

(51) International Patent Classification:
H04R 3/00 (2006.01)(21) International Application Number:
PCT/US2017/030220(22) International Filing Date:
28 April 2017 (28.04.2017)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
15/144,631 02 May 2016 (02.05.2016) US(71) Applicant: **KNOWLES ELECTRONICS, LLC**
[US/US]; 1151 Maplewood Drive, Itasca, Illinois 60143 (US).(72) Inventors: **ROY, Jonathon**; c/o Knowles Electronics, LLC, 1151 Maplewood Drive, Itasca, Illinois 60143 (US). **WOODRUFF, John**; c/o Knowles Electronics, LLC, 1151 Maplewood Drive, Itasca, Illinois 60143 (US). **SAKRI, Shailesh**; c/o Knowles Electronics, LLC, 1151 Maplewood Drive, Itasca, Illinois 60143 (US). **VERMA, Tony**; c/o Knowles Electronics, LLC, 1151 Maplewood Drive, Itasca, Illinois 60143 (US).(74) Agent: **DANIELSON, Mark J.** et al.; Foley & Lardner LLP, 3000 K Street NW, Suite 600, Washington, District of Columbia 20007 (US).(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(54) Title: STEREO SEPARATION AND DIRECTIONAL SUPPRESSION WITH OMNI-DIRECTIONAL MICROPHONES

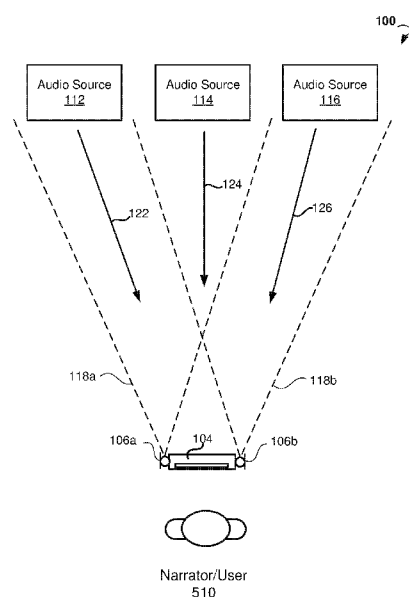


FIG. 1

(57) Abstract: Systems and methods for stereo separation and directional suppression are provided. An example method includes receiving a first audio signal, representing sound captured by a first microphone (106a) associated with a first location, and a second audio signal, representing sound captured by a second microphone (106b) associated with a second location. The microphones comprise omni-directional microphones. The distance between the first and second microphones is limited by the size of a mobile device (104). A first channel signal of a stereo signal is generated by forming, based on the first and second audio signals, a first beam at the first location. A second channel signal of the stereo signal is generated by forming, based on the first and second audio signals, a second beam at the second location. First and second directions, associated respectively with the first and second beams, are fixed relative to a line between the first and second locations.

Declarations under Rule 4.17:

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:

- *with international search report (Art. 21(3))*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

STEREO SEPARATION AND DIRECTIONAL SUPPRESSION WITH OMNI-DIRECTIONAL MICROPHONES

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of and priority to U.S. Patent Application No. 15/144,631, filed May 2, 2016, the entire contents of which are incorporated herein by reference.

FIELD

[0002] The present invention relates generally to audio processing, and, more specifically, to systems and methods for stereo separation and directional suppression with omni-directional microphones.

BACKGROUND

[0003] Recording stereo audio with a mobile device, such as smartphones and tablet computers, may be useful for making video of concerts, performances, and other events. Typical stereo recording devices are designed with either large separation between microphones or with precisely angled directional microphones to utilize acoustic properties of the directional microphones to capture stereo effects. Mobile devices, however, are limited in size and, therefore, the distance between microphones is significantly smaller than a minimum distance required for optimal omni-directional microphone stereo separation. Using directional microphones is not practical due to the size limitations of the mobile devices and may result in an increase in overall costs associated with the mobile devices. Additionally, due to the limited space for placing directional microphones, a user of the mobile device can be a dominant source for the directional microphones, often interfering with target sound sources.

[0004] Another aspect of recording stereo audio using a mobile device is a problem of capturing acoustically representative signals to be used in subsequent processing. Traditional

microphones used for mobile devices may not be able to handle high pressure conditions in which stereo recording is performed, such as a performance, concert, or a windy environment. As a result, signals generated by the microphones can become distorted due to reaching their acoustic overload point (AOP).

SUMMARY

[0005] This summary is provided to introduce a selection of concepts in a simplified form that are further described below in the Detailed Description. This summary is not intended to identify key features or essential features of the claimed subject matter, nor is it intended to be used as an aid in determining the scope of the claimed subject matter.

[0006] Provided are systems and methods for stereo separation and directional suppression with omni-directional microphones. An example method includes receiving at least a first audio signal and a second audio signal. The first audio signal can represent sound captured by a first microphone associated with a first location. The second audio signal can represent sound captured by a second microphone associated with a second location. The first microphone and the second microphone can include omni-directional microphones. The method can include generating a first channel signal of a stereo audio signal by forming, based on the at least first audio signal and second audio signal, a first beam at the first location. The method can also include generating a second channel signal of the stereo audio signal by forming, based on the at least first audio signal and second audio signal, a second beam at the second location.

[0007] In some embodiments, a distance between the first microphone and the second microphone is limited by a size of a mobile device. In certain embodiments, the first microphone is located at the top of the mobile device and the second microphone is located at the bottom of the mobile device. In other embodiments, the first and second microphones (and additional microphones, if any) may be located differently, including but not limited to, the microphones being located along a side of the device, e.g., separated along the side of a tablet having microphones on the side.

[0008] In some embodiments, directions of the first beam and the second beam are fixed relative to a line between the first location and the second location. In some embodiments,

the method further includes receiving at least one other acoustic signal. The other acoustic signal can be captured by another microphone associated with another location. The other microphone includes an omni-directional microphone. In some embodiments, forming the first beam and the second beam is further based on the other acoustic signal. In some embodiments, the other microphone is located off the line between the first microphone and the second microphone.

[0009] In some embodiments, forming the first beam includes reducing signal energy of acoustic signal components associated with sources outside the first beam. Forming the second beam can include reducing signal energy of acoustic signal components associated with further sources off the second beam. In certain embodiments, reducing signal energy is performed by a subtractive suppression. In some embodiments, the first microphone and the second microphone include microphones having an acoustic overload point (AOP) greater than a pre-determined sound pressure level. In certain embodiments, the pre-determined sound pressure level is 120 decibels.

[0010] According to another example embodiment of the present disclosure, the steps of the method for stereo separation and directional suppression with omni-directional microphones are stored on a machine-readable medium comprising instructions, which when implemented by one or more processors perform the recited steps.

[0011] Other example embodiments of the disclosure and aspects will become apparent from the following description taken in conjunction with the following drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0012] Embodiments are illustrated by way of example and not limitation in the figures of the accompanying drawings, in which like references indicate similar elements.

[0013] FIG. 1 is a block diagram of an example environment in which the present technology can be used.

[0014] FIG. 2 is a block diagram of an example audio device.

[0015] FIG. 3 is a block diagram of an example audio processing system.

[0016] FIG. 4 is a block diagram of an example audio processing system suitable for directional audio capture.

[0017] FIG. 5A is a block diagram showing example environment for directional audio signal capture using two omni-directional microphones.

[0018] FIG. 5B is a plot showing directional audio signals being captured with two omni-directional microphones.

[0019] FIG. 6 is a block diagram showing a module for null processing noise subtraction.

[0020] FIG. 7A is a block diagram showing coordinates used in audio zoom audio processing.

[0021] FIG. 7B is a block diagram showing coordinates used in example audio zoom audio processing.

[0022] FIG. 8 is a block diagram showing an example module for null processing noise subtraction.

[0023] FIG. 9 is a block diagram showing a further example environment in which embodiments of the present technology can be practiced.

[0024] FIG. 10 depicts plots of unprocessed and processed example audio signals.

[0025] FIG. 11 is a flow chart of an example method for stereo separation and directional suppression of audio using omni-directional microphones.

[0026] FIG. 12 is a computer system which can be used to implement example embodiment of the present technology.

DETAILED DESCRIPTION

[0027] The technology disclosed herein relates to systems and methods for stereo separation and directional suppression with omni-directional microphones. Embodiments of the present technology may be practiced with audio devices operable at least to capture and

process acoustic signals. In some embodiments, the audio devices may be hand-held devices, such as wired and/or wireless remote controls, notebook computers, tablet computers, phablets, smart phones, personal digital assistants, media players, mobile telephones, and the like. The audio devices can have radio frequency (RF) receivers, transmitters and transceivers; wired and/or wireless telecommunications and/or networking devices; amplifiers; audio and/or video players; encoders; decoders; speakers; inputs; outputs; storage devices; and user input devices. Audio devices may have input devices such as buttons, switches, keys, keyboards, trackballs, sliders, touch screens, one or more microphones, gyroscopes, accelerometers, global positioning system (GPS) receivers, and the like. The audio devices may have outputs, such as LED indicators, video displays, touchscreens, speakers, and the like.

[0028] In various embodiments, the audio devices operate in stationary and portable environments. The stationary environments can include residential and commercial buildings or structures and the like. For example, the stationary embodiments can include concert halls, living rooms, bedrooms, home theaters, conference rooms, auditoriums, business premises, and the like. Portable environments can include moving vehicles, moving persons or other transportation means, and the like.

[0029] According to an example embodiment, a method for stereo separation and directional suppression includes receiving at least a first audio signal and a second audio signal. The first audio signal can represent sound captured by a first microphone associated with a first location. The second audio signal can represent sound captured by a second microphone associated with a second location. The first microphone and the second microphone can comprise omni-directional microphones. The example method includes generating a first stereo signal by forming, based on the at least first audio signal and second audio signal, a first beam at the first location. The method can further include generating a second stereo signal by forming, based on the at least first audio signal and second audio signal, a second beam at the second location.

[0030] FIG. 1 is a block diagram of an example environment 100 in which the embodiments of the present technology can be practiced. The environment 100 of FIG. 1 can

include audio device 104 and audio sources 112, 114, and 116. The audio device can include at least a primary microphone 106a and a secondary microphone 106b.

[0031] The primary microphone 106a and the secondary microphone 106b of the audio device 104 may comprise omni-directional microphones. In some embodiments, the primary microphone 106a is located at the bottom of the audio device 104 and, accordingly, may be referred to as the bottom microphone. Similarly, in some embodiments, the secondary microphone 106b is located at the top of the audio device 104 and, accordingly, may be referred to as the top microphone. In other embodiments, the first and second microphones (and additional microphones, if any) may be located differently, including but not limited to, the microphones being located along a side of the device, e.g., separated along the side of a tablet having microphones on the side.

[0032] Some embodiments of the present disclosure utilize level differences (e.g., energy differences), phase differences, and differences in arrival times between the acoustic signals received by the two microphones 106a and 106b. Because the primary microphone 106a is closer to the audio source 112 than the secondary microphone 106b, the intensity level, for the audio signal from audio source 112 (represented graphically by 122, which may also include noise in addition to desired sounds) is higher for the primary microphone 106a, resulting in a larger energy level received by the primary microphone 106a. Similarly, because the secondary microphone 106b is closer to the audio source 116 than the primary microphone 106a, the intensity level, for the audio signal from audio source 116 (represented graphically by 126, which may also include noise in addition to desired sounds) is higher for the secondary microphone 106, resulting in a larger energy level received by the secondary microphone 106b. On the other hand, the intensity level for the audio signal from audio source 114 (represented graphically by 124, which may also include noise in addition to desired sounds) could be higher for one of the two microphones 106a and 106b, depending on, for example, its location within cones 108a and 108b.

[0033] The level differences can be used to discriminate between speech and noise in the time-frequency domain. Some embodiments may use a combination of energy level differences and differences in arrival times to discriminate between acoustic signals coming

from different directions. In some embodiments, a combination of energy level differences and phase differences is used for directional audio capture.

[0034] Various example embodiments of the present technology utilize level differences (e.g. energy differences), phase differences, and differences in arrival times for stereo separation and directional suppression of acoustic signals captured by microphones 106a and 106b. As shown in FIG. 1, a multi-directional acoustic signal provided by audio sources 112, 114, and 116 can be separated into a left channel signal of a stereo audio signal and a right channel signal of the stereo audio signal (also referred to herein as left and right stereo signals, or left and right channels of the stereo signal). The left channel of the stereo signal can be obtained by focusing on acoustic signals within cone 118a and suppressing acoustic signals outside the cone 118a. The cone 118a can cover audio sources 112 and 114. Similarly, a right channel of the stereo signal can be obtained by focusing on acoustic signals within cone 118b and suppressing acoustic signals outside cone 118b. The cone 118b can cover audio sources 114 and 116. In some embodiments of the present disclosure, audio signals coming from a site associated with user 510 (also referred to as narrator/user 510) are suppressed in both the left channel of the stereo signal and the right channel of the stereo signal. Various embodiments of the present technology can be used for capturing stereo audio when shooting video at home, during concerts, school plays, and so forth.

[0035] FIG. 2 is a block diagram of an example audio device. In some embodiments, the example audio device of FIG. 2 provides additional details for audio device 104 of FIG. 1. In the illustrated embodiment, the audio device 104 includes a receiver 210, a processor 220, the primary microphone 106a, a secondary microphone 106b, an audio processing system 230, and an output device 240. In some embodiments, the audio device 104 includes another, optional tertiary microphone 106c. The audio device 104 may include additional or different components to enable audio device 104 operations. Similarly, the audio device 104 may include fewer components that perform similar or equivalent functions to those depicted in FIG. 2.

[0036] Processor 220 may execute instructions and modules stored in a memory (not illustrated in FIG. 2) of the audio device 104 to perform functionality described herein, including noise reduction for an acoustic signal. Processor 220 may include hardware and

software implemented as a processing unit, which may process floating point and/or fixed point operations and other operations for the processor 220.

[0037] The example receiver 210 can be a sensor configured to receive a signal from a communications network. In some embodiments, the receiver 210 may include an antenna device. The signal may then be forwarded to the audio processing system 230 for noise reduction and other processing using the techniques described herein. The audio processing system 230 may provide a processed signal to the output device 240 for providing an audio output(s) to the user. The present technology may be used in one or both of the transmitting and receiving paths of the audio device 104.

[0038] The audio processing system 230 can be configured to receive acoustic signals that represent sound from acoustic source(s) via the primary microphone 106a and secondary microphone 106b and process the acoustic signals. The processing may include performing noise reduction for an acoustic signal. The example audio processing system 230 is discussed in more detail below. The primary and secondary microphones 106a, 106b may be spaced a distance apart in order to allow for detecting an energy level difference, time arrival difference, or phase difference between them. The acoustic signals received by primary microphone 106a and secondary microphone 106b may be converted into electrical signals (e.g., a primary electrical signal and a secondary electrical signal). The electrical signals may, in turn, be converted by an analog-to-digital converter (not shown) into digital signals, that represent the captured sound, for processing in accordance with some embodiments.

[0039] The output device 240 can include any device which provides an audio output to the user. For example, the output device 240 may include a loudspeaker, an earpiece of a headset or handset, or a memory where the output is stored for video/audio extraction at a later time, e.g., for transfer to computer, video disc or other media for use.

[0040] In various embodiments, where the primary and secondary microphones include omni-directional microphones that are closely-spaced (e.g., 1-2 cm apart), a beamforming technique may be used to simulate forward-facing and backward-facing directional microphones. The energy level difference may be used to discriminate between speech and noise in the time-frequency domain used in noise reduction.

[0041] FIG. 3 is a block diagram of an example audio processing system. The block diagram of FIG. 3 provides additional details for the audio processing system 230 of the example block diagram of FIG. 2. Audio processing system 230 in this example includes various modules including fast cochlea transform (FCT) 302 and 304, beamformer 310, multiplicative gain expansion 320, reverb 330, mixer 340, and zoom control 350.

[0042] FCT 302 and 304 may receive acoustic signals from audio device microphones and convert the acoustic signals into frequency range sub-band signals. In some embodiments, FCT 302 and 304 are implemented as one or more modules operable to generate one or more sub-band signals for each received microphone signal. FCT 302 and 304 can receive an acoustic signal representing sound from each microphone included in audio device 104. These acoustic signals are illustrated as signals $X_1 - X_i$, wherein X_1 represent a primary microphone signal and X_i represents the rest (e.g., $N-1$) of the microphone signals. In some embodiments, the audio processing system 230 of FIG. 3 performs audio zoom on a per frame and per sub-band basis.

[0043] In some embodiments, beamformer 310 receives frequency sub-band signals as well as a zoom indication signal. The zoom indication signal can be received from zoom control 350. The zoom indication signal can be generated in response to user input, analysis of a primary microphone signal, or other acoustic signals received by audio device 104, a video zoom feature selection, or some other data. In operation, beamformer 310 receives sub-band signals, processes the sub-band signals to identify which signals are within a particular area to enhance (or “zoom”), and provide data for the selected signals as output to multiplicative gain expansion module 320. The output may include sub-band signals for the audio source within the area to enhance. Beamformer 310 can also provide a gain factor to multiplicative gain expansion 320. The gain factor may indicate whether multiplicative gain expansion 320 should perform additional gain or reduction to the signals received from beamformer 310. In some embodiments, the gain factor is generated as an energy ratio based on the received microphone signals and components. The gain indication output by beamformer 310 may be a ratio of energy in the energy component of the primary microphone reduced by beamformer 310 to output energy of beamformer 310. Accordingly, the gain may include a boost or cancellation gain expansion factor. An example gain factor is discussed in more detail below.

[0044] Beamformer 310 can be implemented as a null processing noise subtraction (NPNS) module, multiplicative module, or a combination of these modules. When an NPNS module is used in microphones to generate a beam and achieve beamforming, the beam is focused by narrowing constraints of alpha (α) and gamma (σ). Accordingly, a beam may be manipulated by providing a protective range for the preferred direction. Exemplary beamformer 310 modules are further described in United States Patent Application serial number 14/957,447, entitled “Directional Audio Capture,” and United States Patent Application serial number 12/896,725, entitled “Audio Zoom” (issued as United States Patent number 9,210,503 on December 8, 2015), the disclosures of which is incorporated herein by reference in its entirety. Additional techniques for reducing undesired audio components of a signal are discussed in United States Patent Application serial number, 12/693,998, entitled “Adaptive Noise Reduction Using Level Cues” (issued as United States Patent number 8,718,290 on May 6, 2014), the disclosure of which is incorporated herein by reference in its entirety.

[0045] Multiplicative gain expansion module 320 can receive sub-band signals associated with audio sources within the selected beam, the gain factor from beamformer 310, and the zoom indicator signal. Multiplicative gain expansion module 320 can apply a multiplicative gain based on the gain factor received. In effect, multiplicative gain expansion module 320 can filter the beamformer signal provided by beamformer 310.

[0046] The gain factor may be implemented as one of several different energy ratios. For example, the energy ratio may include a ratio of a noise reduced signal to a primary acoustic signal received from a primary microphone, the ratio of a noise reduced signal and a detected noise component within the primary microphone signal, the ratio of a noise reduced signal and a secondary acoustic signal, or the ratio of a noise reduced signal compared to an intra level difference between a primary signal and a further signal. The gain factors may be an indication of signal strength in a target direction versus all other directions. In other words, the gain factor may be indicative of multiplicative expansions and whether these additional expansions should be performed by the multiplicative gain expansion 320. Multiplicative gain expansion 320 can output the modified signal and provide signal to reverb 330 (also referred to herein as reverb (de-reverb) 330).

[0047] Reverb 330 can receive the sub-band signals output by multiplicative gain expansion 320, as well as the microphone signals also received by beamformer 310, and perform reverberation (or dereverberation) of the sub-band signal output by multiplicative gain expansion 320. Reverb 330 may adjust a ratio of direct energy to remaining energy within a signal based on the zoom control indicator provided by zoom control 350. After adjusting the reverberation of the received signal, reverb 330 can provide the modified signal to a mixing component, e.g., mixer 340.

[0048] The mixer 340 can receive the reverberation adjusted signal and mix the signal with the signal from the primary microphone. In some embodiments, mixer 340 increases the energy of the signal appropriately when audio is present in the frame and decreases the energy when there is little audio energy present in the frame.

[0049] FIG. 4 is a block diagram illustrating an audio processing system 400, according to another example embodiment. The audio processing system 400 can include audio zoom audio (AZA), a subsystem augmented with a source estimation subsystem 430. The example AZA subsystem includes limiters 402a, 402b, and 402c, along with various other modules including FCT 404a, 404b, and 404c, analysis 406, zoom control 410, signal modifier 412, plus variable amplifier 418 and a limiter 420. The source estimation subsystem 430 can include a source direction estimator (SDE) 408 (also referred to variously as SDE module 408 or as a target estimator), a gain (module) 416, and an automatic gain control (AGC) (module) 414. In various embodiments, the audio processing system 400 processes acoustic audio signal from microphones 106a, 106b, and optionally a third microphone, 106c.

[0050] In various embodiments, SDE module 408 is operable to localize a source of sound. The SDE module 408 is operable to generate cues based on correlation of phase plots between different microphone inputs. Based on the correlation of the phase plots, the SDE module 408 is operable to compute a vector of salience estimates at different angles. Based on the salience estimates, the SDE module 408 can determine a direction of the source. In other words, a peak in the vector of salience estimates is an indication of direction of a source in a particular direction. At the same time, sources of diffused nature, i.e., non-directional, are represented by poor salience estimates at all the angles. The SDE module 408 can rely upon the cues (estimates of salience) to improve the performance of a directional audio

solution, which is carried out by the analysis module 406, signal modifier 412, and zoom control 410. In some embodiments, the signal modifier 412 includes modules analogous or similar to beamformer 310, multiplicative gain expansion module 320, reverb module 330, and mixer module 340 as shown for audio system 230 in FIG. 3.

[0051] In some embodiments, estimates of salience are used to localize the angle of the source in the range of 0 to 360 degrees in a plane parallel to the ground, when, for example, the audio device 104 is placed on a table top. The estimates of salience can be used to attenuate/amplify the signals at different angles as required by the customer. The characterization of these modes may be driven by a SDE salience parameter. Example AZA and SDE subsystems are described further in United States Patent Application serial number 14/957,447, entitled “Directional Audio Capture,” the disclosure of which is incorporated herein by reference in its entirety.

[0052] FIG. 5A illustrates an example environment 500 for directional audio signal capture using two omni-directional microphones. The example environment 500 can include audio device 104, primary microphone 106a, secondary microphone 106b, a user 510 (also referred to as narrator 510) and a second sound source 520 (also referred to as scene 520). Narrator 510 can be located proximate to primary microphone 106a. Scene 520 can be located proximate to secondary microphone 106b. The audio processing system 400 may provide a dual output including a first signal and a second signal. The first signal can be obtained by focusing on a direction associated with narrator 510. The second signal can be obtained by focusing on a direction associated with scene 520. SDE module 408 (an example of which is shown in FIG. 4) can provide a vector of salience estimates to localize a direction associated with target sources, for example narrator 510 and scene 520. FIG. 5B illustrates a directional audio signal captured using two omni-directional microphones. As target sources or audio device change positions, SDE module 408 (e.g., in the system in FIG. 4) can provide an updated vector of salience estimates to allow audio processing system 400 to keep focusing on the target sources.

[0053] FIG. 6 shows a block diagram of an example NPNS module 600. The NPNS module 600 can be used as a beamformer module in audio processing systems 230 or 400. NPNS module 600 can include analysis modules 602 and 606 (e.g., for applying coefficients

σ_1 and σ_2 respectively), adaptation modules 604 and 608 (e.g., for adapting the beam based on coefficients α_1 and α_2) and summing modules 610, 612, and 614. The NPNS module 600 may provide gain factors based on inputs from a primary microphone, a secondary microphone, and, optionally, a tertiary microphone. Exemplary NPNS modules are further discussed in United States Patent Application serial number, 12/215,980, entitled “System and Method for Providing Noise Suppression Utilizing Null Processing Noise Subtraction” (issued as United States Patent number 9,185,487 on November 10, 2015), the disclosure of which is incorporated herein by reference in its entirety.

[0054] In the example in FIG. 6, the NPNS module 600 is configured to adapt to a target source. Attenuation coefficients σ_1 and σ_2 can be adjusted based on a current direction of a target source as either the target source or the audio device moves.

[0055] FIG. 7A shows an example coordinate system 710 used for determining the source direction in the AZA subsystem. Assuming that the largest side of the audio device 104 is parallel to the ground when, for example, the audio device 104 is placed on a table top, X axis of coordinate system 710 is directed from the bottom to the top of audio device 104. Y axis of coordinate system 710 is directed in such a way that XY plane is parallel to the ground.

[0056] In various embodiments of the present disclosure, the coordinate system 710 used in AZA is rotated to adapt for providing a stereo separation and directional suppression of received acoustic signals. FIG. 7B shows a rotated coordinate system 720 as related to audio device 104. The audio device 104 is oriented in such way that the largest side of the audio device is orthogonal (e.g., perpendicular) to the ground and the longest edge of the audio device is parallel to the ground when, for example, the audio device 104 is held when recording a video. The X axis of coordinate system 720 is directed from the top to the bottom of audio device 104. The Y axis of coordinate system 720 is directed in such a way that XY plane is parallel to the ground.

[0057] According to various embodiments of the present disclosure, at least two channels of a stereo signal (also referred to herein as left and right channel stereo (audio) signals, and a left stereo signal and a right stereo signal) are generated based on acoustic signals captured

by two or more omni-directional microphones. In some embodiments, the omni-directional microphones include the primary microphone 106a and the secondary microphone 106b. As shown in FIG. 1, the left (channel) stereo signal can be provided by creating a first target beam on the left. The right (channel) stereo signal can be provided by creating a second target beam on the right. According to various embodiments, the directions for the beams are fixed and maintained as a target source or audio device changes position. Fixing the directions for the beams allows obtaining a natural stereo effect (having left and right stereo channels) that can be heard by a user. By fixing the direction, the natural stereo effect can be heard when an object moves across the field of view, from one side to the other, for example, a car moving across a movie screen. In some embodiments, the directions for the beams are adjustable but are maintained fixed during beamforming.

[0058] According to some embodiments of the present disclosure, NPNS module 600 (in the example in FIG. 6) is modified so it does not adapt to a target source. A modified NPNS module 800 is shown in FIG. 8. Components of NPNS module 800 are analogous to elements of NPNS module 600 except that the modules 602 and 606 in FIG. 6 are replaced with modules 802 and 806. Unlike in the example in FIG. 6, values for coefficients σ_1 and σ_2 in the example embodiment in FIG. 8 are fixed during forming the beams for creation of stereo signals. By preventing adaptation to the target source, the direction for beams remains fixed, ensuring that the left stereo signal and the right stereo signal do not overlap as sound source(s) or the audio device change position. In some embodiments, the attenuation coefficients σ_1 and σ_2 are determined by calibration and tuning.

[0059] FIG. 9 is an example environment 900, in which example methods for stereo separation and directional suppression can be implemented. The environment 900 includes audio device 104 and audio sources 910, 920, and 930. In some embodiments, the audio device 104 includes two omni-directional microphones 106a and 106b. The primary microphone 106a is located at the bottom of the audio device 104 and the secondary microphone 106b is located at the top of the audio device 104, in this example. When the audio device 104 is oriented to record video, for example, in the direction of audio source 910, the audio processing system of the audio device may be configured to operate in a stereo recording mode. A left channel stereo signal and a right channel stereo signal may be

generated based on inputs from two or more omni-directional microphones by creating a first target beam for audio on the left and a second target beam for audio on the right. The directions for the beams are fixed, according to various embodiments.

[0060] In certain embodiments, only two omni-directional microphones 106a and 106b are used for stereo separation. Using two omni-directional microphones 106a and 106b, one on each end of the audio device, a clear separation between the left side and the right side can be achieved. For example, the secondary microphone 106b is closer to the audio source 920 (at the right in the example in FIG. 9) and receives the wave from the audio source 920 shortly before the primary microphone 106a. The audio source can be then triangulated based on the spacing between the microphones 106a and 106b and the difference in arrival times at the microphones 106a and 106b. However, this exemplary two-microphone system may not distinguish between acoustic signals coming from a scene side (where the user is directing the camera of audio device) and acoustic signals coming from the user side (e.g., opposite the scene side). In the example embodiment shown in FIG. 9, the audio sources 910 and 930 are equidistant from microphones 106a and 106b. From the top view of an audio device 104, the audio source 910 is located in front of the audio device 104 at scene side and the audio source 930 is located behind the audio device at the user side. The microphones 106a and 106b receive the same acoustic signal from the audio source 910 and the same acoustic signal from audio source 930 since there is no delay in the time of arrival between the microphones, in this example. This means that, when using only the two microphones 106a and 106b, locations of audio sources 910 and 930 cannot be distinguished, in this example. Thus, for this example, it cannot be determined which of the audio sources 910 and 930 is located in front and which of the audio sources 910 and 930 is located behind the audio device.

[0061] In some embodiments, an appropriately-placed third microphone can be used to improve differentiation of the scene (audio device camera's view) direction from the direction behind the audio device. Using a third microphone (for example, the tertiary microphone 106c shown in FIG. 9) may help providing a more robust stereo sound. Input from the third microphone can also allow for better attenuation of unwanted content such as speech of the user holding the audio device and people behind the user. In various embodiments, the three microphones 106a, 106b, and 106c are not all located in a straight

line, so that various embodiments can provide a full 360 degree picture of sounds relative to a plane on which the three microphones are located.

[0062] In some embodiments, the microphones 106a, 106b, and 106c include high AOP microphones. The AOP microphones can provide robust inputs for beamforming in loud environments, for example, concerts. Sound levels at some concerts are capable of exceeding 120dB with peak levels exceeding 120dB considerably. Traditional omni-directional microphones may saturate at these sound levels making it impossible to recover any signal captured by the microphone. High AOP microphones are designed for a higher overload point as compared to traditional microphones and, therefore, are capable of capturing an accurate signal under significantly louder environments when compared to traditional microphones. Combining the technology of high AOP microphones with the methods for stereo separation and directional suppression using omni-directional microphones (e.g., using high AOP omni-directional microphones for the combination) according to various embodiments of the present disclosure, can enable users to capture a video providing a much more realistic representation of their experience during, for example, a concert.

[0063] FIG. 10 shows a depiction 1000 of example plots of example directional audio signals. Plot 1010 represents an unprocessed directional audio signal captured by a secondary microphone 106b. Plot 1020 represents an unprocessed directional audio signal captured by a primary microphone 106a. Plot 1030 represents a right channel stereo audio signal obtained by forming a target beam on the right. Plot 1040 represents a left channel stereo audio signal obtained by forming a target beam on the left. Plots 1030 and 1040, in this example, show a clear stereo separation of the unprocessed audio signal depicted in plots 1010 and 1020.

[0064] FIG. 11 is a flow chart showing steps of a method for stereo separation and directional suppression, according to an example embodiment. Method 1100 can commence, in block 1110, with receiving at least a first audio signal and a second audio signal. The first audio signal can represent sound captured by a first microphone associated with a first location. The second audio signal can represent sound captured by a second microphone associated with a second location. The first microphone and the second microphone may comprise omni-directional microphones. In some embodiments, the first microphone and the

second microphone comprise microphones with high AOP. In some embodiments, the distance between the first and the second microphones is limited by size of a mobile device.

[0065] In block 1120, a first stereo signal (e.g., a first channel signal of a stereo audio signal) can be generated by forming a first beam at the first location, based on the first audio signal and the second audio signal. In block 1130, a second stereo signal (e.g., a second channel signal of the stereo audio signal) can be generated by forming a second beam at the second location based on the first audio signal and the second audio signal.

[0066] FIG. 12 illustrates an example computer system 1200 that may be used to implement some embodiments of the present invention. The computer system 1200 of FIG. 12 may be implemented in the contexts of the likes of computing systems, networks, servers, or combinations thereof. The computer system 1200 of FIG. 12 includes one or more processor unit(s) 1210 and main memory 1220. Main memory 1220 stores, in part, instructions and data for execution by processor unit(s) 1210. Main memory 1220 stores the executable code when in operation, in this example. The computer system 1200 of FIG. 12 further includes a mass data storage 1230, portable storage device 1240, output devices 1250, user input devices 1260, a graphics display system 1270, and peripheral devices 1280.

[0067] The components shown in FIG. 12 are depicted as being connected via a single bus 1290. The components may be connected through one or more data transport means. Processor unit(s) 1210 and main memory 1220 is connected via a local microprocessor bus, and the mass data storage 1230, peripheral devices 1280, portable storage device 1240, and graphics display system 1270 are connected via one or more input/output (I/O) buses.

[0068] Mass data storage 1230, which can be implemented with a magnetic disk drive, solid state drive, or an optical disk drive, is a non-volatile storage device for storing data and instructions for use by processor unit(s) 1210. Mass data storage 1230 stores the system software for implementing embodiments of the present disclosure for purposes of loading that software into main memory 1220.

[0069] Portable storage device 1240 operates in conjunction with a portable non-volatile storage medium, such as a flash drive, floppy disk, compact disk, digital video disc, or Universal Serial Bus (USB) storage device, to input and output data and code to and from the

computer system 1200 of FIG. 12. The system software for implementing embodiments of the present disclosure is stored on such a portable medium and input to the computer system 1200 via the portable storage device 1240.

[0070] User input devices 1260 can provide a portion of a user interface. User input devices 1260 may include one or more microphones, an alphanumeric keypad, such as a keyboard, for inputting alphanumeric and other information, or a pointing device, such as a mouse, a trackball, stylus, or cursor direction keys. User input devices 1260 can also include a touchscreen. Additionally, the computer system 1200 as shown in FIG. 12 includes output devices 1250. Suitable output devices 1250 include speakers, printers, network interfaces, and monitors.

[0071] Graphics display system 1270 include a liquid crystal display (LCD) or other suitable display device. Graphics display system 1270 is configurable to receive textual and graphical information and processes the information for output to the display device.

[0072] Peripheral devices 1280 may include any type of computer support device to add additional functionality to the computer system.

[0073] The components provided in the computer system 1200 of FIG. 12 are those typically found in computer systems that may be suitable for use with embodiments of the present disclosure and are intended to represent a broad category of such computer components that are well known in the art. Thus, the computer system 1200 of FIG. 12 can be a personal computer (PC), hand held computer system, telephone, mobile computer system, workstation, tablet, phablet, mobile phone, server, minicomputer, mainframe computer, wearable, or any other computer system. The computer may also include different bus configurations, networked platforms, multi-processor platforms, and the like. Various operating systems may be used including UNIX, LINUX, WINDOWS, MAC OS, PALM OS, QNX ANDROID, IOS, CHROME, TIZEN, and other suitable operating systems.

[0074] The processing for various embodiments may be implemented in software that is cloud-based. In some embodiments, the computer system 1200 is implemented as a cloud-based computing environment, such as a virtual machine operating within a computing cloud. In other embodiments, the computer system 1200 may itself include a cloud-based computing

environment, where the functionalities of the computer system 1200 are executed in a distributed fashion. Thus, the computer system 1200, when configured as a computing cloud, may include pluralities of computing devices in various forms, as will be described in greater detail below.

[0075] In general, a cloud-based computing environment is a resource that typically combines the computational power of a large grouping of processors (such as within web servers) and/or that combines the storage capacity of a large grouping of computer memories or storage devices. Systems that provide cloud-based resources may be utilized exclusively by their owners or such systems may be accessible to outside users who deploy applications within the computing infrastructure to obtain the benefit of large computational or storage resources.

[0076] The cloud may be formed, for example, by a network of web servers that comprise a plurality of computing devices, such as the computer system 1200, with each server (or at least a plurality thereof) providing processor and/or storage resources. These servers may manage workloads provided by multiple users (e.g., cloud resource customers or other users). Typically, each user places workload demands upon the cloud that vary in real-time, sometimes dramatically. The nature and extent of these variations typically depends on the type of business associated with the user.

[0077] The present technology is described above with reference to example embodiments. Therefore, other variations upon the example embodiments are intended to be covered by the present disclosure.

WHAT IS CLAIMED IS:

1. A method for providing stereo separation and directional suppression, the method comprising:

configuring a processor to receive at least a first audio signal and a second audio signal, the first audio signal representing sound captured by a first microphone associated with a first location and the second audio signal representing sound captured by a second microphone associated with a second location, the first microphone and the second microphone comprising omni-directional microphones of a mobile device, the distance between the first microphone and the second microphone being limited by the size of the mobile device;

configuring the processor to generate a first channel signal of a stereo audio signal by forming, based on the first audio signal and the second audio signal, a first beam at the first location; and

configuring the processor to generate a second channel signal of the stereo audio signal by forming, based on the first audio signal and the second audio signal, a second beam at the second location.

2. The method of claim 1, wherein the first microphone is located at the top of the mobile device and the second microphone is located at the bottom of the mobile device.

3. The method of claim 1, wherein a first direction, associated with the first beam, and a second direction, associated with the second beam, are each fixed relative to a line between the first location and the second location.

4. The method of claim 3, wherein the first direction remains fixed even if an audio source at the first location moves from the first location to the second location.

5. The method of claim 4, wherein the second direction remains fixed even if another audio source at the second location moves from the second location to the first location.

6. The method of claim 1, wherein:
forming the first beam includes reducing signal energy of acoustic signal components associated with sources off the first beam; and
forming the second beam includes reducing signal energy of acoustic signal components associated with further sources off the second beam.
7. The method of claim 6, wherein reducing energy components is performed by a subtractive suppression.
8. The method of claim 1, wherein a first audio source at the first location is associated with the first microphone by the first audio source being located closer to the first microphone.
9. The method of claim 8, wherein a second audio source at the second location is associated with the second microphone by the second audio source being located closer to the second microphone.
10. The method of claim 1, wherein the first microphone and the second microphone include microphones having an acoustic overload point (AOP) higher than a pre-determined sound pressure level.
11. The method of claim 10, wherein the pre-determined sound pressure level is 120 decibels.
12. The method of claim 6, further comprising configuring the processor to receive at least one other acoustic signal representing sound captured by another microphone associated with another location, the other microphone comprising an omni-directional microphone, and the forming the first beam and the forming the second beam each being further based on the at least one other acoustic signal.
13. The method of claim 12, wherein the other microphone is located at a position on the mobile device other than on a line between the first microphone and the second microphone.

14. A system for stereo separation and directional suppression, the system comprising:

at least one processor; and

a memory communicatively coupled with the at least one processor, the memory storing instructions, which when executed by the at least one processor, perform a method comprising:

receiving at least a first audio signal and a second audio signal, the first audio signal representing sound captured by a first microphone associated with a first location and the second audio signal representing sound captured by a second microphone associated with a second location, the first microphone and the second microphone comprising omni-directional microphones of a mobile device, the distance between the first microphone and the second microphone being limited by the size of the mobile device;

generating a first channel signal of a stereo audio signal by forming, based on the first audio signal and the second audio signal, a first beam at the first location; and

generating a second channel signal of the stereo audio signal by forming, based on the first audio signal and the second audio signal, a second beam at the second location.

15. The system of claim 14, wherein the first microphone is located at the top of the mobile device and the second microphone is located at the bottom of the mobile device.

16. The system of claim 14, wherein a first direction associated with the first beam and a second direction associated with the second beam are fixed relative to a line between the first location and the second location.

17. The system of claim 14, wherein:

forming the first beam includes reducing signal energy of acoustic signal components associated with sources off the first beam; and

forming the second beam includes reducing signal energy of acoustic signal components associated with further sources off the second beam.

18. The system of claim 17, wherein reducing energy components is performed by a subtractive suppression.

19. The system of claim 17, wherein the method further comprises receiving at least one other acoustic signal representing sound captured by another microphone associated with another location, the other microphone comprising an omni-directional microphone, and the forming the first beam and the forming the second beam each being further based on the other acoustic signal.

20. The system of claim 19, wherein the other microphone is located at a position on the mobile device other than on a line between the first microphone and the second microphone.

21. The system of claim 14, wherein the first audio source at the first location is associated with the first microphone by the first audio source being located closer to the first microphone, and the second audio source at the second location is associated with the second microphone by the second audio source being located closer to the second microphone.

22. The system of claim 14, wherein the first microphone and the second microphone include microphones having an acoustic overload point (AOP) greater than a pre-determined sound pressure level.

23. The system of claim 22, wherein the pre-determined sound pressure level is 120 decibels.

24. A non-transitory computer-readable storage medium having embodied thereon instructions, which when executed by at least one processor, perform steps of a method for stereo separation and directional suppression, the method comprising:

receiving at least a first audio signal and a second audio signal, the first audio signal representing sound captured by a first microphone associated with a first location and the second audio signal representing sound captured by a second microphone associated with a second location, the first microphone and the second microphone comprising omnidirectional microphones of a mobile device, the distance between the first microphone and the second microphone being limited by the size of the mobile device;

generating a first channel signal of a stereo audio signal by forming, based on the first audio signal and the second audio signal, a first beam at the first location; and

generating a second channel signal of the stereo audio signal by forming, based on the first audio signal and the second audio signal, a second beam at the second location.

1/12

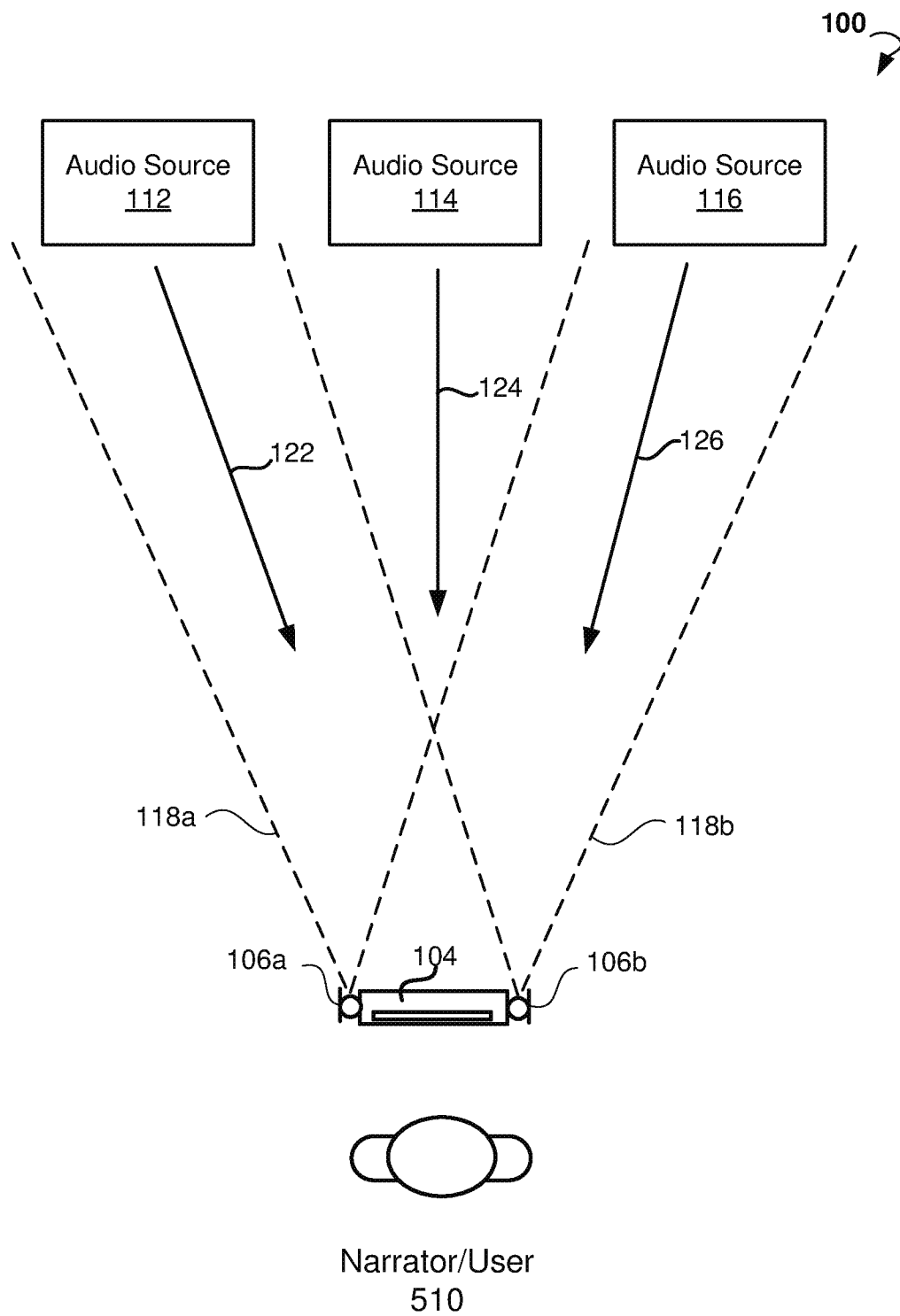


FIG. 1

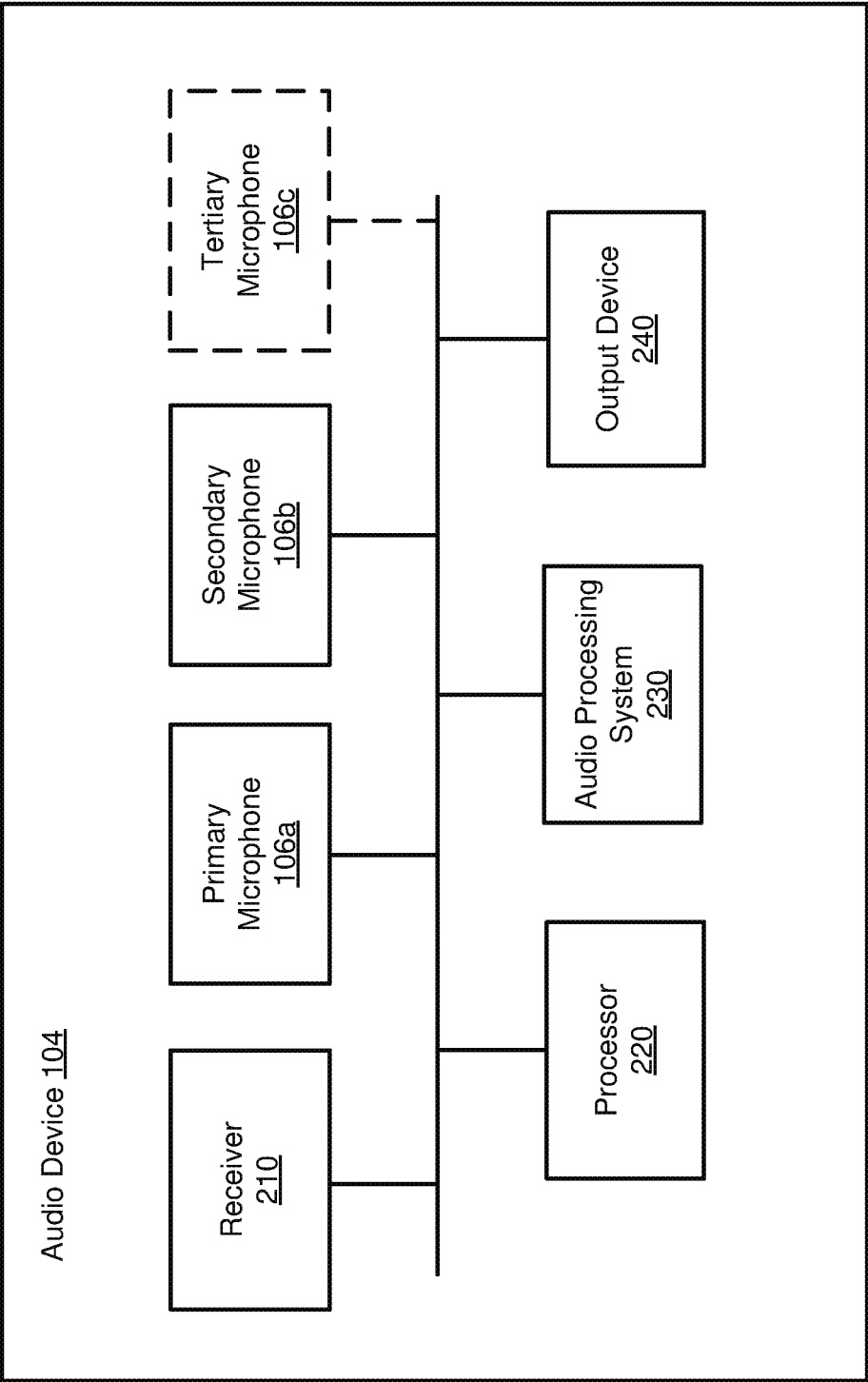


FIG. 2

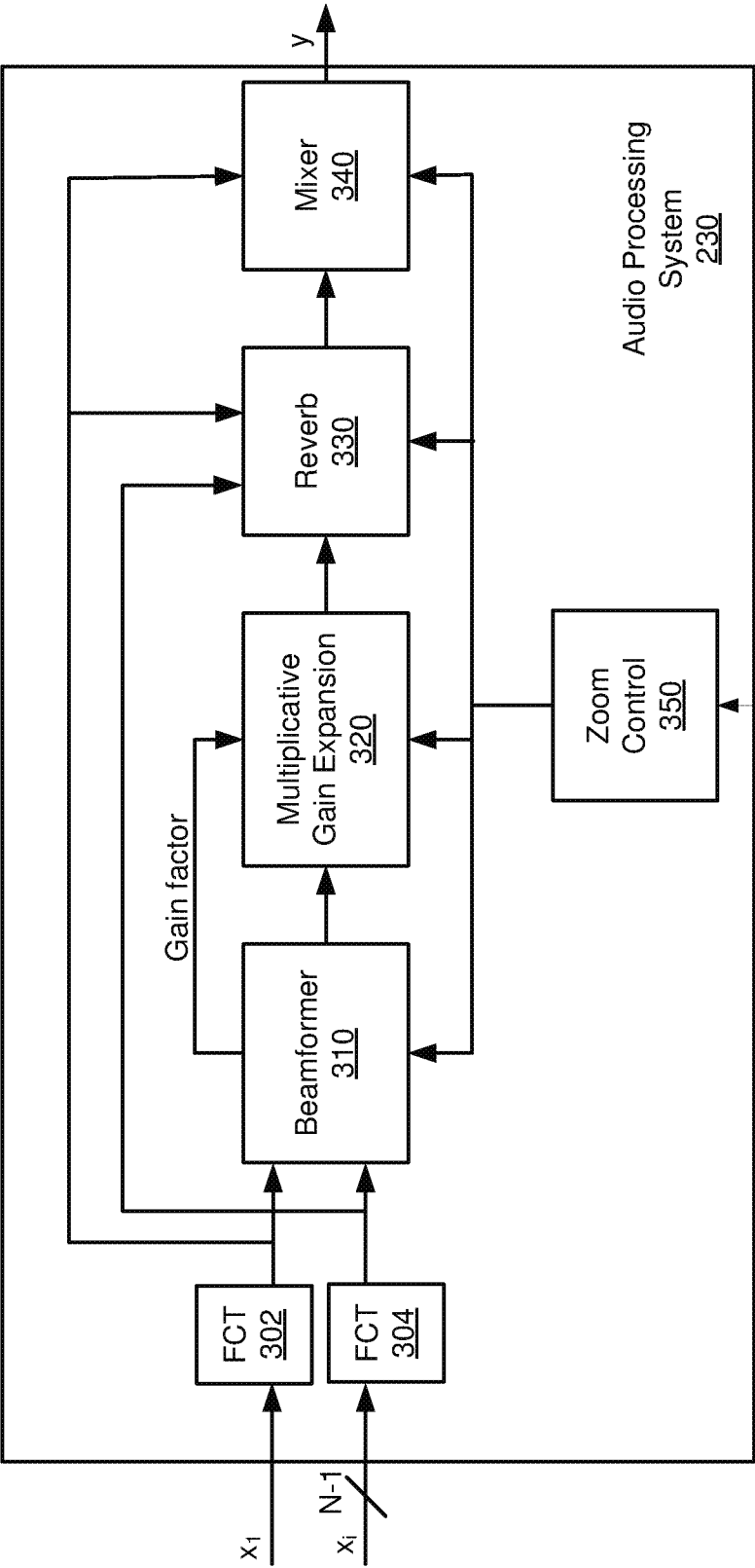


FIG. 3

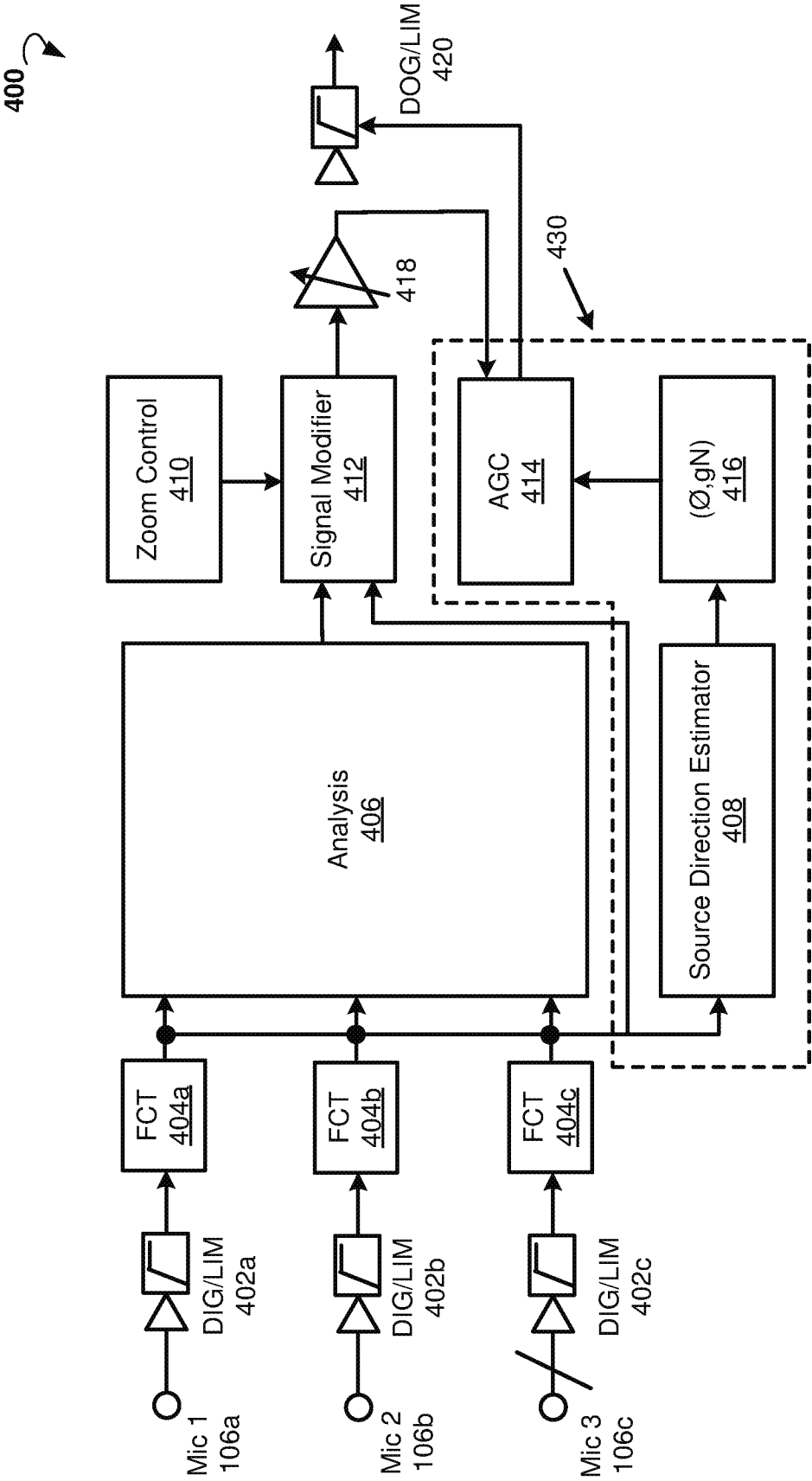


FIG. 4

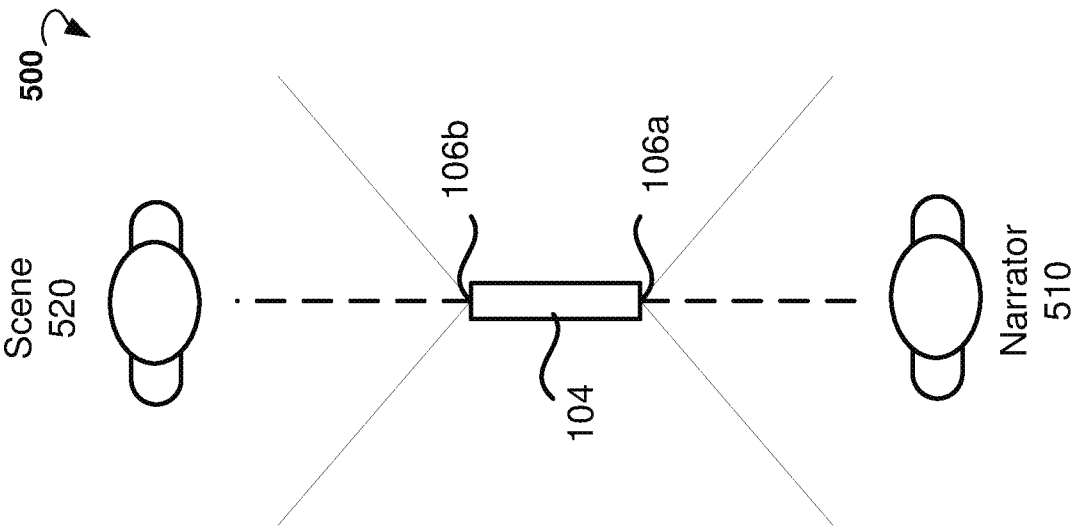


FIG. 5A

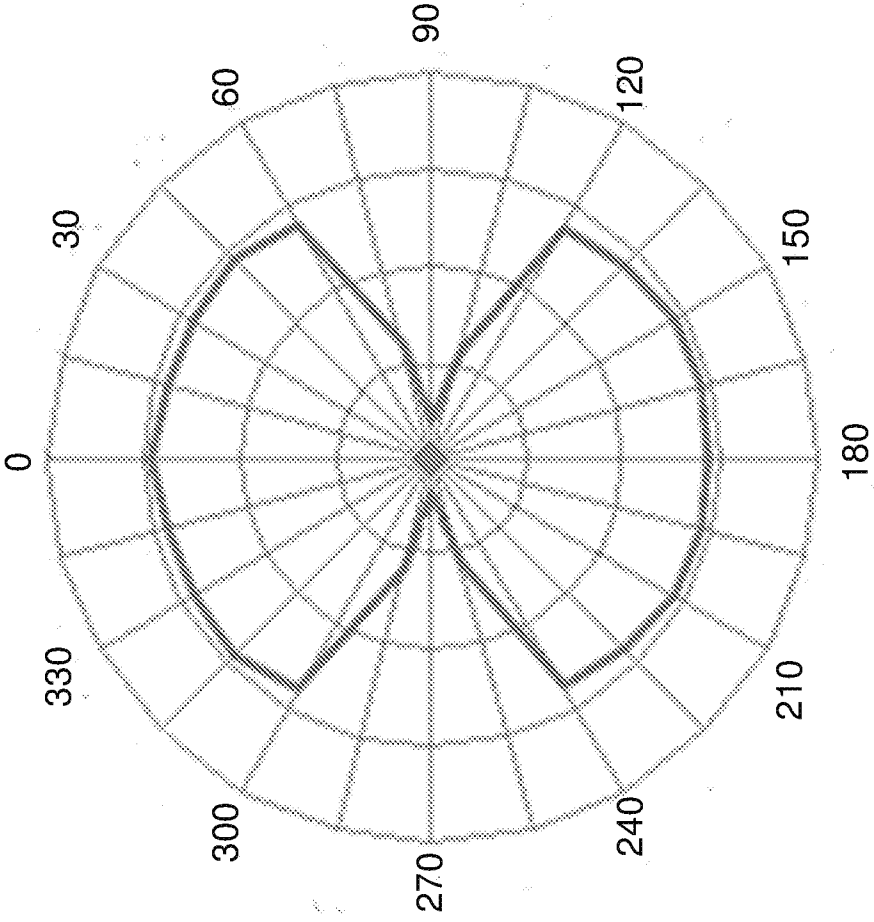


FIG. 5B

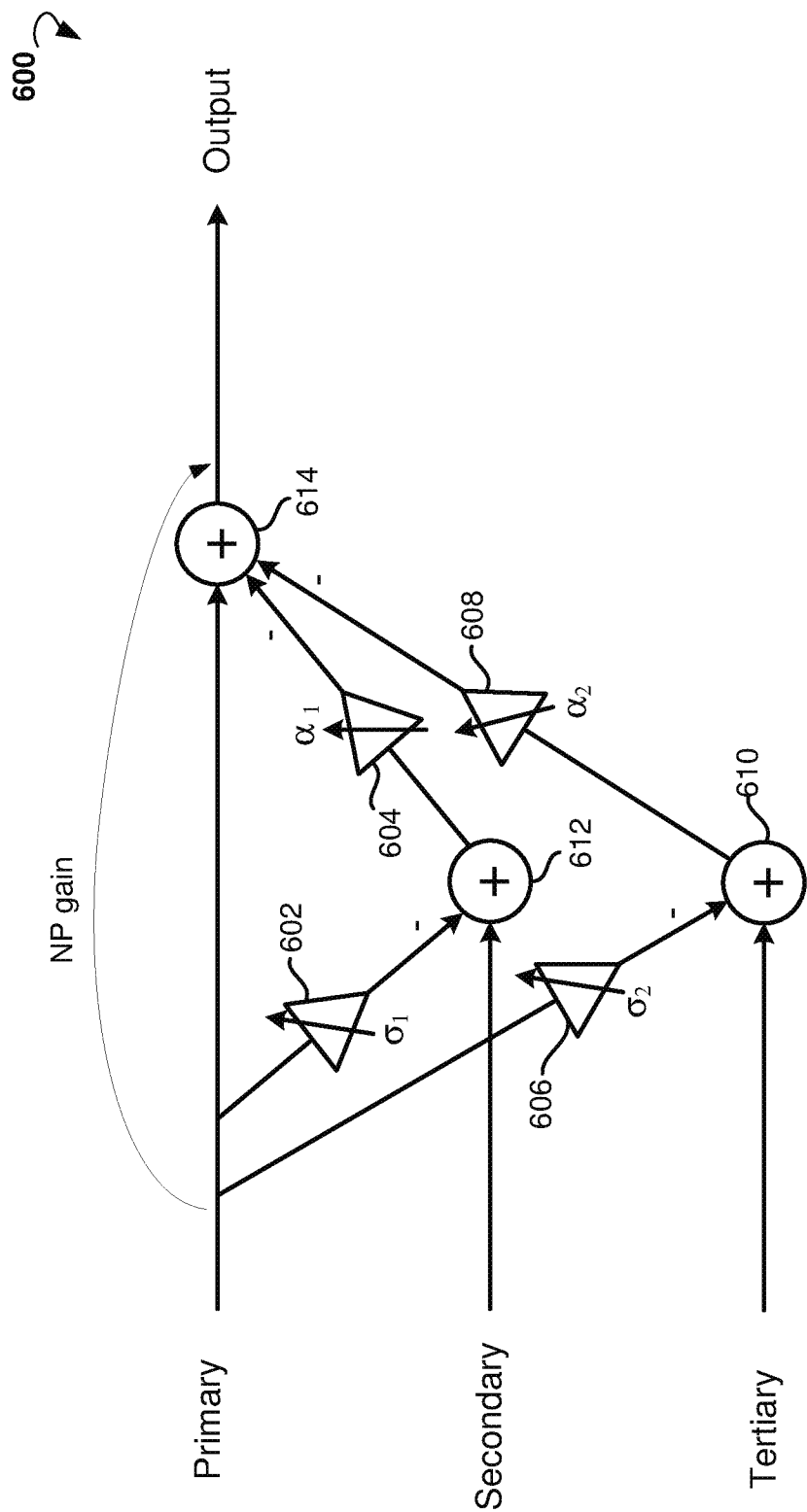


FIG. 6

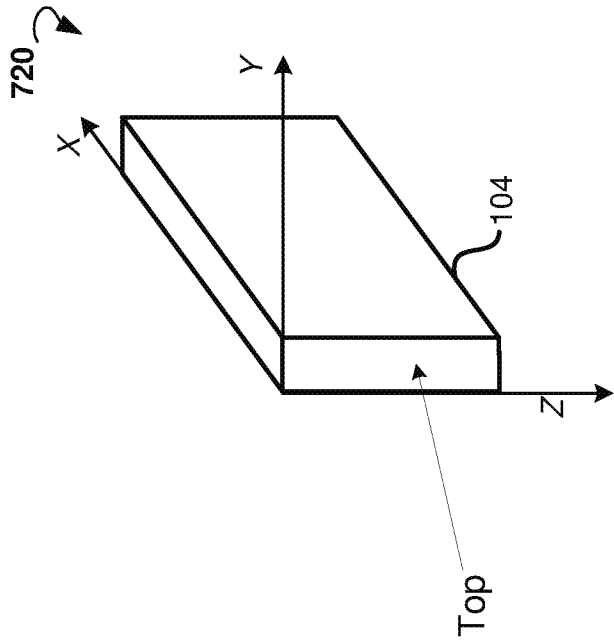


FIG. 7A

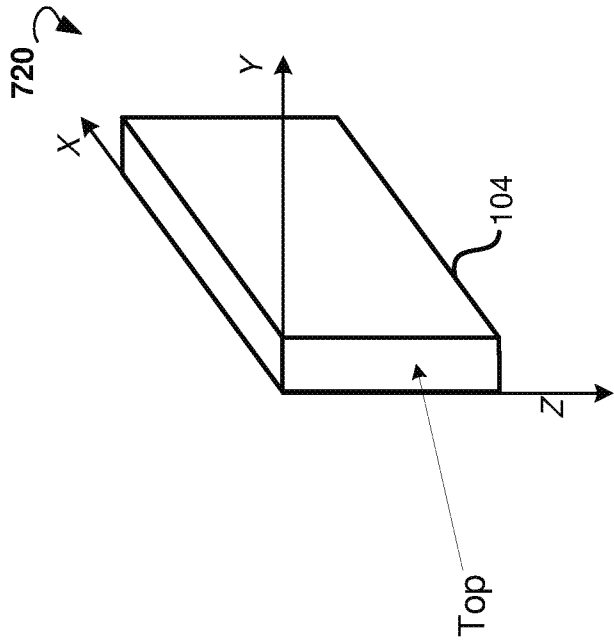


FIG. 7B

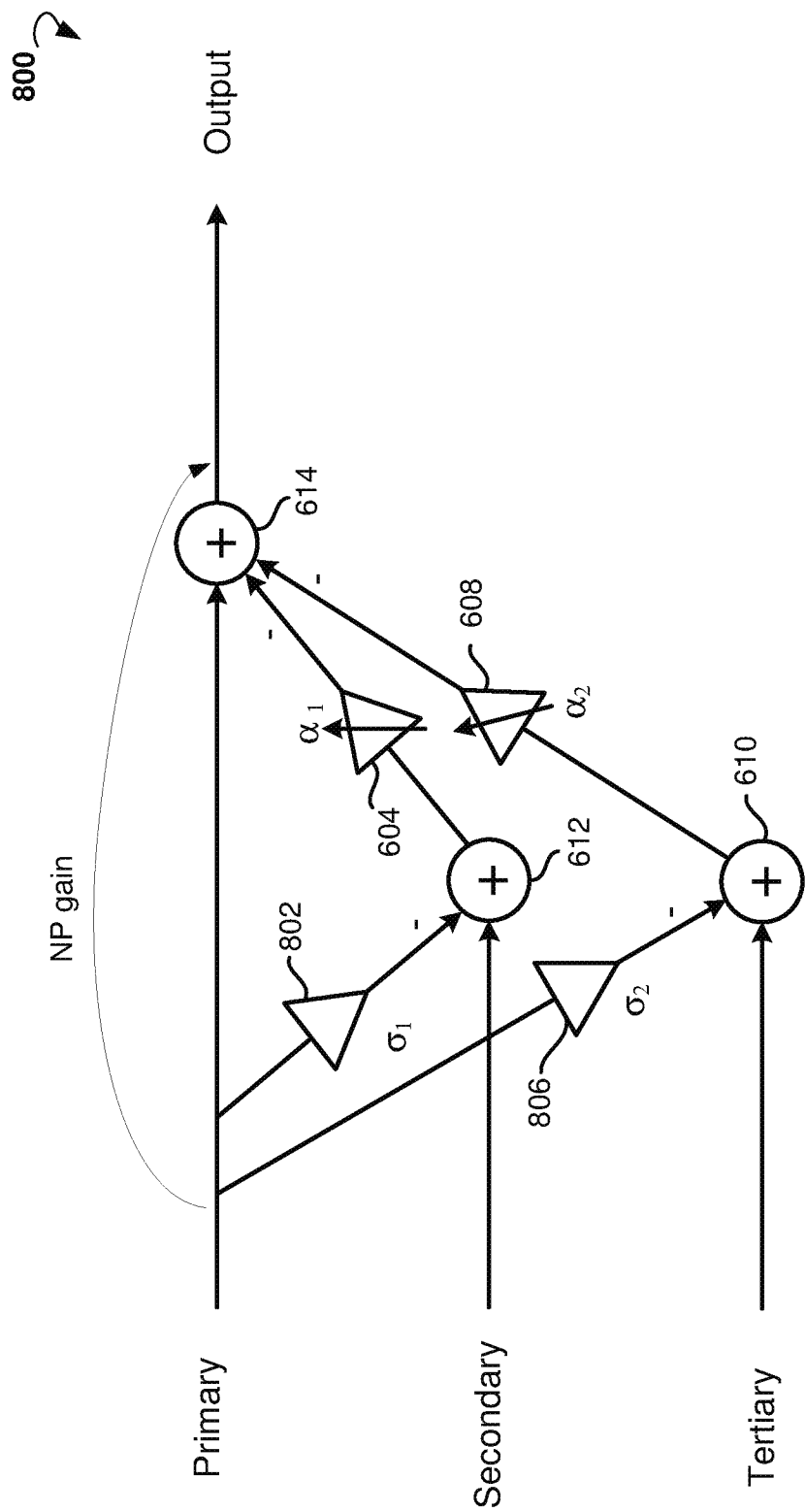


FIG. 8

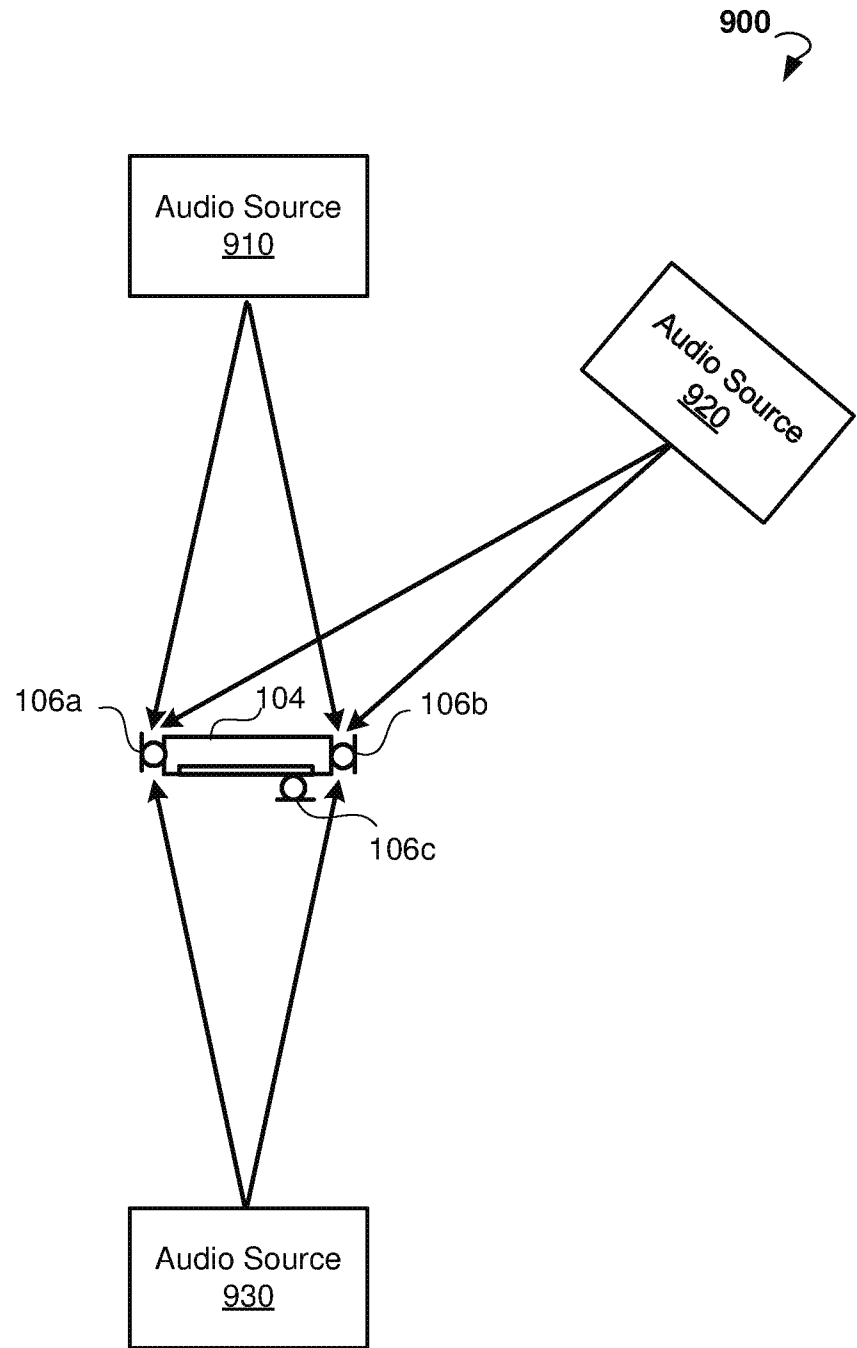


FIG. 9

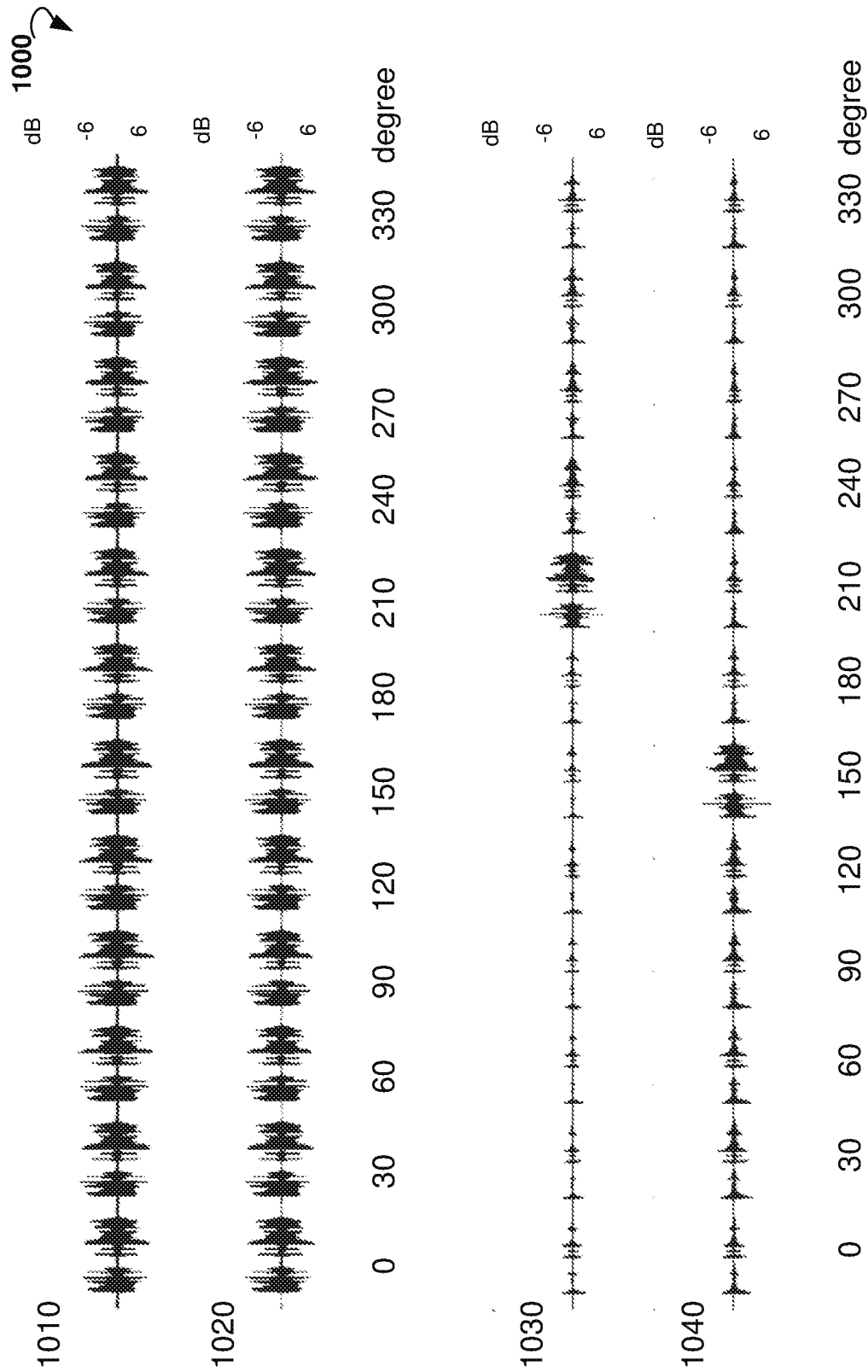
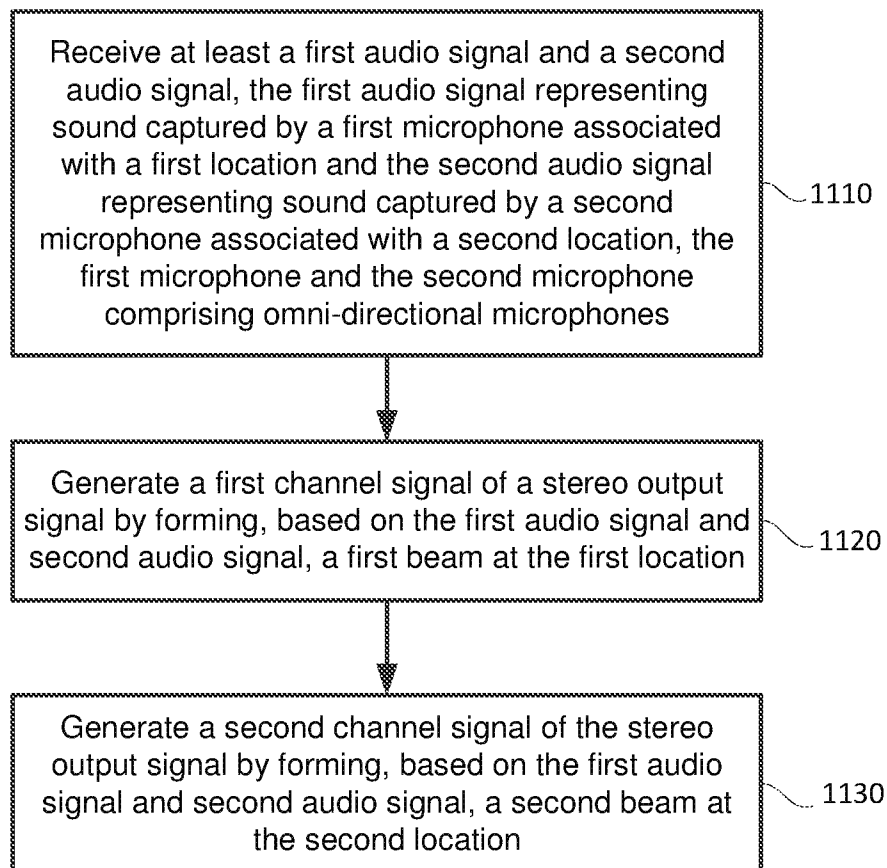
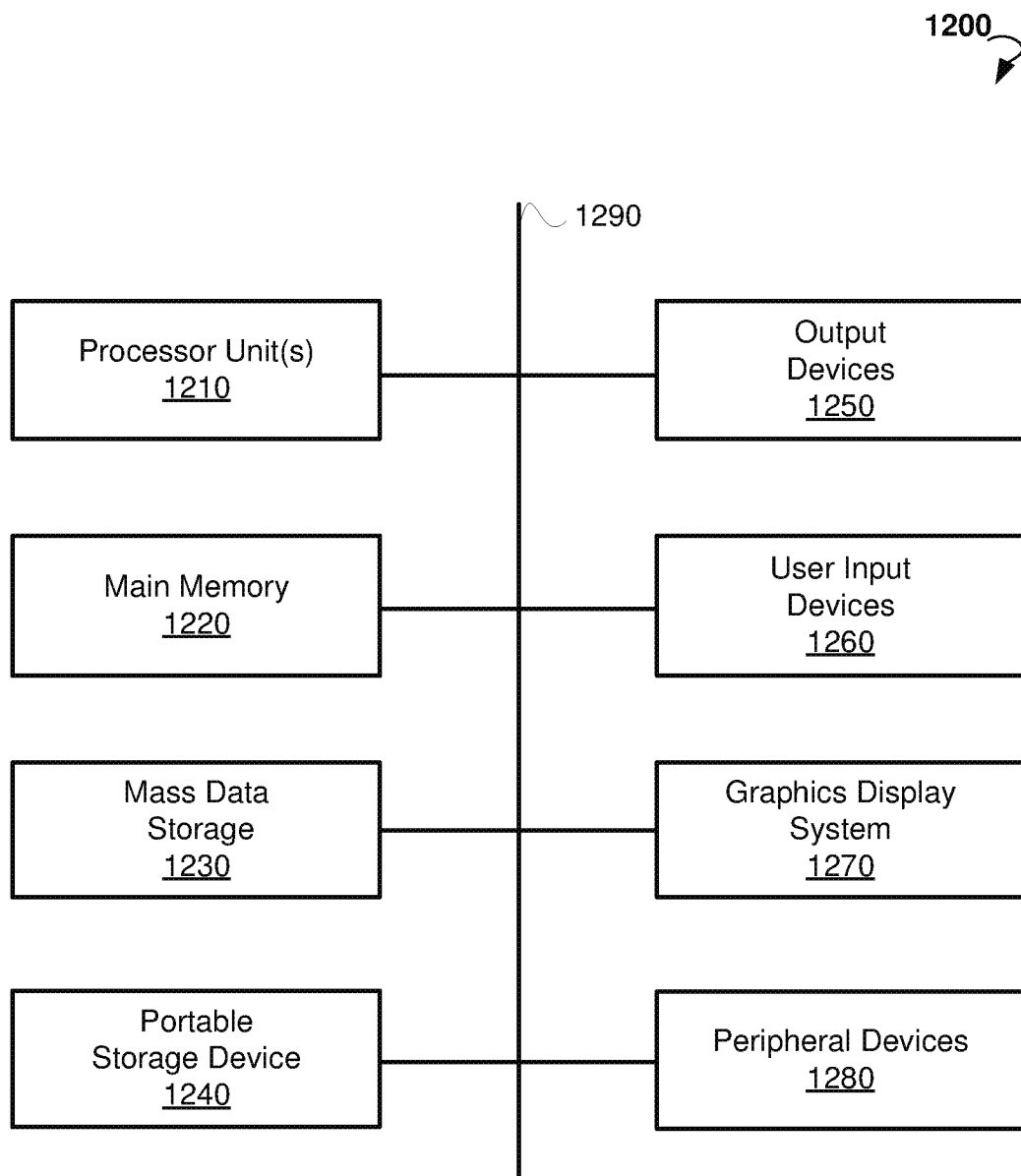


FIG. 10

11/12

1100
**FIG. 11**

12/12

**FIG. 12**

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2017/030220

A. CLASSIFICATION OF SUBJECT MATTER
INV. H04R3/00
ADD.

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
H04R

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

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

EPO-Internal, WPI Data, INSPEC, COMPENDEX, IBM-TDB

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2014/126726 A1 (HEIMAN ARIE [IL] ET AL) 8 May 2014 (2014-05-08) paragraphs [0023] - [0028]; figures 1-5 -----	1-24
X	US 2012/019689 A1 (ZUREK ROBERT [US] ET AL) 26 January 2012 (2012-01-26) figures 3-6 -----	1-24
X	US 2012/013768 A1 (ZUREK ROBERT [US] ET AL) 19 January 2012 (2012-01-19) figures 4,7 -----	1-24
A	US 2011/129095 A1 (AVENDANO CARLOS [US] ET AL) 2 June 2011 (2011-06-02) figures 1-7 ----- -/--	1-24



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search

4 August 2017

Date of mailing of the international search report

30/08/2017

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Radomirescu, B-M

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2017/030220

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2011/182436 A1 (MURGIA CARLO [US] ET AL) 28 July 2011 (2011-07-28) figures 2-5 -----	1-24
A	US 2016/094910 A1 (VALLABHAN HARINARAYANAN ERUMBI [IN] ET AL) 31 March 2016 (2016-03-31) figures 2,3,8,10 -----	1-24
A	US 2009/323982 A1 (SOLBACH LUDGER [US] ET AL) 31 December 2009 (2009-12-31) figures 3-9 -----	1-24

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2017/030220

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2014126726 A1	08-05-2014	CN 103905960 A EP 2731352 A2 JP 2014112830 A KR 20140061256 A US 2014126726 A1	02-07-2014 14-05-2014 19-06-2014 21-05-2014 08-05-2014
US 2012019689 A1	26-01-2012	CN 103026734 A EP 2599328 A1 US 2012019689 A1 WO 2012018445 A1	03-04-2013 05-06-2013 26-01-2012 09-02-2012
US 2012013768 A1	19-01-2012	CN 103004233 A EP 2594087 A1 US 2012013768 A1 WO 2012009107 A1	27-03-2013 22-05-2013 19-01-2012 19-01-2012
US 2011129095 A1	02-06-2011	FI 20125600 A JP 5855571 B2 JP 2013513306 A KR 20120101457 A TW 201143475 A US 8903721 B1 US 2011129095 A1 WO 2011068901 A1	01-06-2012 09-02-2016 18-04-2013 13-09-2012 01-12-2011 02-12-2014 02-06-2011 09-06-2011
US 2011182436 A1	28-07-2011	JP 5675848 B2 JP 2013518477 A KR 20120114327 A TW 201142829 A US 2011182436 A1 US 2014205107 A1 WO 2011094232 A1	25-02-2015 20-05-2013 16-10-2012 01-12-2011 28-07-2011 24-07-2014 04-08-2011
US 2016094910 A1	31-03-2016	NONE	
US 2009323982 A1	31-12-2009	JP 5762956 B2 JP 2011527025 A KR 20110038024 A TW 201009817 A US 2009323982 A1 US 2016027451 A1 WO 2010005493 A1	12-08-2015 20-10-2011 13-04-2011 01-03-2010 31-12-2009 28-01-2016 14-01-2010