for hands-free or handheld speech acquisition devices. The main reason is that the desired sound signal has an extremely wide bandwidth relative to its center frequency, thereby rendering conventional narrowband techniques employed in the conventional microphone arrays unsuitable. In order to cater to such broadband speech applications, the array size needs to be vastly increased, making the conventional microphone arrays large and bulky, and precluding the conventional microphone arrays from having broader applications, for example, in mobile and handheld communication devices. There is a need for a microphone array system that provides an effective response over a wide spectrum of frequencies while being unobtrusive in terms of size.

[0005] Hence, there is a long felt but unresolved need for a broadband microphone array and broadband beamforming system that enhances acoustics of a desired sound signal while suppressing ambient noise signals.

SUMMARY OF THE INVENTION

[0006] This summary is provided to introduce a selection of concepts in a simplified form that are further described in the detailed description of the invention. This summary is not intended to identify key or essential inventive concepts of the claimed subject matter, nor is it intended for determining the scope of the claimed subject matter.

[0007] The method and system disclosed herein addresses the above stated need for enhancing acoustics of a target sound signal received from a target sound source, while suppressing ambient noise signals. As used herein, the term "tar-get sound signal" refers to a sound signal from a desired or target sound source, for example, a person's speech that needs to be enhanced. A microphone array system comprising an array of sound sensors positioned in an arbitrary configuration, a sound source localization unit, an adaptive beamforming unit, and a noise reduction unit, is provided. The sound source localization unit, the adaptive beamforming unit, and the noise reduction unit are in operative communication with the array of sound sensors. The array of sound sensors is, for example, a linear array of sound sensors, a circular array of sound sensors, or an arbitrarily distributed coplanar array of sound sensors. The array of sound sensors herein referred to as a "microphone array" receives sound signals from multiple disparate sound sources. The method disclosed herein can be applied on a microphone array with an arbitrary number of sound sensors having, for example, an arbitrary two dimensional (2D) configuration. The sound signals received by the sound sensors in the microphone array comprise the target sound signal from the target sound source among the disparate sound sources, and ambient noise signals.

[0008] The sound source localization unit estimates a spatial location of the target sound signal from the received sound signals, for example, using a steered response power-phase transform. The adaptive beamforming unit performs adaptive beamforming for steering a directivity pattern o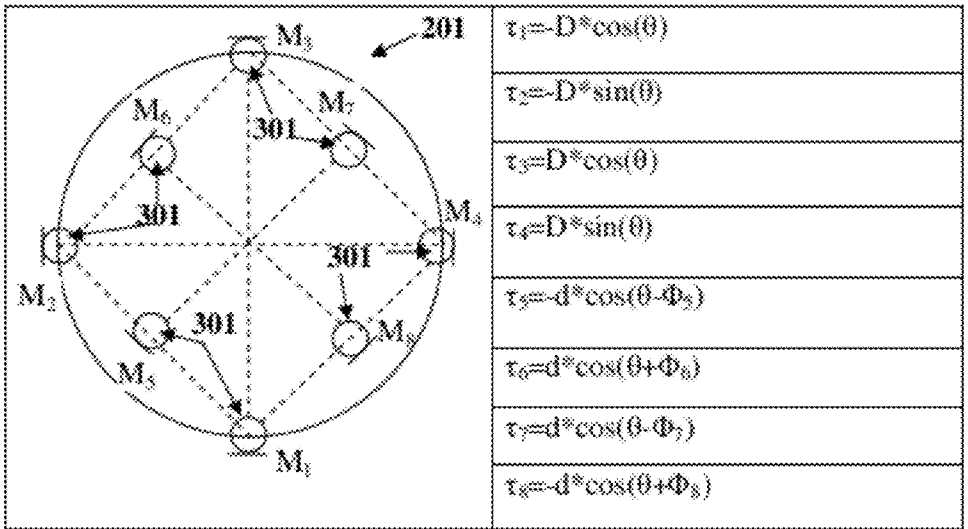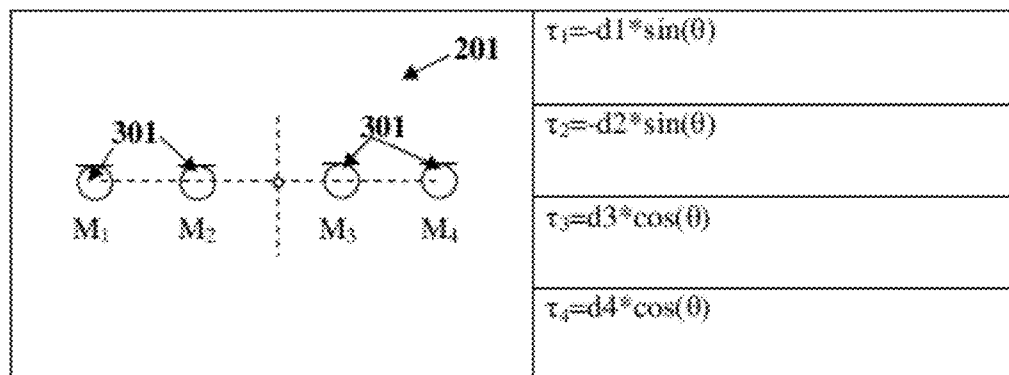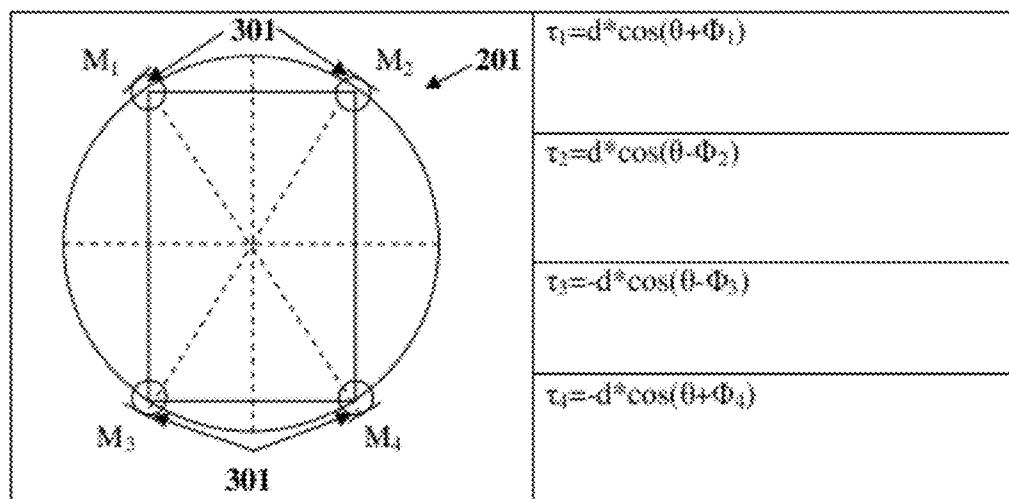f the microphone array in a direction of the spatial location of the target sound signal. The adaptive beamforming unit thereby enhances the target sound signal from the target sound source and partially suppresses the ambient noise signals. The noise reduction unit suppresses the ambient noise signals for further enhancing the target sound signal received from the target sound source.

[0009] In an embodiment where the target sound source that emits the target sound signal is in a two dimensional plane, a delay between each of the sound sensors and an origin of the microphone array is determined as a function of distance between each of the sound sensors and the origin, a predefined angle between each of the sound sensors and a reference axis, and an azimuth angle between the reference axis and the target sound signal. In another embodiment where the target sound source that emits the target sound signal is in a three dimensional plane, the delay between each of the sound sensors and the origin of the microphone array is determined as a function of distance between each of the sound sensors and the origin, a predefined angle between each of the sound sensors and a first reference axis, an elevation angle between a second reference axis and the target sound signal, and an azimuth angle between the first reference axis and the target sound signal. This method of determining the delay enables beamforming for arbitrary numbers of sound sensors and multiple arbitrary microphone array configurations. The delay is determined, for example, in terms of number of samples. Once the delay is determined, the microphone array can be aligned to enhance the target sound signal from a specific direction.

[0010] The adaptive beamforming unit comprises a fixed beamformer, a blocking matrix, and an adaptive filter. The fixed beamformer steers the directivity pattern of the microphone array in the direction of the spatial location of the target sound signal from the target sound source for enhancing the target sound signal, when the target sound source is in motion. The blocking matrix feeds the ambient noise signals to the adaptive filter by blocking the target sound signal from the target sound source. The adaptive filter adaptively filters the

2

ambient noise signals in response to detecting the presence or absence of the target sound signal in the sound signals received from the disparate sound sources. The fixed beamformer performs fixed beamforming, for example, by filtering and summing output sound signals from the sound sensors.

[0011] In an embodiment, the adaptive filtering comprises sub-band adaptive filtering. The adaptive filter comprises an analysis filter bank, an adaptive filter matrix, and a synthesis filter bank. The analysis filter bank splits the enhanced target sound signal from the fixed beamformer and the ambient noise signals from the blocking matrix into multiple frequency sub-bands. The adaptive filter matrix adaptively filters the ambient noise signals in each of the frequency sub-bands in response to detecting the presence or absence of the target sound signal in the sound signals received from the disparate sound sources. The synthesis filter bank synthesizes a full-band sound signal using the frequency sub-bands of the enhanced target sound signal. In an embodiment, the adaptive beamforming unit further comprises an adaptation control unit for detecting the presence of the target sound signal and adjusting a step size for the adaptive filtering in response to detecting the presence or the absence of the target sound signal in the sound signals received from the disparate sound sources.

[0012] The noise reduction unit suppresses the ambient noise signals for further enhancing the target sound signal from the target sound source. The noise reduction unit performs noise reduction, for example, by using a Wiener-filter based noise reduction algorithm, a spectral subtraction noise reduction algorithm, an auditory transform based noise reduction algorithm, or a model based noise reduction algorithm. The noise reduction unit performs noise reduction in multiple frequency sub-bands employed for sub-band adaptive beamforming by the analysis filter bank of the adaptive beamforming unit.

[0013] The microphone array system disclosed herein comprising the microphone array with an arbitrary number of sound sensors positioned in arbitrary configurations can be implemented in handheld devices, for example, the iPad® of Apple Inc., the iPhone® of Apple Inc., smart phones, tablet computers, laptop computers, etc. The microphone array system disclosed herein can further be implemented in conference phones, video conferencing applications, or any device or equipment that needs better speech inputs.

BRIEF DESCRIPTION OF THE DRAWINGS

[0014] The foregoing summary, as well as the following detailed description of the invention, is better understood when read in conjunction with the appended drawings. For the purpose of illustrating the invention, exemplary constructions of the invention are shown in the drawings. However, the invention is not limited to the specific methods and instrumentalities disclosed herein.

[0015] FIG. 1 illustrates a method for enhancing a target sound signal from multiple sound signals.

[0016] FIG. 2 illustrates a system for enhancing a target sound signal from multiple sound signals.

[0017] FIG. 3 exemplarily illustrates a microphone array configuration showing a microphone array having N sound sensors arbitrarily distributed on a circle.

[0018] FIG. 4 exemplarily illustrates a graphical representation of a filter-and-sum beamforming algorithm for determining output of the microphone array having N sound sensors.

[0019] FIG. 5 exemplarily illustrates distances between an origin of the microphone array and sound sensor M_1 and sound sensor M_3 in the circular microphone array configuration, when the target sound signal is at an angle θ from the Y-axis.

[0020] FIG. 6A exemplarily illustrates a table showing the distance between each sound sensor in a circular microphone array configuration from the origin of the microphone array, when the target sound source is in the same plane as that of the microphone array.

[0021] FIG. 6B exemplarily illustrates a table showing the relationship of the position of each sound sensor in the circular microphone array configuration and its distance to the origin of the microphone array, when the target sound source is in the same plane as that of the microphone array.

[0022] FIG. 7A exemplarily illustrates a graphical representation of a microphone array, when the target sound source is in a three dimensional plane.

[0023] FIG. 7B exemplarily illustrates a table showing delay between each sound sensor in a circular microphone array configuration and the origin of the microphone array, when the target sound source is in a three dimensional plane.

[0024] FIG. 7C exemplarily illustrates a three dimensional working space of the microphone array, where the target sound signal is incident at an elevation angle $\Psi < \Omega$

[0025] FIG. 8 exemplarily illustrates a method for estimating a spatial location of the target sound signal from the target sound source by a sound source localization unit using a steered response power-phase transform.

[0026] FIG. 9A exemplarily illustrates a graph showing the value of the steered response power-phase transform for every 10°.

[0027] FIG. 9B exemplarily illustrates a graph representing the estimated target sound signal from the target sound source.

[0028] FIG. 10 exemplarily illustrates a system for performing adaptive beamforming by an adaptive beamforming unit.

[0029] FIG. 11 exemplarily illustrates a system for sub-band adaptive filtering.

[0030] FIG. 12 exemplarily illustrates a graphical representation showing the performance of a perfect reconstruction filter bank.

[0031] FIG. 13 exemplarily illustrates a block diagram of a noise reduction unit that performs noise reduction using a Wiener-filter based noise reduction algorithm.

[0032] FIG. 14 exemplarily illustrates a hardware implementation of the microphone array system.

[0033] FIGS. 15A-15C exemplarily illustrate a conference phone comprising an eight-sensor microphone array.

[0034] FIG. 16A exemplarily illustrates a layout of an eight-sensor microphone array for a conference phone.

[0035] FIG. 16B exemplarily illustrates a graphical representation of eight spatial regions to which the eight-sensor microphone array of FIG. 16A responds.

[0036] FIGS. 16C-16D exemplarily illustrate computer simulations showing the steering of the directivity patterns of the eight-sensor microphone array of FIG. 16A in the directions of 15° and 60° respectively, in the frequency range 300 Hz to 5 kHz.

[0037] FIGS. 16E-16L exemplarily illustrate graphical representations showing the directivity patterns of the eight-sensor microphone array of FIG. 16A in each of the eight

spatial regions, where each directivity pattern is an average response from 300 Hz to 5000 Hz.

[0038] FIG. 17A exemplarily illustrates a graphical representation of four spatial regions to which a four-sensor microphone array for a wireless handheld device responds.

[0039] FIGS. 17B-17I exemplarily illustrate computer simulations showing the directivity patterns of the four-sensor microphone array of FIG. 17A with respect to azimuth and frequency.

[0040] FIGS. 18A-18B exemplarily illustrate a microphone array configuration for a tablet computer.

[0041] FIG. 18C exemplarily illustrates an acoustic beam formed using the microphone array configuration of FIGS. 18A-18B according to the method and system disclosed herein.

[0042] FIGS. 18D-18G exemplarily illustrate graphs showing processing results of the adaptive beamforming unit and the noise reduction unit for the microphone array configuration of FIG. 18B, in both a time domain and a spectral domain for the tablet computer.

[0043] FIGS. 19A-19F exemplarily illustrate tables showing different microphone array configurations and the corresponding values of delay τ_n, for the sound sensors in each of the microphone array configurations.

DETAILED DESCRIPTION OF THE INVENTION

[0044] FIG. 1 illustrates a method for enhancing a target sound signal from multiple sound signals. As used herein, the term "target sound signal" refers to a desired sound signal from a desired or target sound source, for example, a person's speech that needs to be enhanced. The method disclosed herein provides 101a microphone array system comprising an array of sound sensors positioned in an arbitrary configuration, a sound source localization unit, an adaptive beamforming unit, and a noise reduction unit. The sound source localization unit, the adaptive beamforming unit, and the noise reduction unit are in operative communication with the array of sound sensors. The microphone array system disclosed herein employs the array of sound sensors positioned in an arbitrary configuration, the sound source localization unit, the adaptive beamforming unit, and the noise reduction unit for enhancing a target sound signal by acoustic beam forming in the direction of the target sound signal in the presence of ambient noise signals.

[0045] The array of sound sensors herein referred to as a "microphone array" comprises multiple or an arbitrary number of sound sensors, for example, microphones, operating in tandem. The microphone array refers to an array of an arbitrary number of sound sensors positioned in an arbitrary configuration. The sound sensors are transducers that detect sound and convert the sound into electrical signals. The sound sensors are, for example, condenser microphones, piezoelectric microphones, etc.

[0046] The sound sensors receive 102 sound signals from multiple disparate sound sources and directions. The target sound source that emits the target sound signal is one of the disparate sound sources. As used herein, the term "sound signals" refers to composite sound energy from multiple disparate sound sources in an environment of the microphone array. The sound signals comprise the target sound signal from the target sound source and the ambient noise signals. The sound sensors are positioned in an arbitrary planar configuration herein referred to as a "microphone array configuration", for example, a linear configuration, a circular configuration, any arbitrarily distributed coplanar array configuration, etc. By employing beamforming according to the method disclosed herein, the microphone array provides a

higher response to the target sound signal received from a particular direction than to the sound signals from other directions. A plot of the response of the microphone array versus frequency and direction of arrival of the sound signals is referred to as a directivity pattern of the microphone array.

[0047] The sound source localization unit estimates 103 a spatial location of the target sound signal from the received sound signals. In an embodiment, the sound source localization unit estimates the spatial location of the target sound signal from the target sound source, for example, using a steered response power-phase transform as disclosed in the detailed description of FIG. 8.

[0048] The adaptive beamforming unit performs adaptive beamforming 104 by steering the directivity pattern of the microphone array in a direction of the spatial location of the target sound signal, thereby enhancing the target sound signal, and partially suppressing the ambient noise signals. Beamforming refers to a signal processing technique used in the microphone array for directional signal reception, that is, spatial filtering. This spatial filtering is achieved by using adaptive or fixed methods. Spatial filtering refers to separating two signals with overlapping frequency content that originate from different spatial locations.

[0049] The noise reduction unit performs noise reduction by further suppressing 105 the ambient noise signals and thereby further enhancing the target sound signal. The noise reduction unit performs the noise reduction, for example, by using a Wiener-filter based noise reduction algorithm, a spectral subtraction noise reduction algorithm, an auditory transform based noise reduction algorithm, or a model based noise reduction algorithm.

[0050] FIG. 2 illustrates a system 200 for enhancing a target sound signal from multiple sound signals. The system 200, herein referred to as a "microphone array system", comprises the array 201 of sound sensors positioned in an arbitrary configuration, the sound source localization unit 202, the adaptive beamforming unit 203, and the noise reduction unit 207.

[0051] The array 201 of sound sensors, herein referred to as the "microphone array" is in operative communication with the sound source localization unit 202, the adaptive beamforming unit 203, and the noise reduction unit 207. The microphone array 201 is, for example, a linear array of sound sensors, a circular array of sound sensors, or an arbitrarily distributed coplanar array of sound sensors. The microphone array 201 achieves directional gain in any preferred spatial direction and frequency band while suppressing signals from other spatial directions and frequency bands. The sound sensors receive the sound signals comprising the target sound signal and ambient noise signals from multiple disparate sound sources, where one of the disparate sound sources is the target sound source that emits the target sound signal.

[0052] The sound source localization unit 202 estimates the spatial location of the target sound signal from the received sound signals. In an embodiment, the sound source localization unit 202 uses, for example, a steered response power-phase transform, for estimating the spatial location of the target sound signal from the target sound source.

[0053] The adaptive beamforming unit 203 steers the directivity pattern of the microphone array 201 in a direction of the spatial location of the target sound signal, thereby enhancing the target sound signal and partially suppressing the ambient noise signals. The adaptive beamforming unit 203 comprises a fixed beamformer 204, a blocking matrix 205, and an adaptive filter 206 as disclosed in the detailed description of FIG. 10. The fixed beamformer 204 performs fixed beamforming by filtering and summing output sound signals from each of

4

the sound sensors in the microphone array **201** as disclosed in the detailed description of FIG. **4**. In an embodiment, the adaptive filter **206** is implemented as a set of sub-band adaptive filters. The adaptive filter **206** comprises an analysis filter bank **206a**, an adaptive filter matrix **206b**, and a synthesis filter bank **206c** as disclosed in the detailed description of FIG. **11**.

[0054] The noise reduction unit **207** further suppresses the ambient noise signals for further enhancing the target sound signal. The noise reduction unit **207** is, for example, a Wiener-filter based noise reduction unit, a spectral subtraction noise reduction unit, an auditory transform based noise reduction unit, or a model based noise reduction unit.

[0055] FIG. **3** exemplarily illustrates a microphone array configuration showing a microphone array **201** having N sound sensors **301** arbitrarily distributed on a circle **302** with a diameter "d", where "N" refers to the number of sound sensors **301** in the microphone array **201**. Consider an example where N=4, that is, there are four sound sensors **301** M_0, M_1, M_2, and M_3 in the microphone array **201**. Each of the sound sensors **301** is positioned at an acute angle "Φ_n" from a Y-axis, where $\Phi_1 \geqq 0$ and n=0, 1, 2, ... N–1. In an example, the sound sensor **301** M_0 is positioned at an acute angle Φ_0 from the Y-axis; the sound sensor **301** M_1 is positioned at an acute angle Φ_1 from the Y-axis; the sound sensor **301** M_2 is positioned at an acute angle Φ_2 from the Y-axis; and the sound sensor **301** M_3 is positioned at an acute angle Φ_3 from the Y-axis. A filter-and-sum beamforming algorithm determines the output "y" of the microphone array **201** having N sound sensors **301** as disclosed in the detailed description of FIG. **4**.

[0056] FIG. **4** exemplarily illustrates a graphical representation of the filter-and-sum beamforming algorithm for determining the output of the microphone array **201** having N sound sensors **301**. Consider an example where the target sound signal from the target sound source is at an angle θ with a normalized frequency w. The microphone array configuration is arbitrary in a two dimensional plane, for example, a circular array configuration where the sound sensors **301** M_0, M_1, M_2, ..., M_N, M_{N-1} of the microphone array **201** are arbitrarily positioned on a circle **302**. The sound signals received by each of the sound sensors **301** in the microphone array **201** are inputs to the microphone array **201**. The adaptive beamforming unit **203** employs the filter-and-sum beamforming algorithm that applies independent weights to each of the inputs to the microphone array **201** such that directivity pattern of the microphone array **201** is steered to the spatial location of the target sound signal as determined by the sound source localization unit **202**.

[0057] The output "y" of the microphone array **201** having N sound sensors **301** is the filter-and-sum of the outputs of the N sound sensors **301**. That is, $y = \sum_{n=0}^{N-1} w_n^T x_n$, where x_n is the output of the $(n+1)^{th}$ sound sensor **301**, and w_n^T denotes a transpose of a length-L filter applied to the $(n+1)^{th}$ sound sensor **301**.

[0058] The spatial directivity pattern H (ω, θ) for the target sound signal from angle θ with normalized frequency w is defined as:

$$H(\omega, \theta) = \frac{Y(\omega, \theta)}{\overline{X}(\omega, \theta)} = \frac{\sum_{n=0}^{N-1} W_n(\omega) X_n(\omega, \theta)}{\overline{X}(\omega, \theta)} \qquad (1)$$

where $\overline{X}$ is the signal received at the origin of the circular microphone array **201** and W is the frequency response of the real-valued finite impulse response (FIR) filter w. If the target

sound source is far enough away from the microphone array **201**, the difference between the signal received by the $(n+1)^{th}$ sound sensor **301** "x_n" and the origin of the microphone array **201** is a delay τ_n; that is, $X_n(\omega,\tau)=\overline{X}(\omega, \theta)e^{-j\omega\tau_n}$.

[0059] FIG. **5** exemplarily illustrates distances between an origin of the microphone array **201** and the sound sensor **301** M_1 and the sound sensor **301** M_3 in the circular microphone array configuration, when the target sound signal is at an angle θ from the Y-axis. The microphone array system **200** disclosed herein can be used with an arbitrary directivity pattern for arbitrarily distributed sound sensors **301**. For any specific microphone array configuration, the parameter that is defined to achieve beamformer coefficients is the value of delay τ_n for each sound sensor **301**. To define the value of τ_n, an origin or a reference point of the microphone array **201** is defined; and then the distance d_n between each sound sensor **301** and the origin is measured, and then the angle Φ_n of each sound sensor **301** biased from a vertical axis is measured.

[0060] For example, the angle between the Y-axis and the line joining the origin and the sound sensor **301** M_0 is Φ_0, the angle between the Y-axis and the line joining the origin and the sound sensor **301** M_1 is Φ_1, the angle between the Y-axis and the line joining the origin and the sound sensor **301** M_2 is Φ_2, and the angle between the Y-axis and the line joining the origin and the sound sensor **301** M_3 is Φ_3. The distance between the origin O and the sound sensor **301** M_1, and the origin O and the sound sensor **301** M_3 when the incoming target sound signal from the target sound source is at an angle θ from the Y-axis is denoted as τ_1 and τ_3, respectively.

[0061] For purposes of illustration, the detailed description refers to a circular microphone array configuration; however, the scope of the microphone array system **200** disclosed herein is not limited to the circular microphone array configuration but may be extended to include a linear array configuration, an arbitrarily distributed coplanar array configuration, or a microphone array configuration with any arbitrary geometry.

[0062] FIG. **6A** exemplarily illustrates a table showing the distance between each sound sensor **301** in a circular microphone array configuration from the origin of the microphone array **201**, when the target sound source is in the same plane as that of the microphone array **201**. The distance measured in meters and the corresponding delay (τ) measured in number of samples is exemplarily illustrated in FIG. **6A**. In an embodiment where the target sound source that emits the target sound signal is in a two dimensional plane, the delay (τ) between each of the sound sensors **301** and the origin of the microphone array **201** is determined as a function of distance (d) between each of the sound sensors **301** and the origin, a predefined angle (Φ) between each of the sound sensors **301** and a reference axis (Y) as exemplarily illustrated in FIG. **5**, and an azimuth angle (θ) between the reference axis (Y) and the target sound signal. The determined delay (τ) is represented in terms of number of samples.

[0063] If the target sound source is far enough from the microphone array **201**, the time delay between the signal received by the $(n+1)^{th}$ sound sensor **301** "x_n" and the origin of the microphone array **201** is herein denoted as "t" measured in seconds. The sound signals received by the microphone array **201**, which are in analog form are converted into digital sound signals by sampling the analog sound signals at a particular frequency, for example, 8000 Hz. That is, the number of samples in each second is 8000. The delay τ can be represented as the product of the sampling frequency (f_s) and

the time delay (t). That is, $\tau=f_s*t$. Therefore, the distance between the sound sensors **301** in the microphone array **201** corresponds to the time used for the target sound signal to travel the distance and is measured by the number of samples within that time period.

[0064] Consider an example where "d" is the radius of the circle **302** of the circular microphone array configuration, "f_s" is the sampling frequency, and "c" is the speed of sound. FIG. **6**B exemplarily illustrates a table showing the relationship of the position of each sound sensor **301** in the circular microphone array configuration and its distance to the origin of the microphone array **201**, when the target sound source is in the same plane as that of the microphone array **201**. The distance measured in meters and the corresponding delay (τ) measured in number of samples is exemplarily illustrated in FIG. **6**B.

[0065] The method of determining the delay (τ) enables beamforming for arbitrary numbers of sound sensors **301** and multiple arbitrary microphone array configurations. Once the delay (τ) is determined, the microphone array **201** can be aligned to enhance the target sound signal from a specific direction.

[0066] Therefore, the spatial directivity pattern H can be re-written as:

$$H(\omega,\theta)=\Sigma_{n=0}^{N-1}W_n(\omega)e^{-j\omega\tau_n(\theta)}=w^Tg(\omega,\theta) \qquad (2)$$

where $w^T=[w_0{}^T, w_1{}^T, w_2{}^T, w_3{}^T, \ldots, w_{N-1}{}^T]$ and $g(\omega,\theta)=\{g^i(\omega, \theta)\}_{i=1 \ldots NL}=\{e^{-j\omega(k+\tau_n(\theta))}\}_{i=1 \ldots NL}$ is the steering vector, i=1 ... NL, and k=mod(i−1,L) and n=floor ((i−1)/L).

[0067] FIGS. **7**A-**7**C exemplarily illustrate an embodiment of a microphone array **201** when the target sound source is in a three dimensional plane. In an embodiment where the target sound source that emits the target sound signal is in a three dimensional plane, the delay (τ) between each of the sound sensors **301** and the origin of the microphone array **201** is determined as a function of distance (d) between each of the sound sensors **301** and the origin, a predefined angle (Φ) between each of the sound sensors **301** and a first reference axis (Y), an elevation angle (Ψ) between a second reference axis (Z) and the target sound signal, and an azimuth angle (θ) between the first reference axis (Y) and the target sound signal. The determined delay (τ) is represented in terms of number of samples. The determination of the delay enables beamforming for arbitrary numbers of the sound sensors **301** and multiple arbitrary configurations of the microphone array **201**.

[0068] Consider an example of a microphone array configuration with four sound sensors **301** M_0, M_1, M_2, and M_3. FIG. **7**A exemplarily illustrates a graphical representation of a microphone array **201**, when the target sound source in a three dimensional plane. As exemplarily illustrated in FIG. **7**A, the target sound signal from the target sound source is received from the direction (Ψ, θ) with reference to the origin of the microphone array **201**, where Ψ is the elevation angle and θ is the azimuth.

[0069] FIG. **7**B exemplarily illustrates a table showing delay between each sound sensor **301** in a circular microphone array configuration and the origin of the microphone array **201**, when the target sound source is in a three dimensional plane. The target sound source in a three dimensional plane emits a target sound signal from a spatial location (Ψ, θ). The distances between the origin O and the sound sensors **301** M_0, M_1, M_2, and M_3 when the incoming target sound signal from the target sound source is at an angle (Ψ, θ) from

the Z-axis and the Y-axis respectively, are denoted as τ_0, τ_1, τ_2, and τ_3 respectively. When the spatial location of the target sound signal moves from the location $\Psi=90°$ to a location $\Psi=0°$, sin(Ψ) changes from 1 to 0, and as a result, the difference between each sound sensor **301** in the microphone array **201** becomes smaller and smaller. When $\Psi=0°$, there is no difference between the sound sensors **301**, which implies that the target sound signal reaches each sound sensor **301** at the same time. Taking into account that the sample delay between the sound sensors **301** can only be an integer, the range where all the sound sensors **301** are identical is determined.

[0070] FIG. **7**C exemplarily illustrates a three dimensional working space of the microphone array **201**, where the target sound signal is incident at an elevation angle $\Psi<\Omega$, where Ω is a specific angle and is a variable representing the elevation angle. When the target sound signal is incident at an elevation angle $\Psi<\Omega$, all four sound sensors **301** M_0, M_1, M_2, and M_3 receive the same target sound signal for 0°<0<360°. The delay τ is a function of both the elevation angle Ψ and the azimuth angle θ. That is, $\tau=\tau(\theta, \Psi)$. As used herein, Ω refers to the elevation angle such that all $\tau_i (\theta, \Omega)$ are equal to each other, where i=0, 1, 2, 3, etc. The value of Ω is determined by the sample delay between each of the sound sensors **301** and the origin of the microphone array **201**. The adaptive beamforming unit **203** enhances sound from this range and suppresses sound signals from other directions, for example, S_1 and S_2 treating them as ambient noise signals.

[0071] Consider a least mean square solution for beamforming according to the method disclosed herein. Let the spatial directivity pattern be 1 in the passband and 0 in the stopband. The least square cost function is defined as:

$$J(w) = \int_{\Omega_p}\int_{\Theta_p} |H(\omega, \theta) - 1|^2 d\omega d\theta + \qquad (3)$$

$$\alpha \int_{\Omega_s}\int_{\Theta_s} |H(\omega, \theta)|^2 d\omega d\theta$$

$$= \int_{\Omega_p}\int_{\Theta_p} |H(\omega, \theta)|^2 d\omega d\theta +$$

$$\alpha \int_{\Omega_S}\int_{\Theta_X} |H(\omega, \theta)|^2 d\omega d\theta -$$

$$2 \int_{\Omega_p}\int_{\Theta_p} \mathrm{Re}(H(\omega, \theta)) d\omega d\theta +$$

$$\int_{\Omega_p}\int_{\Theta_p} 1 d\omega d\theta$$

Replacing

[0072]

$|H(\omega,\theta)|^2=w^Tg(\omega,\theta)g^H(\omega,\theta)w=w^T(G_R(\omega,\theta)+jG_1(\omega,\theta))w=w^TG_R(\omega,\theta)w$ and $Re(H(\omega,\theta))=w^Tg_R(\omega,\theta),J(\omega)$ becomes

$J(\omega)=w^TQw-2w^T\alpha+d,$ where

$Q=\int_{\Omega_p}\int_{73\,p}G_R(\omega,\theta)d\omega d\theta+a\theta_{\Omega_s}\int_{\Theta_S}G_R(\omega,\theta)d\omega d\theta$

$\alpha=\int_{\Omega_p}\int_{\Theta_p}g_R(\omega,\theta)d\omega d\theta$

$d=\int_{\Omega_p}\int_{\Theta_p}1 d\omega d\theta \qquad (4)$

where $g_R(\omega,\theta)=\cos[\omega(k+\tau_n)]$ and $G_R(\omega,\theta)=\cos[\omega(k-1+\tau_n-\tau_m)]$.

When $\partial J/\partial w=0$, the cost function J is minimized. The least-square estimate of w is obtained by:

$$w=Q^{-1}\alpha \tag{5}$$

[0073] Applying linear constrains Cw=b, the spatial response is further constrained to a predefined value b at angle θ_f using following equation:

$$\begin{bmatrix} g_R^T(\omega_{start}, \theta_f) \\ \dots \\ g_R^T(\omega_{end}, \theta_f) \end{bmatrix} w = \begin{bmatrix} b_{start} \\ \dots \\ b_{end} \end{bmatrix} \tag{6}$$

Now, the design problem becomes:

$$\min_w w^T Q w - 2w^T a + d \text{ subject to } Cw = b \tag{7}$$

and the solution of the constrained minimization problem is equal to:

$$w=Q^{-1}C^T(CQ^{-1}C^T)^{-1}(b-CQ^{-1}\alpha)+Q^{-1}\alpha \tag{8}$$

where w is the filter parameter for the designed adaptive beamforming unit **203**.

[0074] In an embodiment, the beamforming is performed by a delay-sum method. In another embodiment, the beamforming is performed by a filter-sum method.

[0075] FIG. **8** exemplarily illustrates a method for estimating a spatial location of the target sound signal from the target sound source by the sound source localization unit **202** using a steered response power-phase transform (SRP-PHAT). The SRP-PHAT combines the advantages of sound source localization methods, for example, the time difference of arrival (TDOA) method and the steered response power (SRP) method. The TDOA method performs the time delay estimation of the sound signals relative to a pair of spatially separated sound sensors **301**. The estimated time delay is a function of both the location of the target sound source and the position of each of the sound sensors **301** in the microphone array **201**. Because the position of each of the sound sensors **301** in the microphone array **201** is predefined, once the time delay is estimated, the location of the target sound source can be determined. In the SRP method, a filter-and-sum beamforming algorithm is applied to the microphone array **201** for sound signals in the direction of each of the disparate sound sources. The location of the target sound source corresponds to the direction in which the output of the filter-and-sum beamforming has the largest response power. The TDOA based localization is suitable under low to moderate reverberation conditions. The SRP method requires shorter analysis intervals and exhibits an elevated insensitivity to environmental conditions while not allowing for use under excessive multi-path. The SRP-PHAT method disclosed herein combines the advantages of the TDOA method and the SRP method, has a decreased sensitivity to noise and reverberations compared to the TDOA method, and provides more precise location estimates than existing localization methods.

[0076] For direction i $(0 \leq i \leq 360)$, the delay D_{it} is calculated **801** between the tth pair of the sound sensors **301** (t=1: all pairs). The correlation value corr(D_{it}) between the tth pair

of the sound sensors **301** corresponding to the delay of D_{it} is then calculated **802**. For the direction i $(0 \leq i \leq 360)$, the correlation value is given **803** by:

$$CORR_i = \sum_{t=1}^{ALL\ PAIR} \text{corr}(D_{it})$$

Therefore, the spatial location of the target sound signal is given **804** by:

$$S = \underset{0 \leq i \leq 360}{\arg\max} CORR_i.$$

[0077] FIGS. **9**A-**9**B exemplarily illustrate graphs showing the results of sound source localization performed using the steered response power-phase transform (SRP-PHAT). FIG. **9**A exemplarily illustrates a graph showing the value of the SRP-PHAT for every 10° The maximum value corresponds to the location of the target sound signal from the target sound source. FIG. **9**B exemplarily illustrates a graph representing the estimated target sound signal from the target sound source and a ground truth.

[0078] FIG. **10** exemplarily illustrates a system for performing adaptive beamforming by the adaptive beamforming unit **203**. The algorithm for fixed beamforming is disclosed with reference to equations (3) through (8) in the detailed description of FIG. **4**, FIGS. **6**A-**6**B, and FIGS. **7**A-**7**C, which is extended herein to adaptive beamforming. Adaptive beamforming refers to a beamforming process where the directivity pattern of the microphone array **201** is adaptively steered in the direction of a target sound signal emitted by a target sound source in motion. Adaptive beamforming achieves better ambient noise suppression than fixed beamforming. This is because the target direction of arrival, which is assumed to be stable in fixed beamforming, changes with the movement of the target sound source. Moreover, the gains of the sound sensors **301** which are assumed uniform in fixed beamforming, exhibit significant distribution. All these factors reduce speech quality. On the other hand, adaptive beamforming adaptively performs beam steering and null steering; therefore, the adaptive beamforming method is more robust against steering error caused by the array imperfection mentioned above.

[0079] As exemplarily illustrated in FIG. **10**, the adaptive beamforming unit **203** disclosed herein comprises a fixed beamformer **204**, a blocking matrix **205**, an adaptation control unit **208**, and an adaptive filter **206**. The fixed beamformer **204** adaptively steers the directivity pattern of the microphone array **201** in the direction of the spatial location of the target sound signal from the target sound source for enhancing the target sound signal, when the target sound source is in motion. The sound sensors **301** in the microphone array **201** receive the sound signals $S_1, \dots, S_4$, which comprise both the target sound signal from the target sound source and the ambient noise signals. The received sound signals are fed as input to the fixed beamformer **204** and the blocking matrix **205**. The fixed beamformer **204** outputs a signal "b". In an embodiment, the fixed beamformer **204** performs fixed beamforming by filtering and summing output sound signals from the sound sensors **301**. The blocking matrix **205** outputs a signal "z" which primarily comprises the ambient noise sig-

nals. The blocking matrix 205 blocks the target sound signal from the target sound source and feeds the ambient noise signals to the adaptive filter 206 to minimize the effect of the ambient noise signals on the enhanced target sound signal.

[0080] The output "z" of the blocking matrix 205 may contain some weak target sound signals due to signal leakage. If the adaptation is active when the target sound signal, for example, speech is present, the speech is cancelled out with the noise. Therefore, the adaptation control unit 208 determines when the adaptation should be applied. The adaptation control unit 208 comprises a target sound signal detector 208a and a step size adjusting module 208b. The target sound signal detector 208a of the adaptation control unit 208 detects the presence or absence of the target sound signal, for example, speech. The step size adjusting module 208b adjusts the step size for the adaptation process such that when the target sound signal is present, the adaptation is slow for preserving the target sound signal, and when the target sound signal is absent, adaptation is quick for better cancellation of the ambient noise signals.

[0081] The adaptive filter 206 is a filter that adaptively updates filter coefficients of the adaptive filter 206 so that the adaptive filter 206 can be operated in an unknown and changing environment. The adaptive filter 206 adaptively filters the ambient noise signals in response to detecting presence or absence of the target sound signal in the sound signals received from the disparate sound sources. The adaptive filter 206 adapts its filter coefficients with the changes in the ambient noise signals, thereby eliminating distortion in the target sound signal, when the target sound source and the ambient noise signals are in motion. In an embodiment, the adaptive filtering is performed by a set of sub-band adaptive filters using sub-band adaptive filtering as disclosed in the detailed description of FIG. 11.

[0082] FIG. 11 exemplarily illustrates a system for sub-band adaptive filtering. Sub-band adaptive filtering involves separating a full-band signal into different frequency ranges called sub-bands prior to the filtering process. The sub-band adaptive filtering using sub-band adaptive filters lead to a higher convergence speed compared to using a full-band adaptive filter. Moreover, the noise reduction unit 207 disclosed herein is developed in a sub-band, whereby applying sub-band adaptive filtering provides the same sub-band framework for both beamforming and noise reduction, and thus saves on computational cost.

[0083] As exemplarily illustrated in FIG. 11, the adaptive filter 206 comprises an analysis filter bank 206a, an adaptive filter matrix 206b, and a synthesis filter bank 206c. The analysis filter bank 206a splits the enhanced target sound signal (b) from the fixed beamformer 204 and the ambient noise signals (z) from the blocking matrix 205 exemplarily illustrated in FIG. 10 into multiple frequency sub-bands. The analysis filter bank 206a performs an analysis step where the outputs of the fixed beamformer 204 and the blocking matrix 205 are split into frequency sub bands. The sub-band adaptive filter 206 typically has a shorter impulse response than its full band counterpart. The step size of the sub-bands can be adjusted individually for each sub-band by the step-size adjusting module 208b, which leads to a higher convergence speed compared to using a full band adaptive filter.

[0084] The adaptive filter matrix 206b adaptively filters the ambient noise signals in each of the frequency sub-bands in response to detecting the presence or absence of the target sound signal in the sound signals received from the disparate sound sources. The adaptive filter matrix 206b performs an adaptation step, where the adaptive filter 206 is adapted such that the filter output only contains the target sound signal, for example, speech. The synthesis filter bank 206c synthesizes a full-band sound signal using the frequency sub-bands of the enhanced target sound signal. The synthesis filter bank 206c performs a synthesis step where the sub-band sound signal is synthesized into a full-band sound signal. Since the noise reduction and the beamforming are performed in the same sub-band framework, the noise reduction as disclosed in the detailed description of FIG. 13, by the noise reduction unit 207 is performed prior to the synthesis step, thereby reducing computation.

[0085] In an embodiment, the analysis filter bank 206a is implemented as a perfect-reconstruction filter bank, where the output of the synthesis filter bank 206c after the analysis and synthesis steps perfectly matches the input to the analysis filter bank 206a. That is, all the sub-band analysis filter banks 206a are factorized to operate on prototype filter coefficients and a modulation matrix is used to take advantage of the fast Fourier transform (FFT). Both analysis and synthesize steps require performing frequency shifts in each sub-band, which involves complex value computations with cosines and sinusoids. The method disclosed herein employs the FFT to perform the frequency shifts required in each sub-band, thereby minimizing the amount of multiply-accumulate operations. The implementation of the sub-band analysis filter bank 206a as a perfect-reconstruction filter bank ensures the quality of the target sound signal by ensuring that the sub-band analysis filter banks 206a do not distort the target sound signal itself.

[0086] FIG. 12 exemplarily illustrates a graphical representation showing the performance of a perfect-reconstruction filter bank. The solid line represents the input signal to the analysis filter bank 206a, and the circles represent the output of the synthesis filter bank 206c after analysis and synthesis. As exemplarily illustrated in FIG. 12, the output of the synthesis filter bank 206c perfectly matches the input, and is therefore referred to as the perfect-reconstruction filter bank.

[0087] FIG. 13 exemplarily illustrates a block diagram of a noise reduction unit 207 for performing noise reduction using, for example, a Wiener-filter based noise reduction algorithm. The noise reduction unit 207 performs noise reduction for further suppressing the ambient noise signals after adaptive beamforming, for example, by using a Wiener-filter based noise reduction algorithm, a spectral subtraction noise reduction algorithm, an auditory transform based noise reduction algorithm, or a model based noise reduction algorithm. In an embodiment, the noise reduction unit 207 performs noise reduction in multiple frequency sub-bands employed by an analysis filter bank 206a of the adaptive beamforming unit 203 for sub-band adaptive beamforming.

[0088] In an embodiment, the noise reduction is performed using the Wiener-filter based noise reduction algorithm. The noise reduction unit 207 explores the short-term and long-term statistics of the target sound signal, for example, speech, and the ambient noise signals, and the wide-band and narrow-band signal-to-noise ratio (SNR) to support a Wiener gain filtering. The noise reduction unit 207 comprises a target sound signal statistics analyzer 207a, a noise statistics analyzer 207b, a signal-to-noise ratio (SNR) analyzer 207c, and a Wiener filter 207d. The target sound signal statistics analyzer 207a explores the short-term and long-term statistics of the target sound signal, for example, speech. Similarly, the noise statistics analyzer 207b explores the short-term and

long-term statistics of the ambient noise signals. The SNR analyzer **207***c* of the noise reduction unit **207** explores the wide-band and narrow-band signal-to-noise ratio (SNR). After the spectrum of noisy-speech passes through the Wiener filter **207***d*, an estimation of the clean-speech spectrum is generated. The synthesis filter bank **206***c*, by an inverse process of the analysis filter bank **206***a*, reconstructs the signals of the clean speech into a full-band signal, given the estimated spectrum of the clean speech.

[0089]   FIG. **14** exemplarily illustrates a hardware implementation of the microphone array system **200** disclosed herein. The hardware implementation of the microphone array system **200** disclosed in the detailed description of FIG. **2** comprises the microphone array **201** having an arbitrary number of sound sensors **301** positioned in an arbitrary configuration, multiple microphone amplifiers **1401**, one or more audio codecs **1402**, a digital signal processor (DSP) **1403**, a flash memory **1404**, one or more power regulators **1405** and **1406**, a battery **1407**, a loudspeaker or a headphone **1408**, and a communication interface **1409**. The microphone array **201** comprises, for example, four or eight sound sensors **301** arranged in a linear or a circular microphone array configuration. The microphone array **201** receives the sound signals.

[0090]   Consider an example where the microphone array **201** comprises four sound sensors **301** that pick up the sound signals. Four microphone amplifiers **1401** receive the output sound signals from the four sound sensors **301**. The microphone amplifiers **1401** also referred to as preamplifiers provide a gain to boost the power of the received sound signals for enhancing the sensitivity of the sound sensors **301**. In an example, the gain of the preamplifiers is 20 dB.

[0091]   The audio codec **1402** receives the amplified output from the microphone amplifiers **1401**. The audio codec **1402** provides an adjustable gain level, for example, from about −74 dB to about 6 dB. The received sound signals are in an analog form. The audio codec **1402** converts the four channels of the sound signals in the analog form into digital sound signals. The pre-amplifiers may not be required for some applications. The audio codec **1402** then transmits the digital sound signals to the DSP **1403** for processing of the digital sound signals. The DSP **1403** implements the sound source localization unit **202**, the adaptive beamforming unit **203**, and the noise reduction unit **207**.

[0092]   After the processing, the DSP **1403** either stores the processed signal from the DSP **1403** in a memory device for a recording application, or transmits the processed signal to the communication interface **1409**. The recording application comprises, for example, storing the processed signal onto the memory device for the purposes of playing back the processed signal at a later time. The communication interface **1409** transmits the processed signal, for example, to a computer, the internet, or a radio for communicating the processed signal. In an embodiment, the microphone array system **200** disclosed herein implements a two-way communication device where the signal received from the communication interface **1409** is processed by the DSP **1403** and the processed signal is then played through the loudspeaker or the headphone **1408**.

[0093]   The flash memory **1404** stores the code for the DSP **1403** and compressed audio signals. When the microphone array system **200** boots up, the DSP **1403** reads the code from the flash memory **1404** into an internal memory of the DSP **1403** and then starts executing the code. In an embodiment, the audio codec **1402** can be configured for encoding and

decoding audio or sound signals during the start up stage by writing to registers of the DSP **1403**. For an eight-sensor microphone array **201**, two four-channel audio codec **1402** chips may be used. The power regulators **1405** and **1406**, for example, linear power regulators **1405** and switch power regulators **1406** provide appropriate voltage and current supply for all the components, for example, **201**, **1401**, **1402**, **1403**, etc., mechanically supported and electrically connected on a circuit board. A universal serial bus (USB) control is built into the DSP **1403**. The battery **1407** is used for powering the microphone array system **200**.

[0094]   Consider an example where the microphone array system **200** disclosed herein is implemented on a mixed signal circuit board having a six-layer printed circuit board (PCB). Noisy digital signals easily contaminate the low voltage analog sound signals from the sound sensors **301**. Therefore, the layout of the mixed signal circuit board is carefully partitioned to isolate the analog circuits from the digital circuits. Although both the inputs and outputs of the microphone amplifiers **1401** are in analog form, the microphone amplifiers **1401** are placed in a digital region of the mixed signal circuit board because of their high power consumption **1401** and switch amplifier nature.

[0095]   The linear power regulators **1405** are deployed in an analog region of the mixed signal circuit board due to the low noise property exhibited by the linear power regulators **1405**. Five power regulators, for example, **1405** are designed in the microphone array system **200** circuits to ensure quality. The switch power regulators **1406** achieve an efficiency of about 95% of the input power and have high output current capacity; however their outputs are too noisy for analog circuits. The efficiency of the linear power regulators **1405** is determined by the ratio of the output voltage to the input voltage, which is lower than that of the switch power regulators **1406** in most cases. The regulator outputs utilized in the microphone array system **200** circuits are stable, quiet, and suitable for the low power analog circuits.

[0096]   In an example, the microphone array system **200** is designed with a microphone array **201** having dimensions of 10 cm×2.5 cm×1.5 cm, a USB interface, and an assembled PCB supporting the microphone array **201** and a DSP **1403** having a low power consumption design devised for portable devices, a four-channel codec **1402**, and a flash memory **1404**. The DSP **1403** chip is powerful enough to handle the DSP **1403** computations in the microphone array system **200** disclosed herein. The hardware configuration of this example can be used for any microphone array configuration, with suitable modifications to the software. In an embodiment, the adaptive beamforming unit **203** of the microphone array system **200** is implemented as hardware with software instructions programmed on the DSP **1403**. The DSP **1403** is programmed for beamforming, noise reduction, echo cancellation, and USB interfacing according to the method disclosed herein, and fine tuned for optimal performance.

[0097]   FIGS. **15A-15C** exemplarily illustrate a conference phone **1500** comprising an eight-sensor microphone array **201**. The eight-sensor microphone array **201** comprises eight sound sensors **301** arranged in a configuration as exemplarily illustrated in FIG. **15A**. A top view of the conference phone **1500** comprising the eight-sensor microphone array **201** is exemplarily illustrated in FIG. **15A**. A front view of the conference phone **1500** comprising the eight-sensor microphone array **201** is exemplarily illustrated in FIG. **15B**. A headset **1502** that can be placed in a base holder **1501** of the confer-

9

ence phone **1500** having the eight-sensor microphone array **201** is exemplarily illustrated in FIG. **15**C. In addition to a conference phone **1500**, the microphone array system **200** disclosed herein with broadband beamforming can be configured for a mobile phone, a tablet computer, etc., for speech enhancement and noise reduction.

[0098] FIG. **16**A exemplarily illustrates a layout of an eight-sensor microphone array **201** for a conference phone **1500**. Consider an example of a circular microphone array **201** in which eight sound sensors **301** are mounted on the surface of the conference phone **1500** as exemplarily illustrated in FIG. **15**A. The conference phone **1500** has a removable handset **1502** on top, and hence the microphone array system **200** is configured to accommodate the handset **1502** as exemplarily illustrated in FIGS. **15**A-**15**C. In an example, the circular microphone array **201** has a diameter of about four inches. Eight sound sensors **301**, for example, microphones, M_0, M_1, M_2, M_3, M_4, M_5, M_6, and M_7 are distributed along a circle **302** on the conference phone **1500**. Microphones M_4-M_7 are separated by 90 degrees from each other, and microphones M_0-M_3 are rotated counterclockwise by 60 degrees from microphone M_4-M_7 respectively.

[0099] FIG. **16**B exemplarily illustrates a graphical representation of eight spatial regions to which the eight-sensor microphone array **201** of FIG. **16**A responds. The space is divided into eight spatial regions with equal spaces centered at 15°, 60°, 105°, 150°, 195°, 240°, 285°, and 330° respectively. The adaptive beamforming unit **203** configures the eight-sensor microphone array **201** to automatically point to one of these eight spatial regions according to the location of the target sound signal from the target sound source as estimated by the sound source localization unit **202**.

[0100] FIGS. **16**C-**16**D exemplarily illustrate computer simulations showing the steering of the directivity patterns of the eight-sensor microphone array **201** of FIG. **16**A, in the directions 15° and 60° respectively, in the frequency range 300 Hz to 5 kHz. FIG. **16**C exemplarily illustrates the computer simulation result showing the directivity pattern of the microphone array **201** when the target sound signal is received from the target sound source in the spatial region centered at 15°.

[0101] The computer simulation for verifying the performance of the adaptive beamforming unit **203** when the target sound signal is received from the target sound source in the spatial region centered at 15° uses the following parameters:

Sampling frequency fs=16 k,

FIR filter taper length L=20

Passband (Θ_p, Ω_p)={300-5000 Hz, -5°-35°}, designed spatial directivity pattern is 1.

Stopband (Θ_s, Ω_s)={300~5000 Hz, -180°~-15°+45°~180°}, the designed spatial directivity pattern is 0.

[0102] It can be seen that the directivity pattern of the microphone array **201** in the spatial region centered at 15° is enhanced while the sound signals from all other spatial regions are suppressed.

[0103] FIG. **16**D exemplarily illustrates the computer simulation result showing the directivity pattern of the microphone array **201** when the target sound signal is received from the target sound source in the spatial region centered at 60°. The computer simulation for verifying the performance of the adaptive beamforming unit **203** when the target sound signal is received from the target sound source in the spatial region centered at 60° uses the following parameters:

Sampling frequency fs=16 k,

FIR filter taper length L=20

Passband (Θ_p, Ω_p)={300-5000 Hz, 40°-80°}, designed spatial directivity pattern is 1.

Stopband (Θ_s, Ω_s)={300~5000 Hz, -180°~30°+90°~180°}, the designed spatial directivity pattern is 0.

[0104] It can be seen that the directivity pattern of the microphone array **201** in the spatial region centered at 60° is enhanced while the sound signals from all other spatial regions are suppressed. The other six spatial regions have similar parameters. Moreover, in all frequencies, the main lobe has the same level, which means the target sound signal has little distortion in frequency.

[0105] FIGS. **16**E-**16**L exemplarily illustrate graphical representations showing the directivity patterns of the eight-sensor microphone array **201** of FIG. **16**A in each of the eight spatial regions, where each directivity pattern is an average response from 300 Hz to 5000 Hz. The main lobe is about 10 dB higher than the side lobe, and therefore the ambient noise signals from other directions are highly suppressed compared to the target sound signal in the pass direction. The microphone array system **200** calculates the filter coefficients for the target sound signal, for example, speech signals from each sound sensor **301** and combines the filtered signals to enhance the speech from any specific direction. Since speech covers a large range of frequencies, the method and system **200** disclosed herein covers broadband signals from 300 Hz to 5000 Hz.

[0106] FIG. **16**E exemplarily illustrates a graphical representation showing the directivity pattern of the eight-sensor microphone array **201** when the target sound signal is received from the target sound source in the spatial region centered at 15°. FIG. **16**F exemplarily illustrates a graphical representation showing the directivity pattern of the eight-sensor microphone array **201** when the target sound signal is received from the target sound source in the spatial region centered at 60°. FIG. **16**G exemplarily illustrates a graphical representation showing the directivity pattern of the eight-sensor microphone array **201** when the target sound signal is received from the target sound source in the spatial region centered at 105°. FIG. **16**H exemplarily illustrates a graphical representation showing the directivity pattern of the eight-sensor microphone array **201** when the target sound signal is received from the target sound source in the spatial region centered at 150°. FIG. **16**I exemplarily illustrates a graphical representation showing the directivity pattern of the eight-sensor microphone array **201** when the target sound signal is received from the target sound source in the spatial region centered at 195°. FIG. **16**J exemplarily illustrates a graphical representation showing the directivity pattern of the eight-sensor microphone array **201** when the target sound signal is received from the target sound source in the spatial region centered at 240°. FIG. **16**K exemplarily illustrates a graphical representation showing the directivity pattern of the eight-sensor microphone array **201** when the target sound signal is received from the target sound source in the spatial region centered at 285°. FIG. **16**L exemplarily illustrates a graphical representation showing the directivity pattern of the eight-sensor microphone array **201** when the target sound signal is received from the target sound source in the spatial region centered at 330°. The microphone array system **200** disclosed herein enhances the target sound signal from each of the

directions 15°, 60°, 105°, 150°, 195°, 240°, 285°, and 330°, while suppressing the ambient noise signals from the other directions.

[0107] The microphone array system 200 disclosed herein can be implemented for a square microphone array configuration and a rectangular array configuration where a sound sensor 301 is positioned in each corner of the four-cornered array. The microphone array system 200 disclosed herein implements beamforming from plane to three dimensional sound sources.

[0108] FIG. 17A exemplarily illustrates a graphical representation of four spatial regions to which a four-sensor microphone array 201 for a wireless handheld device responds. The wireless handheld device is, for example, a mobile phone. Consider an example where the microphone array 201 comprises four sound sensors 301, for example, microphones, uniformly distributed around a circle 302 having diameter equal to about two inches. This configuration is identical to positioning four sound sensors 301 or microphones on four corners of a square. The space is divided into four spatial regions with equal space centered at −90°, 0°, 90°, and 180° respectively. The adaptive beamforming unit 203 configures the four-sensor microphone array 201 to automatically point to one of these spatial regions according to the location of the target sound signal from the target sound source as estimated by the sound source localization unit 202.

[0109] FIGS. 17B-17I exemplarily illustrate computer simulations showing the directivity patterns of the four-sensor microphone array 201 of FIG. 17A with respect to azimuth and frequency. The results of the computer simulations performed for verifying the performance of the adaptive beamforming unit 203 of the microphone array system 200 disclosed herein for a sampling frequency f_s=16 k and FIR filter taper length L=20, are as follows:
For the spatial region centered at 0°:
Passband (Θ_p, Ω_p)={300-4000 Hz, −20°-20°}, designed spatial directivity pattern is 1.

[0110] Stopband (Θ, Ω_s)={300~4000 Hz, −180°~−30°+30°~180°}, the designed spatial directivity pattern is 0.

[0111] For the spatial region centered at 90°:
Passband (Θ_p, Ω_p)={300-4000 Hz, 70°-110°}, designed spatial directivity pattern is 1.
Stopband (Θ_s, Ω_s)={300~4000 Hz, −180°~−60°+120°~180°}, the designed spatial directivity pattern is 0. The directivity patterns for the spatial regions centered at −90° and 180° are similarly obtained.

[0112] FIG. 17B exemplarily illustrates the computer simulation result representing a three dimensional (3D) display of the directivity pattern of the four-sensor microphone array 201 when the target sound signal is received from the target sound source in the spatial region centered at −90°. FIG. 17C exemplarily illustrates the computer simulation result representing a 2D display of the directivity pattern of the four-sensor microphone array 201 when the target sound signal is received from the target sound source in the spatial region centered at −90°.

[0113] FIG. 17D exemplarily illustrates the computer simulation result representing a 3D display of the directivity pattern of the four-sensor microphone array 201 when the target sound signal is received from the target sound source in the spatial region centered at 0°. FIG. 17E exemplarily illustrates the computer simulation result representing a 2D display of the directivity pattern of the four-sensor microphone

array 201 when the target sound signal is received from the target sound source in the spatial region centered at 0°.

[0114] FIG. 17F exemplarily illustrates the computer simulation result representing a 3D display of the directivity pattern of the four-sensor microphone array 201 when the target sound signal is received from the target sound source in the spatial region centered at 90°. FIG. 17G exemplarily illustrates the computer simulation result representing a 2D display of the directivity pattern of the four-sensor microphone array 201 when the target sound signal is received from the target sound source in the spatial region centered at 90°.

[0115] FIG. 17H exemplarily illustrates the computer simulation result representing a 3D display of the directivity pattern of the four-sensor microphone array 201 when the target sound source is received from the target sound source in the spatial region centered at 180°. FIG. 17I exemplarily illustrates the computer simulation result representing a 2D display of the directivity pattern of the four-sensor microphone array 201 when the target sound source is received from the target sound source in the spatial region centered at 180°. The 3D displays of the directivity patterns in FIG. 17B, FIG. 17D, FIG. 17F, and FIG. 17H demonstrate that the passbands have the same height. The 2D displays of the directivity patterns in FIG. 17C, FIG. 17E, FIG. 17G, and FIG. 17I demonstrate that the passbands have the same width along the frequency and demonstrates the broadband properties of the microphone array 201.

[0116] FIGS. 18A-18B exemplarily illustrates a microphone array configuration for a tablet computer. In this example, four sound sensors 301 of the microphone array 201 are positioned on a frame 1801 of the tablet computer, for example, the iPad® of Apple Inc. Geometrically, the sound sensors 301 are distributed on the circle 302 as exemplarily in FIG. 18B. The radius of the circle 302 is equal to the width of the tablet computer. The angle θ between the sound sensors 301 M_2 and M_3 is determined to avoid spatial aliasing up to 4000 Hz. This microphone array configuration enhances a front speaker's voice and suppresses background ambient noise. The adaptive beamforming unit 203 configures the microphone array 201 to form an acoustic beam 1802 pointing frontwards using the method and system 200 disclosed herein. The target sound signal, that is, the front speaker's voice within the range of Φ<30° is enhanced compared to the sound signals from other directions.

[0117] FIG. 18C exemplarily illustrates an acoustic beam 1802 formed using the microphone array configuration of FIGS. 18A-18B according to the method and system 200 disclosed herein.

[0118] FIGS. 18D-18G exemplarily illustrates graphs showing processing results of the adaptive beamforming unit 203 and the noise reduction unit 207 for the microphone array configuration of FIG. 18B, in both a time domain and a spectral domain for the tablet computer. Consider an example where a speaker is talking in front of the tablet computer with ambient noise signals on the side. FIG. 18D exemplarily illustrates a graph showing the performance of the microphone array 201 before performing beamforming and noise reduction with a signal-to-noise ratio (SNR) of 15 dB. FIG. 18E exemplarily illustrates a graph showing the performance of the microphone array 201 after performing beamforming and noise reduction, according to the method disclosed herein, with an SNR of 15 dB. FIG. 18F exemplarily illustrates a graph showing the performance of the microphone array 201 before performing beamforming and noise reduc-

tion with an SNR of 0 dB. FIG. **18**G exemplarily illustrates a graph showing the performance of the microphone array **201** after performing beamforming and noise reduction, according to the method disclosed herein, with an SNR of 0 dB.

[0119] It can be seen from FIGS. **18**D-**18**G that the performance graph is noisier for the microphone array **201** before the beamforming and noise reduction is performed. Therefore, the adaptive beamforming unit **203** and the noise reduction unit **207** of the microphone array system **200** disclosed herein suppresses ambient noise signals while maintaining the clarity of the target sound signal, for example, the speech signal.

[0120] FIGS. **19**A-**19**F exemplarily illustrate tables showing different microphone array configurations and the corresponding values of delay τ_n for the sound sensors **301** in each of the microphone array configurations. The broadband beamforming method disclosed herein can be used for microphone arrays **201** with arbitrary numbers of sound sensors **301** and arbitrary locations of the sound sensors **301**. The sound sensors **301** can be mounted on surfaces or edges of any speech acquisition device. For any specific microphone array configuration, the only parameter that needs to be defined to achieve the beamformer coefficients is the value of τ_n for each sound sensor **301** as disclosed in the detailed description of FIG. **5**, FIGS. **6**A-**6**B, and FIGS. **7**A-**7**C and as exemplarily illustrated in FIGS. **19**A-**19**F. In an example, the microphone array configuration exemplarily illustrated in FIG. **19**F is implemented on a handheld device for hands-free speech acquisition. In a hands-free and non-close talking scenario, a user prefers to talk in distance rather than speaking close to the sound sensor **301** and may want to talk while watching a screen of the handheld device. The microphone array system **200** disclosed herein allows the handheld device to pick up sound signals from the direction of the speaker's mouth and suppress noise from other directions. The method and system **200** disclosed herein may be implemented on any device or equipment, for example, a voice recorder where a target sound signal or speech needs to be enhanced.

[0121] The foregoing examples have been provided merely for the purpose of explanation and are in no way to be construed as limiting of the present invention disclosed herein. While the invention has been described with reference to various embodiments, it is understood that the words, which have been used herein, are words of description and illustration, rather than words of limitation. Further, although the invention has been described herein with reference to particular means, materials and embodiments, the invention is not intended to be limited to the particulars disclosed herein; rather, the invention extends to all functionally equivalent structures, methods and uses, such as are within the scope of the appended claims. Those skilled in the art, having the benefit of the teachings of this specification, may affect numerous modifications thereto and changes may be made without departing from the scope and spirit of the invention in its aspects.

We claim:

1. A method for enhancing a target sound signal from a plurality of sound signals, comprising:

providing a microphone array system comprising an array of sound sensors positioned in an arbitrary configuration, a sound source localization unit, an adaptive beamforming unit, and a noise reduction unit, wherein said sound source localization unit, said adaptive beamform-

ing unit, and said noise reduction unit are in operative communication with said array of said sound sensors;

receiving said sound signals from a plurality of disparate sound sources by said sound sensors, wherein said received sound signals comprise said target sound signal from a target sound source among said disparate sound sources, and ambient noise signals;

estimating a spatial location of said target sound signal from said received sound signals by said sound source localization unit;

performing adaptive beamforming for steering a directivity pattern of said array of said sound sensors in a direction of said spatial location of said target sound signal by said adaptive beamforming unit, wherein said adaptive beamforming unit enhances said target sound signal and partially suppresses said ambient noise signals; and

suppressing said ambient noise signals by said noise reduction unit for further enhancing said target sound signal.

2. The method of claim **1**, wherein said spatial location of said target sound signal from said target sound source is estimated using a steered response power-phase transform by said sound source localization unit.

3. The method of claim **1**, further comprising determining a delay between each of said sound sensors and origin of said array of said sound sensors as a function of distance between each of said sound sensors and said origin, a predefined angle between each of said sound sensors and a reference axis, and an azimuth angle between said reference axis and said target sound signal, when said target sound source that emits said target sound signal is in a two dimensional plane, wherein said delay is represented in terms of number of samples, wherein said determination of said delay enables beamforming for arbitrary numbers of said sound sensors and a plurality of arbitrary configurations of said array of said sound sensors.

4. The method of claim **1**, further comprising determining a delay between each of said sound sensors and origin of said array of said sound sensors as a function of distance between each of said sound sensors and said origin, a predefined angle between each of said sound sensors and a first reference axis, an elevation angle between a second reference axis and said target sound signal, and an azimuth angle between said first reference axis and said target sound signal, when said target sound source that emits said target sound signal is in a three dimensional plane, wherein said delay is represented in terms of number of samples, wherein said determination of said delay enables beamforming for arbitrary numbers of said sound sensors and a plurality of arbitrary configurations of said array of said sound sensors.

5. The method of claim **1**, wherein said adaptive beamforming comprises:

providing a fixed beamformer, a blocking matrix, and an adaptive filter in said adaptive beamforming unit;

steering said directivity pattern of said array of said sound sensors in said direction of said spatial location of said target sound signal from said target sound source by said fixed beamformer for enhancing said target sound signal, when said target sound source is in motion;

feeding said ambient noise signals to said adaptive filter by blocking said target sound signal received from said target sound source using said blocking matrix; and

adaptively filtering said ambient noise signals by said adaptive filter in response to detecting one of presence and absence of said target sound signal in said sound signals received from said disparate sound sources.

6. The method of claim **5**, wherein said fixed beamformer performs fixed beamforming by filtering and summing output sound signals from said sound sensors.

7. The method of claim **5**, wherein said adaptive filtering comprises sub-band adaptive filtering performed by said adaptive filter, wherein said sub-band adaptive filtering comprises:

provides an analysis filter bank, an adaptive filter matrix, and a synthesis filter bank in said adaptive filter;

splitting said enhanced target sound signal from said fixed beamformer and said ambient noise signals from said blocking matrix into a plurality of frequency sub-bands by said analysis filter bank;

adaptively filtering said ambient noise signals in each of said frequency sub-bands by said adaptive filter matrix in response to detecting one of presence and absence of said target sound signal in said sound signals received from said disparate sound sources; and

synthesizing a full-band sound signal using said frequency sub-bands of said enhanced target sound signal by said synthesis filter bank.

8. The method of claim **5**, wherein said adaptive beamforming further comprises detecting said presence of said target sound signal by an adaptation control unit provided in said adaptive beamforming unit and adjusting a step size for said adaptive filtering in response to detecting one of said presence and said absence of said target sound signal in said sound signals received from said disparate sound sources.

9. The method of claim **1**, wherein said noise reduction unit performs noise reduction by using one of a Wiener-filter based noise reduction algorithm, a spectral subtraction noise reduction algorithm, an auditory transform based noise reduction algorithm, and a model based noise reduction algorithm.

10. The method of claim **1**, wherein said noise reduction unit performs noise reduction in a plurality of frequency sub-bands, wherein said frequency sub-bands are employed by an analysis filter bank of said adaptive beamforming unit for sub-band adaptive beamforming.

11. A system for enhancing a target sound signal from a plurality of sound signals, comprising:

an array of sound sensors positioned in an arbitrary configuration, wherein said sound sensors receive said sound signals from a plurality of disparate sound sources, wherein said received sound signals comprise said target sound signal from a target sound source among said disparate sound sources, and ambient noise signals;

a sound source localization unit that estimates a spatial location of said target sound signal from said received sound signals;

an adaptive beamforming unit that steers directivity pattern of said array of said sound sensors in a direction of said spatial location of said target sound signal, wherein said adaptive beamforming unit enhances said target sound signal and partially suppresses said ambient noise signals; and

a noise reduction unit that suppresses said ambient noise signals for further enhancing said target sound signal.

12. The system of claim **11**, wherein said sound source localization unit estimates said spatial location of said target

sound signal from said target sound source using a steered response power-phase transform.

13. The system of claim **11**, wherein said adaptive beamforming unit comprises:

a fixed beamformer that steers said directivity pattern of said array of said sound sensors in said direction of said spatial location of said target sound signal from said target sound source for enhancing said target sound signal, when said target sound source is in motion;

a blocking matrix that feeds said ambient noise signals to an adaptive filter by blocking said target sound signal received from said target sound source; and

said adaptive filter that adaptively filters said ambient noise signals in response to detecting one of presence and absence of said target sound signal in said sound signals received from said disparate sound sources.

14. The system of claim **13**, wherein said fixed beamformer performs fixed beamforming by filtering and summing output sound signals from said sound sensors.

15. The system of claim **13**, wherein said adaptive filter comprises a set of sub-band adaptive filters comprising:

an analysis filter bank that splits said enhanced target sound signal from said fixed beamformer and said ambient noise signals from said blocking matrix into a plurality of frequency sub-bands;

an adaptive filter matrix that adaptively filters said ambient noise signals in each of said frequency sub-bands in response to detecting one of presence and absence of said target sound signal in said sound signals received from said disparate sound sources; and

a synthesis filter bank that synthesizes a full-band sound signal using said frequency sub-bands of said enhanced target sound signal.

16. The system of claim **11**, wherein said adaptive beamforming unit further comprises an adaptation control unit that detects said presence of said target sound signal and adjusts a step size for said adaptive filtering in response to detecting one of said presence and said absence of said target sound signal in said sound signals received from said disparate sound sources.

17. The system of claim **11**, wherein said noise reduction unit is one of a Wiener-filter based noise reduction unit, a spectral subtraction noise reduction unit, an auditory transform based noise reduction unit, and a model based noise reduction unit.

18. The system of claim **11**, further comprising one or more audio codecs that convert said sound signals in an analog form of said sound signals into digital sound signals and reconverts said digital sound signals into said analog form of said sound signals.

19. The system of claim **11**, wherein said noise reduction unit performs noise reduction in a plurality of frequency sub-bands employed by an analysis filter bank of said adaptive beamforming unit for sub-band adaptive beamforming.

20. The system of claim **11**, wherein said array of said sound sensors is one of a linear array of said sound sensors, a circular array of said sound sensors, and an arbitrarily distributed coplanar array of said sound sensors.

* * * * *