

(51) International Patent Classification:
H04R 3/00 (2006.01)(21) International Application Number:
PCT/US2015/063519(22) International Filing Date:
2 December 2015 (02.12.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/098,247 30 December 2014 (30.12.2014) US(71) Applicant: **KNOWLES ELECTRONICS, LLC**
[US/US]; 1151 Maplewood Drive, Itasca, IL 60143 (US).(72) Inventors: **VALLABHAN, Harinarayanan Erumbi**;
Knowles Electronics, LLC, 1151 Maplewood Drive, Itasca,
IL 60143 (US). **SAKRI, Shailesh**; Knowles Electronics,
LLC, 1151 Maplewood Drive, Itasca, IL 60143 (US).
AVENDANO, Carlos; Knowles Electronics, LLC, 1151
Maplewood Drive, Itasca, IL 60143 (US). **SOLBACH,**
Ludger; Knowles Electronics, LLC, 1151 Maplewood
Drive, Itasca, IL 60143 (US).(74) Agents: **DRAPINSKI, James W.** et al.; Carr & Ferrell
LLP, 120 Constitution Drive, Menlo Park, California
94025 (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, 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, KN, KP, KR,
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).

Published:

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

(54) Title: DIRECTIONAL AUDIO CAPTURE

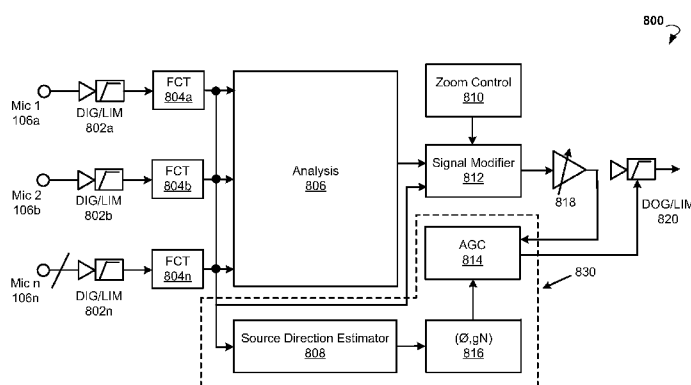


FIG. 8

(57) Abstract: Systems and methods for improving performance of a directional audio capture system are provided. An example method includes correlating phase plots of at least two audio inputs, with the audio inputs being captured by at least two microphones. The method can further include generating, based on the correlation, estimates of salience at different directional angles to localize a direction of a source of sound. The method can allow providing cues to the directional audio capture system based on the estimates. The cues include attenuation levels. A rate of change of the levels of attenuation is controlled by attack and release time constants to avoid sound artifacts. The method also includes determining a mode based on an absence or presence of one or more peaks in the estimates of salience. The method also provides for configuring the directional audio capture system based on the determined mode.

DIRECTIONAL AUDIO CAPTURE

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] The present application is a Continuation-in-Part of U.S. Patent Application No. 12/896,725, filed October 1, 2010, which claims the benefit of U.S. Provisional Application No. 61/266,131, filed December 2, 2009; the present application also claims the benefit of U.S. Provisional Application No. 62/098,247, filed December 30, 2014. The subject matter of the aforementioned applications is incorporated herein by reference for all purposes.

FIELD

[0002] The present disclosure relates generally to audio processing and, more particularly, to systems and methods for improving performance of directional audio capture.

BACKGROUND

[0003] Existing systems for directional audio capture are typically configured to capture an audio signal within an area of interest (e.g., within a lobe) and to suppress anything outside the lobe. Furthermore, the existing systems for directional audio capture do not utilize the directionality of the speaker being recorded. This results in non-uniform suppression throughout the lobe. The robustness of such systems can be compromised, especially in cases of varying distances between a talker (i.e., speaker) and an audio capturing device for a given angle. If the talker moves closer to or farther away from the device, the suppression can become non-uniform.

[0004] In the existing solutions for directional audio capture, off-the box/calibration and customer requirements may disagree. This disagreement may result in more or less

suppression needed at a certain range of angles. With the non-uniform suppression, deploying such solutions can become even more challenging where suppressing/boosting certain angles is desirable to maintain uniform noise suppression across the lobe.

[0005] The existing directional audio capture solutions can also be very sensitive to microphone sealing. Better microphone sealing results in more uniform suppression and poor microphone sealing results in non-uniform suppression. Microphone sealing, in general, can make one device different from another device even when the same batch of manufacturing is used. A solution that makes microphone sealing robust during a change in distance between a talker and an audio capture system is desirable.

SUMMARY

[0006] 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.

[0007] Provided are systems and methods for improving performance of a directional audio capturing system. An example method includes correlating phase plots of at least two audio inputs. The method allows for generating, based on the correlation, estimates of salience at different directional angles to localize at least one direction associated with at least one source of a sound. The method also includes determining cues, based on the estimates of salience, and providing the cues to the directional audio capture system.

[0008] In some embodiments, the cues are used by the directional audio capture system to attenuate or amplify the at least two audio inputs at the different directional angles. In certain embodiments, the cues include at least attenuation levels for the different directional angles. In some embodiments, the estimates of salience include a vector of saliences at directional angles from 0 to 360 in a plane parallel to a ground.

[0009] In some embodiments, generating the cues includes mapping the different directional angles to relative levels of attenuation for the directional audio capture system. In certain embodiments, the method includes controlling the rate of changing of the levels of attenuation in a real time by attack and release time constants to avoid sound artifacts.

[0010] In some embodiments, the method includes determining, based on absence or presence of one or more peaks in the estimates of salience, a mode from a plurality of the operational modes. The method allows configuring, based on the determined

mode, the directional audio capture system. In certain embodiments, the method allows controlling a rate of switching between modes from the plurality of the operational modes in real time by applying attack and release time constants. In some embodiments, the audio inputs are captured by at least two microphones having different qualities of sealing.

[0011] According to another example embodiment of the present disclosure, the steps of the method for improving performance of directional audio capture systems are stored on a machine-readable medium comprising instructions, which when implemented by one or more processors, perform the recited steps.

[0012] 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

[0013] Embodiments are illustrated by way of example and not limitation in the figures of the accompanying drawings, in which like references indicate similar elements.

[0014] FIG. 1 is a block diagram of an exemplary environment in which the present technology can be used.

[0015] FIG. 2 is a block diagram of an exemplary audio device.

[0016] FIG. 3 is a block diagram of an exemplary audio processing system.

[0017] FIG. 4 is a block diagram of an exemplary beam former module.

[0018] FIG. 5 is a flow chart of an exemplary method for performing an audio zoom.

[0019] FIG. 6 is a flow chart of an exemplary method for enhancing acoustic signal components.

[0020] FIG. 7 is a flow chart of an exemplary method for generating a multiplicative mask.

[0021] FIG. 8 is a block diagram of an exemplary audio processing system suitable for improving performance of directional audio capture.

[0022] FIG. 9 is a flow chart of an exemplary method for improving performance of directional audio capture.

[0023] FIG. 10 is a computer system that can be used to implement methods disclosed herein, according to various example embodiments.

DETAILED DESCRIPTION

[0024] The technology disclosed herein relates to systems and methods for improving performance of directional audio capture. Embodiments of the present technology may be practiced with audio devices operable at least to capture and process acoustic signals. The audio devices can include: 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 include 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 include outputs, such as Light-Emitting Diode (LED) indicators, video displays, touchscreens, speakers, and the like. In some embodiments, the audio devices include 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. In certain embodiments, the audio devices include Television (TV) sets, car control and audio systems, smart thermostats, light switches, dimmers, and so on.

[0025] In various embodiments, the audio devices operate in stationary and portable environments. Stationary environments can include residential and commercial buildings or structures, and the like. For example, the stationary embodiments can include living rooms, bedrooms, home theaters, conference rooms, auditoriums, business premises, and the like. Portable environments can include moving vehicles, moving persons, other transportation means, and the like.

[0026] According to an example embodiment, a method for improving a directional audio capture system includes correlating phase plots of at least two audio inputs. The method allows for generating, based on the correlation, estimates of salience at different

directional angles to localize at least one direction associated with at least one source of a sound. The cues include at least levels of attenuation. The method includes determining cues, based on the estimates of salience, and providing the cues to the directional audio capture system.

[0027] FIG. 1 is a block diagram of an exemplary environment 100 in which the present technology can be used. The environment 100 of FIG. 1 includes audio device 104, and audio sources 112, 114 and 116, all within an environment 100 having walls 132 and 134.

[0028] A user of the audio device 104 may choose to focus on or “zoom” into a particular audio source from the multiple audio sources within environment 100. Environment 100 includes audio sources 112, 114, and 116 which all provide audio in multidirections, including towards audio device 104. Additionally, reflections from audio sources 112 and 116 as well as other audio sources may provide audio which reflects off the walls 132 and 134 of the environment 100 and is directed at audio device 104. For example, reflection 128 is a reflection of an audio signal provided by audio source 112 and reflected from wall 132, and reflection 129 is a reflection of an audio signal provided by audio source 116 and reflected from wall 134, both of which travel towards audio device 104.

[0029] The present technology allows the user to select an area to “zoom.” By performing an audio zoom on a particular area, the present technology detects audio signals having a source within the particular area and enhances those signals with respect to signals from audio sources outside the particular area. The area may be defined using a beam, such as, for example, beam 140 in FIG. 1. In FIG. 1, beam 140 contains an area that includes audio source 114. Audio sources 112 and 116 are contained outside the beam area. As such, the present technology would emphasize or “zoom” into the audio signal provided by audio source 114 and de-emphasize the audio

provided by audio sources 112 and 116, including any reflections provided by environment 100, such as reflections 128 and 129.

[0030] A primary microphone 106 and secondary microphone 108 of audio device 104 may be omni-directional microphones. Alternate embodiments may utilize other forms of microphones or acoustic sensors, such as directional microphones.

[0031] While the microphones 106 and 108 receive sound (i.e., acoustic signals) from the audio source 114, the microphones 106 and 108 also pick up noise from audio source 112. Although the noise 122 is shown coming from a single location in FIG. 1, the noise 122 may include any sounds from one or more locations that differ from the location of audio source 114, and may include reverberations and echoes. The noise 124 may be stationary, non-stationary, and/or a combination of both stationary and non-stationary noise.

[0032] Some embodiments may utilize level differences (e.g., energy differences) between the acoustic signals received by the two microphones 106 and 108. Because the primary microphone 106 is much closer to the audio source 116 than the secondary microphone 108 in a close-talk use case, the intensity level for noise 126 is higher for the primary microphone 106, resulting in a larger energy level received by the primary microphone 106 during a speech/voice segment, for example.

[0033] The level difference may then be used to discriminate speech and noise in the time-frequency domain. Further embodiments may use a combination of energy level differences and time delays to discriminate speech. Based on binaural cue encoding, speech signal extraction or speech enhancement may be performed.

[0034] FIG. 2 is a block diagram of an exemplary audio device. In some embodiments, the audio device of FIG. 2 provides more detail for audio device 104 of FIG. 1.

[0035] In the illustrated embodiment, the audio device 104 includes a receiver 210, a processor 220, the primary microphone 106, an optional secondary microphone 108, an

audio processing system 230, and an output device 240. The audio device 104 may include further or other components needed for 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) in 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 operations and other operations for the processor 220.

[0037] The exemplary receiver 210 is an acoustic 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 to reduce noise using the techniques described herein, and provide an audio signal to the output device 240. 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 is configured to receive the acoustic signals from an acoustic source via the primary microphone 106 and secondary microphone 108 and process the acoustic signals. Processing may include performing noise reduction within an acoustic signal. The audio processing system 230 is discussed in more detail below. The primary and secondary microphones 106, 108 may be spaced a distance apart in order to allow for detecting an energy level difference, time difference, or phase difference between them. The acoustic signals received by primary microphone 106 and secondary microphone 108 may be converted into electrical signals (i.e., a primary electrical signal and a secondary electrical signal). The electrical signals may themselves be converted by an analog-to-digital converter (not shown) into digital signals for processing, in accordance with some embodiments. In order to differentiate the acoustic signals for clarity purposes, the acoustic signal received by the primary

microphone 106 is herein referred to as the primary acoustic signal, while the acoustic signal received from by the secondary microphone 108 is herein referred to as the secondary acoustic signal. The primary acoustic signal and the secondary acoustic signal may be processed by the audio processing system 230 to produce a signal with an improved signal-to-noise ratio. It should be noted that embodiments of the technology described herein may be practiced utilizing only the primary microphone 106.

[0039] The output device 240 is any device that provides an audio output to the user. For example, the output device 240 may include a speaker, an earpiece of a headset or handset, or a speaker on a conference device.

[0040] In various embodiments, where the primary and secondary microphones 106 and 108 are omni-directional microphones that are closely-spaced (e.g., 1-2 cm apart), a beamforming technique may be used to simulate forwards-facing and backwards-facing directional microphones. The level difference may be used to discriminate speech and noise in the time-frequency domain, which can be used in noise reduction.

[0041] FIG. 3 is a block diagram of an exemplary audio processing system. The block diagram of FIG. 3 provides more detail for the audio processing system 230 in the block diagram of FIG. 2. Audio processing system 230 includes fast cosine transform (FCT) modules 302 and 304, beam former module 310, multiplicative gain expansion module 320, reverb module 330, mixer module 340, and zoom control module 350.

[0042] FCT modules 302 and 304 may receive acoustic signals from audio device microphones and convert the acoustic signals to frequency range sub-band signals. In some embodiments, FCT modules 302 and 304 are implemented as one or more modules that create one or more sub-band signals for each received microphone signal. FCT modules 302 and 304 receive an acoustic signal from each microphone contained in audio device 104. These received signals are illustrated as signals X_1 - X_L , wherein X_1 is a primary microphone signal and X_L represents the remaining microphone signals. In

some embodiments, the audio processing system 230 of FIG. 3 performs an audio zoom on a per frame and per sub band basis.

[0043] In some embodiments, beam former module 310 receives the frequency sub-band signals as well as a zoom indication signal. The zoom indication is received from zoom control module 350. The zoom indication communicated by zoom indicator signal K may 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, beam former module 310 receives sub-band signals, processes the sub-band signals to identify which signals are within a particular area to enhance (or “zoom”), and provides 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. Beam former module 310 also provides a gain factor to multiplicative gain expansion module 320. The gain factor may indicate whether multiplicative gain expansion module 320 should perform additional gain or reduction to the signals received from beam former module 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 beam former module 310 may be a ratio of how much energy is reduced in the signal from the primary microphone versus the energy in the signals from the other microphones. Hence, the gain may be a boost or cancellation gain expansion factor. The gain factor is discussed in more detail below.

[0044] Beam former module 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 beam forming, the beam is focused by narrowing constraints of alpha and gamma. For a rider beam, the constraints may be made larger. Hence, a beam may be manipulated by putting a protective range around the preferred direction. Beam former module 310 may be implemented by a system described in the U.S. Patent Application No.

61/325,764, entitled "Multi-Microphone Robust Noise Suppression System," the disclosure of which is incorporated herein by reference. Additional techniques for reducing undesired audio components of a signal are discussed in U.S. Patent Application No. 12/693,998 (now U.S. Patent No. 8,718,290), entitled "Adaptive Noise Reduction Using Level Cues," the disclosure of which is incorporated herein by reference.

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

[0046] The gain factor may be implemented as one of several different energy ratios. For example, the energy ratio may be the 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 the intra level difference between a primary signal and another signal. The gain factors may be an indication of signal strength in a target direction versus all other directions. Put another way, the gain factor may be an indication of multiplicative expansions due, and whether additional expansion or subtraction should be performed at the multiplicative gain expansion module 320. Multiplicative gain expansion module 320 outputs the modified signal and provides signal to reverb module 330 (which may also function to de-reverb).

[0047] Reverb module 330 receives the sub-band signals output by multiplicative gain expansion module 320, as well as the microphone signals which were also received by beam former module 310, and performs reverberation or dereverberation to the sub-

band signals output by multiplicative gain expansion module 320. Reverb module 330 may adjust a ratio of direct energy to remaining energy within a signal based on the zoom control indicator provided by zoom control module 350.

[0048] Adjusting the reverb for a signal may involve adjusting the energy of different components of the signal. An audio signal has several components in a frequency domain, including a direct component, early reflections, and a tail component. A direct component typically has the highest energy level, followed by a somewhat lower energy level of reflections within the signal. Also included within a very particular signal is a tail, which may include noise and other low energy data or low energy audio. A reverberation is defined as reflections of the direct audio component. Hence, a reverberation with many reflections over a broad frequency range results in a more noticeable reverberation. A signal with fewer reflection components has a smaller reverberation component.

[0049] Typically, the further away a listener is from an audio source, the larger the reverberation in the signal, and the closer a listener is to an audio source, the smaller the intensity of the reverberation signal (reflection components). Hence, based on the zoom indicator received from zoom control module 350, reverb module 330 may adjust the reverberation components in the signal received from multiplicative gain expansion module 320. Hence, if the zoom indicator received indicates that a zoom in operation is to be performed on the audio, the reverberation will be decreased by minimizing the reflection components of the received signal. If the zoom indicator indicates that a zoom out is to be performed on the audio signal, the early reflection components are gained to increase these components to make it appear as if there is additional reverberation within the signal. After adjusting the reverberation of the received signal, reverb module 330 provides the modified signal to mixer module 340.

[0050] The mixer module 340 receives the reverberation adjusted signal and mixes the signal with the signal from the primary microphone. In some embodiments, mixer

module 340 increases the energy of the signal appropriately when there is audio present in the frame and decreases it where there is little audio energy present in the frame.

[0051] FIG. 4 is a block diagram of an exemplary beam former module. The beam former module 310 may be implemented per tap (i.e., per sub-band). Beam former module 310 receives FCT output signals for a first microphone (such as a primary microphone) and a second microphone. The first microphone FCT signal is processed by module 410 according to the function:

$$b_{\omega}e^{j\omega\tau_f}$$

to generate a first differential array with parameters.

[0052] The secondary microphone FCT signal is processed by module 420 according to the function:

$$a_{\omega}e^{j\omega\tau_b}$$

to generate a second differential array with parameters. Further details regarding the generation of exemplary first and second differential arrays are described in the U.S. Patent Application No. 11/699,732,764, entitled "System and Method for Utilizing Omni-Directional Microphones for Speech Enhancement," now U.S. Patent No. 8,194,880, issued June 5, 2012, the disclosure of which is incorporated herein by reference.

[0053] The output of module 410 is then subtracted from the secondary microphone FCT signal at combiner 440 and the output of module 420 is then subtracted by the primary microphone FCT signal at combiner 430. A cardioid signal C_f is output from combiner 430 and provided to module 450 where the following function is applied:

$$\text{Log}(|C_f|^2).$$

[0054] A cardioid signal C_b is output from combiner 440 and provided to module 460 where the following function is applied:

$$\text{Log}(|C_b|^2).$$

[0055] The difference of the outputs of modules 450 and 460 is determined by element 470 and output as an ILD cue. The ILD cue may be output by beam former module 310 to a post filter (for example, a filter implemented by multiplicative gain expansion module 320).

[0056] FIG. 5 is a flow chart of an exemplary method for performing an audio zoom. An acoustic signal is received from one or more sources at step 510. In some embodiments, the acoustic signals are received through one or more microphones on audio device 104. For example, acoustic signals from audio sources 112-116 and reflections 128-129 are received through microphones 106 and 108 of audio device 104.

[0057] A zoom indication is then received for a spatial area at step 520. In some embodiments, the zoom indication is received from a user or determined based on other data. For example, the zoom indication is received from a user via a video zoom setting, pointing an audio device in a particular direction, an input for video zoom, or in some other manner.

[0058] Acoustic signal component energy levels are enhanced based on the zoom indication at step 530. In some embodiments, acoustic signal component energy levels are enhanced by increasing the energy levels for audio source sub-band signals that originate from a source device within a selected beam area. Audio signals from a device outside a selected beam area are de-emphasized. Enhancing acoustic signal component energy levels is discussed in more detail below with respect to the method of FIG. 6.

[0059] Reverberation signal components associated with a position inside the spatial area are adjusted based on the received indication at step 540. As discussed above, the adjustments may include modifying the ratio of a direct component with respect to reflection components for the particular signal. When a zoom in function is to be performed, reverberation should be decreased by increasing the ratio of the direct component to the reflection components in the audio signal. When a zoom out function

is performed for the audio signal, the direct component is reduced with respect to the reflection components to decrease the ratio of direct to reflection components of the audio signal.

[0060] A modulated gain is applied to the signal component at step 550. The gain may be applied by mixing a reverb processed acoustic signal with a primary acoustic signal (or another audio signal received by audio device 104). The mixed signal that has been processed by audio zoom is output at step 560.

[0061] As discussed above, sub-band signals are enhanced based on a zoom indication. FIG. 6 is a flow chart of an exemplary method for enhancing acoustic signal components. In some embodiments, the method in FIG. 6 provides more detail for step 530 of the method in FIG. 5. An audio source is detected in the direction of a beam at step 610. This detection may be performed by a null-processing noise subtraction mechanism or some other module that is able to identify a spatial position of a source based on audio signals received by two or more microphones.

[0062] Acoustic signal sources located outside the spatial area are attenuated at step 620. In various embodiments, the acoustic sources outside the spatial area include certain audio sources (e.g., 112 in FIG. 1) and reflected audio signals such as reflections 128 and 129. Adaptation constraints are then used to steer the beam based on the zoom indication at step 630. In some embodiments, the adaptation constraints include α and σ constraints used in a null processing noise suppression system. The adaptation constraints may also be derived from multiplicative expansion or selection of a region around a preferred direction based on a beam pattern.

[0063] Energy ratios are then determined at step 640. The energy ratios may be used to derive multiplicative masks that boost or reduce a beam former cancellation gain for signal components. Next, multiplicative masks are generated based on energy ratios at step 650. Generating multiplicative masks based on an energy ratio is discussed in more detail below with respect to the method of FIG. 7.

[0064] FIG. 7 is a flow chart of an exemplary method for generating a multiplicative mask. The method of FIG. 7 provides more detail for step 650 in the method of FIG. 6. Differential arrays are generated from microphone signals at step 710. The arrays may be generated as part of a beam former module 310. The beam pattern may be a cardioid pattern generated based at least in part from the differential output signals. Next, a beam pattern is generated from the differential arrays at step 720. Energy ratios are then generated from beam patterns at step 730. The energy ratios may be generated as any of a combination of signals. Once generated, an ILD map may be generated per frequency from energy ratios. An ILD range corresponding to the desired selection may be selected. An ILD window may then be applied to a map by boosting the signal components within the window and attenuating the signal components positioned outside the window. A filter, such as a post filter, may be derived from the energy ratio at step 740.

[0065] The above described modules, including those discussed with respect to FIG. 3, may include instructions stored in a storage media such as a machine readable medium (e.g., computer readable medium). These instructions may be retrieved and executed by the processor 220 to perform the functionality discussed herein. Some examples of instructions include software, program code, and firmware. Some examples of storage media include memory devices and integrated circuits.

[0066] FIG. 8 is a block diagram illustrating an audio processing system 800, according to another example embodiment. The example audio processing system 800 includes a source estimation subsystem 830 coupled to various elements of an example AZA subsystem. The example AZA subsystem includes limiters 802a, 802b, ..., and 802n, FCT modules 804a, 804b, ..., and 804n, analysis module 806, zoom control module 810, signal modifier 812, element 818, and a limiter 820. The source estimation subsystem 830 may include a source direction estimator (SDE) module 808, also referred to as a target estimator, a gain module 816, and an automatic gain control

(AGC) module 814. The example audio processing system 800 processes acoustic audio signal from microphones 106a, 106b, ..., and 106n.

[0067] In various exemplary embodiments, SDE module 808 is operable to localize a source of sound. The SDE module 808 may generate cues based on correlation of phase plots between different microphone inputs. Based on the correlation of the phase plots, the example SDE module 808 can compute a vector of salience estimates at different angles. Based on the salience estimates, the SDE module 808 can determine a direction of the source. In other words, according to various embodiments, a peak in the vector of salience estimates is an indication of direction of source in a particular direction. At the same time, sources of diffused nature, i.e., non-directional, may be represented by poor salience estimates at all the angles. Various embodiments may rely upon the cues (estimates of salience) to improve the performance of an existing directional audio solution, which is carried out by the analysis module 806, signal modifier 812, and zoom control module 810.

[0068] According to an example embodiment, 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/user.

[0069] In various embodiments, the SDE module 808 is configured to operate in two and more modes. The modes of operation can include "normal," "noisy," and "simultaneous talkers." The characterization of these modes is driven by a SDE salience parameter.

Normal mode

[0070] A "normal" mode of operation is defined by a single directional speech source without the presence of any kind of strong speech distractors with or without the presence of noise. A vector of salience estimates in such case can be characterized by a

single peak (above a salience threshold). The single peak can indicate a presence of a single source of sound. The location of the peak, in the vector of salience estimates, may characterize the angle of the source. In such cases, both a diffused source detector and a simultaneous talker detector may be set to a "no" state. Based on these states, the target estimator, in various embodiments, drives the level of suppression/amplification as desired by the user on a per angle basis.

[0071] In some embodiments, the target estimator generates a mapping of angle to relative levels of attenuation in the AZA subsystem. For example, a range of angles 240-270 degrees may require 10dB of incremental suppression relative to AZA's performance target estimator containing an array with 0dB throughout except for the entries between 240 and 270 degrees.

[0072] Although an immediate relative suppression level of 10dB is achievable on detection, in a real time speech system, such suppression may cause audible distortion to a listener due to sudden jumps in signal levels. In some embodiments, to alleviate the distortion problem, the AGC module 814 can control the rate of roll-off by means of attack and release time constants. A smooth roll-off can effectively stabilize the speech system without audible distortions in the audio. In some embodiments, noise, if present along with the directional speech, is alleviated by the AZA subsystem.

Noisy mode

[0073] A noisy mode of operation can be characterized by a diffused noise source with no directional speech. The noisy mode can result in poor salience estimates for all angles. Since there is no directional information of the source of such data, the signal can be processed solely by the AZA subsystem. In some embodiments, interactions between the noisy mode and the normal mode of operations are handled smoothly without sudden switch overs to avoid pumping or any gain related artifacts. For a smooth handover, a target estimator can provide a target of 0dB to the AGC module

814. By appropriately handling the attack and release time, a smooth handover can be achieved. It should be noted, however, that the attack and release time in the noisy mode are different from the attack and release time used in the normal mode.

Simultaneous talkers mode

[0074] A simultaneous talkers mode is characterized by simultaneous multiple talkers/side distractors either with or without noise. The salience vector for the simultaneous talkers mode can be characterized by multiple peaks (above a salience threshold). The simultaneous talkers mode can be handled in a way similar to the noisy mode. When SDE module operates in the simultaneous talkers mode, acoustic signals from the microphones can be processed solely by the AZA subsystem. In various embodiments, a handover between the above modes can be carried out in a gracious manner with the help of the AGC subsystem.

[0075] Various embodiments of the technology described herein, having the AZA subsystem augmented with a source estimation subsystem, can avoid the problem of microphone sealing by ignoring any inter-microphone signal level differences. Various embodiments focus instead on the time of arrival/phase cues between the microphones. However, it should be noted that even though various embodiments can be insensitive to the microphone sealing, the underlying AZA sub-system may still be sensitive to the microphone sealing, and therefore the overall system performance may depend on the microphone sealing. In some embodiments, to alleviate the microphone sealing problem, an AZA sub-system may be tuned based on characteristics of the sealing of the microphones utilized to reduce the sensitivity on the microphone sealing. Further details regarding exemplary tuning of the AZA sub-system may be found in U.S. Patent Application No. 12/896,725, filed October 1, 2010, incorporated by reference herein.

[0076] Various embodiments of the present technology may utilize the fact that SDE salience varies very little with the change of a distance between a talker/speaker and an

audio device, when the distance is in the range of 0.5m-2m and the speaker's mouth is at around 30cm above the audio device. This can make the audio processing system 800 more robust to the distance variance and can result in an even/similar performance for a talker talking at these distances. In some embodiments, the AZA subsystem may be tuned to take full-advantage of the robustness to the distance.

[0077] The target estimator block (also referred to as SDE module) 808 can provide relative levels of suppression based on the angle of arrival of sounds independently of the AZA subsystem. In some embodiments, the target estimator block can be controlled independently without any interactions with other subsystems. This independently controllable (e.g. "island") architecture can empower the field tuning engineers to match the performance desired by a customer/user.

[0078] As described regarding various embodiments, the array of the target estimators during the "normal" mode of operation provides a powerful tool which can allow implementing the above architecture by manipulating the angle of the suppression level array in the target-estimator block.

[0079] FIG. 9 is a flow chart showing steps of a method 900 of improving performance of directional audio capture system, according to an example embodiment. In block 910, the example method 900 includes correlating phase plots of at least two audio inputs. In some embodiments, the audio inputs can be captured by at least two microphones having different sealing.

[0080] In block 920, the example method 900 allows generating, based on the correlation, estimates of salience at different directional angles to localize at least one direction associated with at least one source of a sound. In some embodiments, the estimates of salience include a vector of saliences at directional angles from 0 to 360 in a plane parallel to a ground.

[0081] In block 930, the example method 900 includes determining cues based on the estimates of salience. In block 940, the example method 900 includes providing those “estimates of salience”-based cues to a directional audio capture system.

[0082] In further embodiments, the example method 900 includes determining, based on the estimates of salience (e.g., absence or presence of one or more peaks in the estimates of salience), a mode from a plurality of the operational modes. In certain embodiments, the operational modes include a “normal” mode characterized by a single directional speech source, a “simultaneous talkers” mode characterized by the presence of at least two single directional speech sources, and a noisy mode characterized by a diffused noise source with no directional speech.

[0083] In block 960, the example method 900 includes configuring, based on the determined mode, the directional audio capture system.

[0084] In block 970, the example method 900 includes determining, based on the estimates of salience and the determined mode, other cues including at least levels of attenuation.

[0085] In block 980, the example method 900 includes controlling a rate of switching between modes from the plurality of the operational modes in real time by applying attack and release time constants.

[0086] FIG. 10 illustrates an exemplary computer system 1000 that may be used to implement some embodiments of the present disclosure. The computer system 1000 of FIG. 10 may be implemented in the contexts of the likes of computing systems, networks, servers, or combinations thereof. The computer system 1000 of FIG. 10 includes one or more processor units 1010 and main memory 1020. Main memory 1020 stores, in part, instructions and data for execution by processor units 1010. Main memory 1020 stores the executable code when in operation, in this example. The computer system 1000 of FIG. 10 further includes a mass data storage 1030, portable

storage device 1040, output devices 1050, user input devices 1060, a graphics display system 1070, and peripheral devices 1080.

[0087] The components shown in FIG. 10 are depicted as being connected via a single bus 1090. The components may be connected through one or more data transport means. Processor unit 1010 and main memory 1020 are connected via a local microprocessor bus, and the mass data storage 1030, peripheral device(s) 1080, portable storage device 1040, and graphics display system 1070 are connected via one or more input/output (I/O) buses.

[0088] Mass data storage 1030, 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 1010. Mass data storage 1030 stores the system software for implementing embodiments of the present disclosure for purposes of loading that software into main memory 1020.

[0089] Portable storage device 1040 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 1000 of FIG. 10. The system software for implementing embodiments of the present disclosure is stored on such a portable medium and input to the computer system 1000 via the portable storage device 1040.

[0090] User input devices 1060 can provide a portion of a user interface. User input devices 1060 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 1060 can also include a touchscreen. Additionally, the computer system 1000 as shown in FIG. 10 includes output devices 1050. Suitable output devices 1050 include speakers, printers, network interfaces, and monitors.

[0091] Graphics display system 1070 include a liquid crystal display (LCD) or other suitable display device. Graphics display system 1070 is configurable to receive textual and graphical information and process the information for output to the display device.

[0092] Peripheral devices 1080 may include any type of computer support device to add additional functionality to the computer system.

[0093] The components provided in the computer system 1000 of FIG. 10 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 1000 of FIG. 10 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.

[0094] The processing for various embodiments may be implemented in software that is cloud-based. In some embodiments, the computer system 1000 is implemented as a cloud-based computing environment, such as a virtual machine operating within a computing cloud. In other embodiments, the computer system 1000 may itself include a cloud-based computing environment, where the functionalities of the computer system 1000 are executed in a distributed fashion. Thus, the computer system 1000, when configured as a computing cloud, may include pluralities of computing devices in various forms, as will be described in greater detail below.

[0095] 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.

[0096] 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 1000, 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.

[0097] 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.

CLAIMS

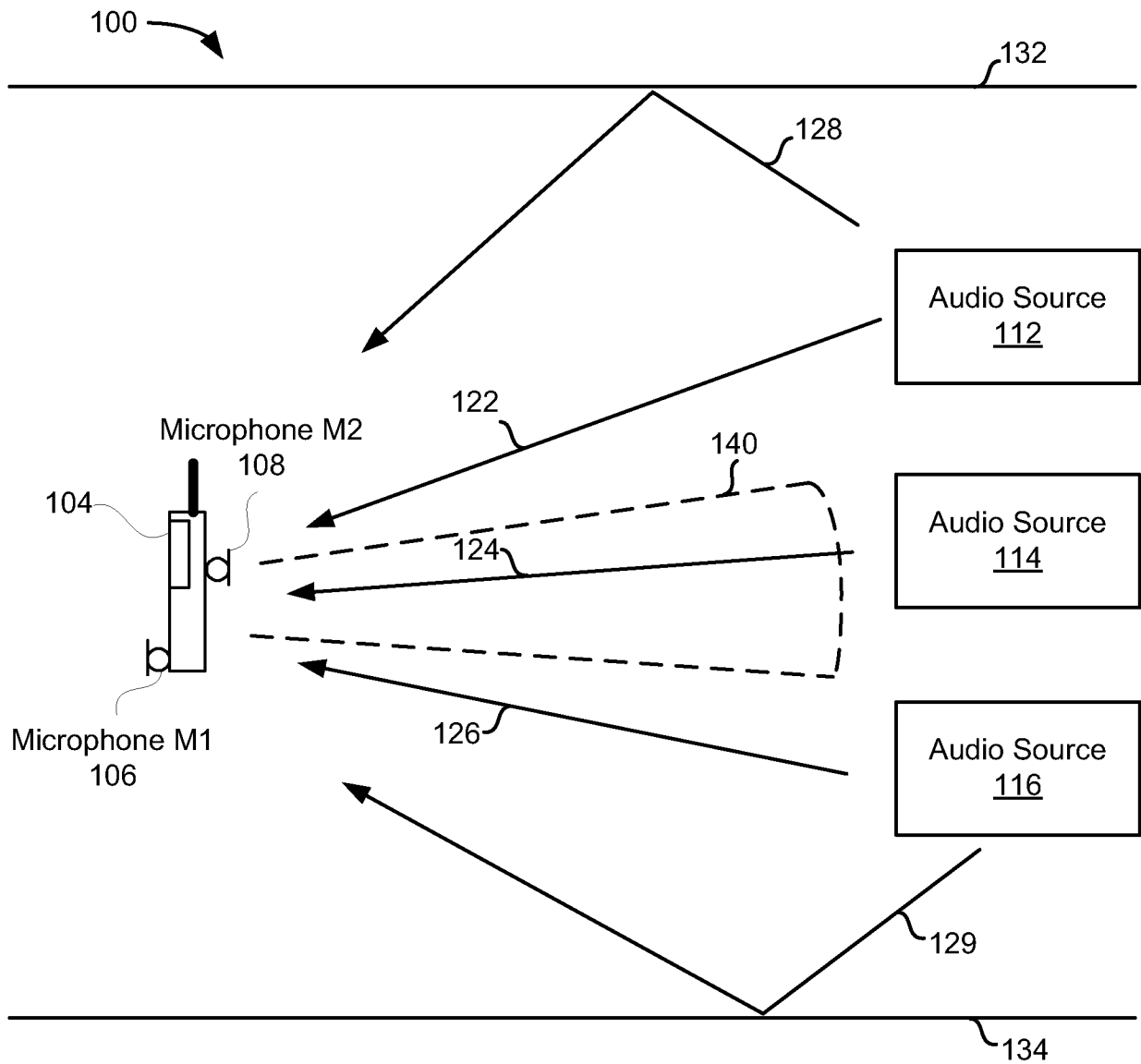
What is claimed is:

1. A method for improving performance of a directional audio capture system, the method comprising:
 - correlating phase plots of at least two audio inputs;
 - generating, based on the correlation, estimates of salience at different directional angles to localize at least one direction associated with at least one source of a sound;
 - determining cues based on the estimates of salience; and
 - providing the cues to the directional audio capture system.
2. The method of claim 1, wherein the cues are used by the directional audio capture system to attenuate or amplify the at least two audio inputs at the different directional angles.
3. The method of claim 1, wherein the cues include at least attenuation levels of the different directional angles.
4. The method of claim 1, wherein the estimates of salience include a vector of saliences at directional angles from 0 to 360 in a plane parallel to a ground.
5. The method of claim 1, wherein generating the cues includes mapping the different directional angles to relative levels of attenuation for the directional audio capture system.

6. The method of claim 5, further comprising controlling the rate of changing the levels of attenuation in a real time by attack and release time constants to avoid sound artifacts.
7. The method of claim 1, further comprising determining, based on the estimates of salience, a mode selected from a plurality of operational modes, the plurality of operational modes including a first operational mode wherein the at least one source of sound includes a single directional speech source.
8. The method of claim 7, wherein the plurality of operational modes further includes a second operational mode wherein the at least one source of sound includes at least two single directional speech sources, and a third operational mode wherein the at least one source of sound includes a diffused noise source having no directional speech.
9. The method of claim 7, wherein determining the mode is based on absence or presence of one or more peaks in the estimates of salience.
10. The method of claim 9, further comprising configuring, based on the determined mode, the directional audio capture system.
11. The method of claim 7, further comprising controlling a rate of switching between modes from the plurality of the operational modes in a real time by applying attack and release time constants.
12. The method of claim 1, wherein the at least two audio inputs are captured by at least two microphones.

13. The method of claim 12, wherein one of the at least two microphones is sealed better than other ones of the at least two microphones.
14. A system for improving performance of a directional audio capture system, 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 performs a method comprising:
 - correlating phase plots of at least two audio inputs;
 - generating, based on the correlation, estimates of salience at different directional angles to localize at least one direction associated with at least one source of a sound;
 - determining cues based on the estimates of salience; and
 - providing the cues to the directional audio capture system.
15. The system of claim 14, wherein the cues are used by the directional audio capture system to attenuate or amplify the at least two audio inputs at the different directional angles.
16. The system of claim 14, wherein the cues include at least attenuation levels for the different directional angles.
17. The system of claim 14, wherein generating the cues includes mapping the different directional angles to relative levels of attenuation for the directional audio capture system.

18. The system of claim 14, further comprising determining, based on absence or presence of one or more peaks in the estimates of salience, a mode selected from a plurality of operational modes, the plurality of operational modes including a first operational mode wherein the at least one source of sound includes a single directional speech source, a second operational mode wherein the at least one source of sound includes at least two single directional speech sources, and a third operational mode wherein the at least one source of sound includes a diffused noise source having no directional speech.
19. The system of claim 18, further comprising:
 configuring, based on the determined mode, the directional audio capture system, and
 controlling a rate of switching between modes in a real time by applying attack and release time constants.
20. A non-transitory computer-readable storage medium having embodied thereon instructions, which when executed by at least one processor, perform steps of a method, the method comprising:
 correlating phase plots of at least two audio inputs;
 generating, based on the correlation, estimates of salience at different directional angles to localize at least one direction associated with at least one source of a sound;
 determining cues based on the estimates of salience; and
 providing the cues to the directional audio capture system.

**FIG. 1**

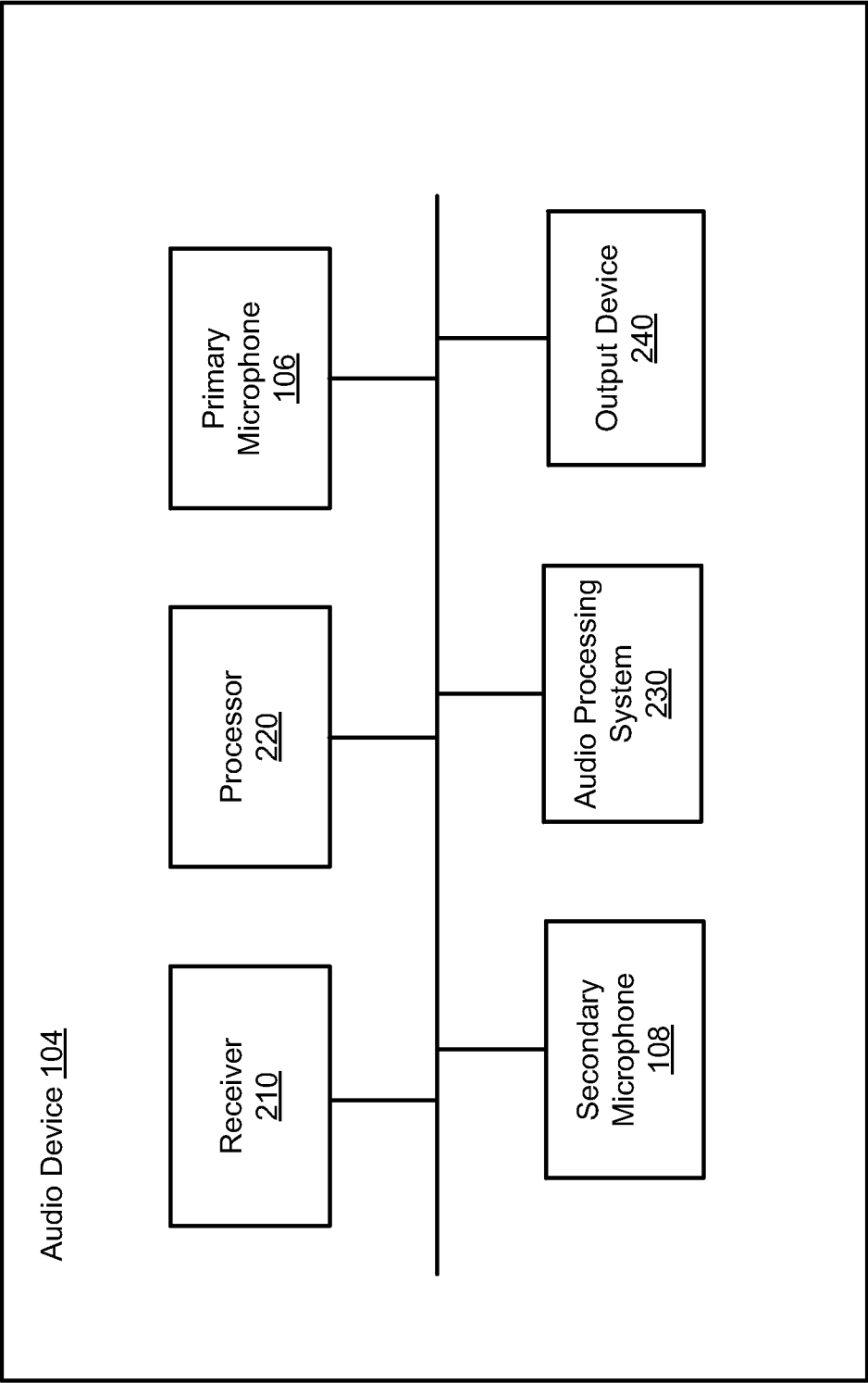


FIG. 2

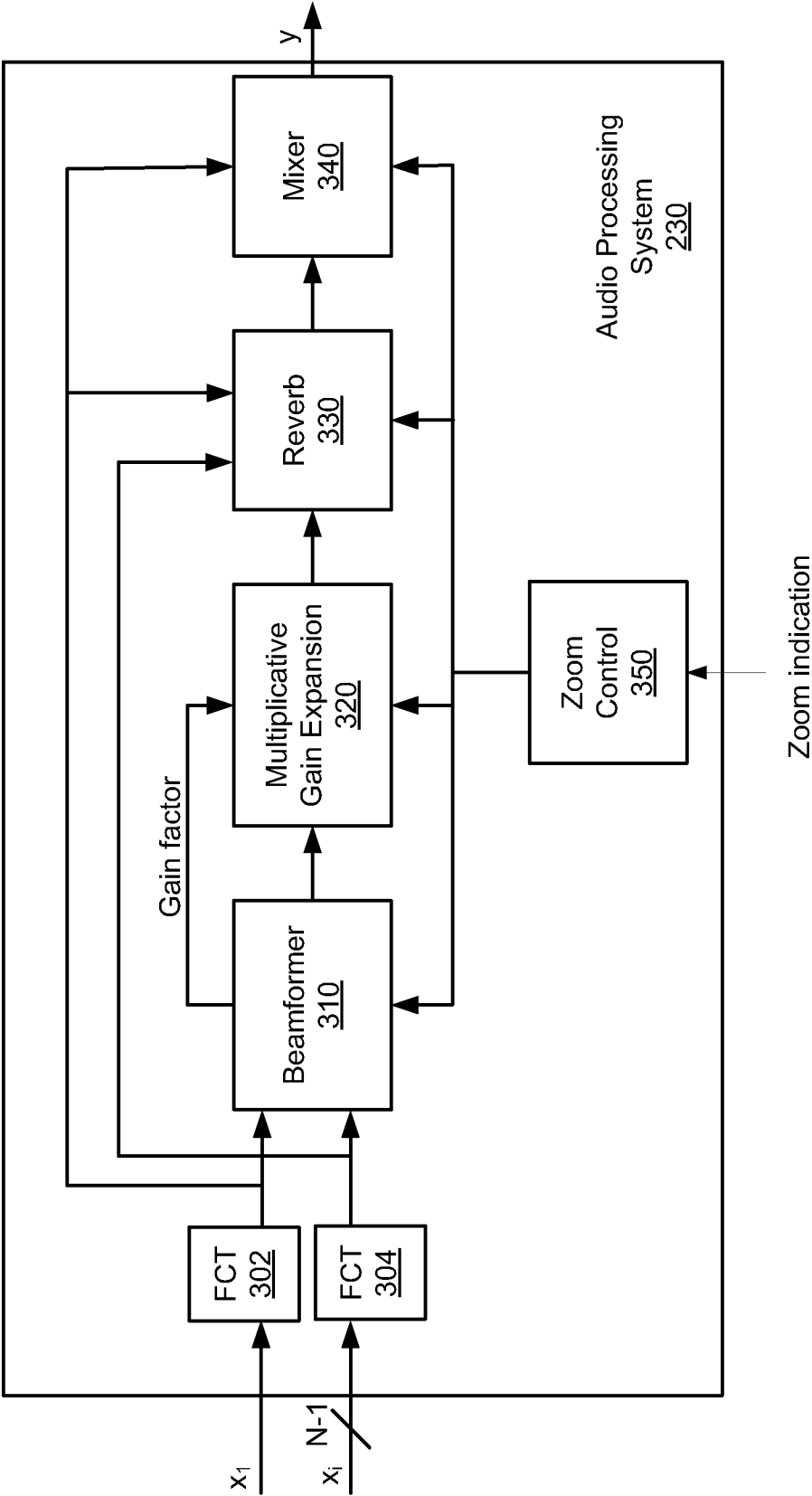
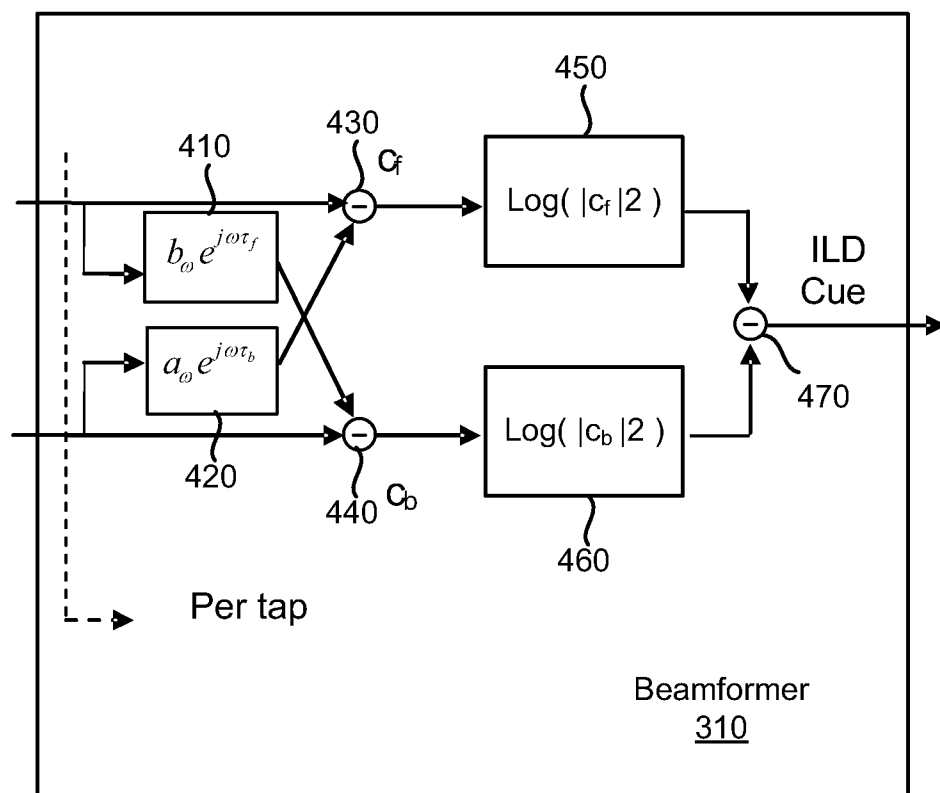
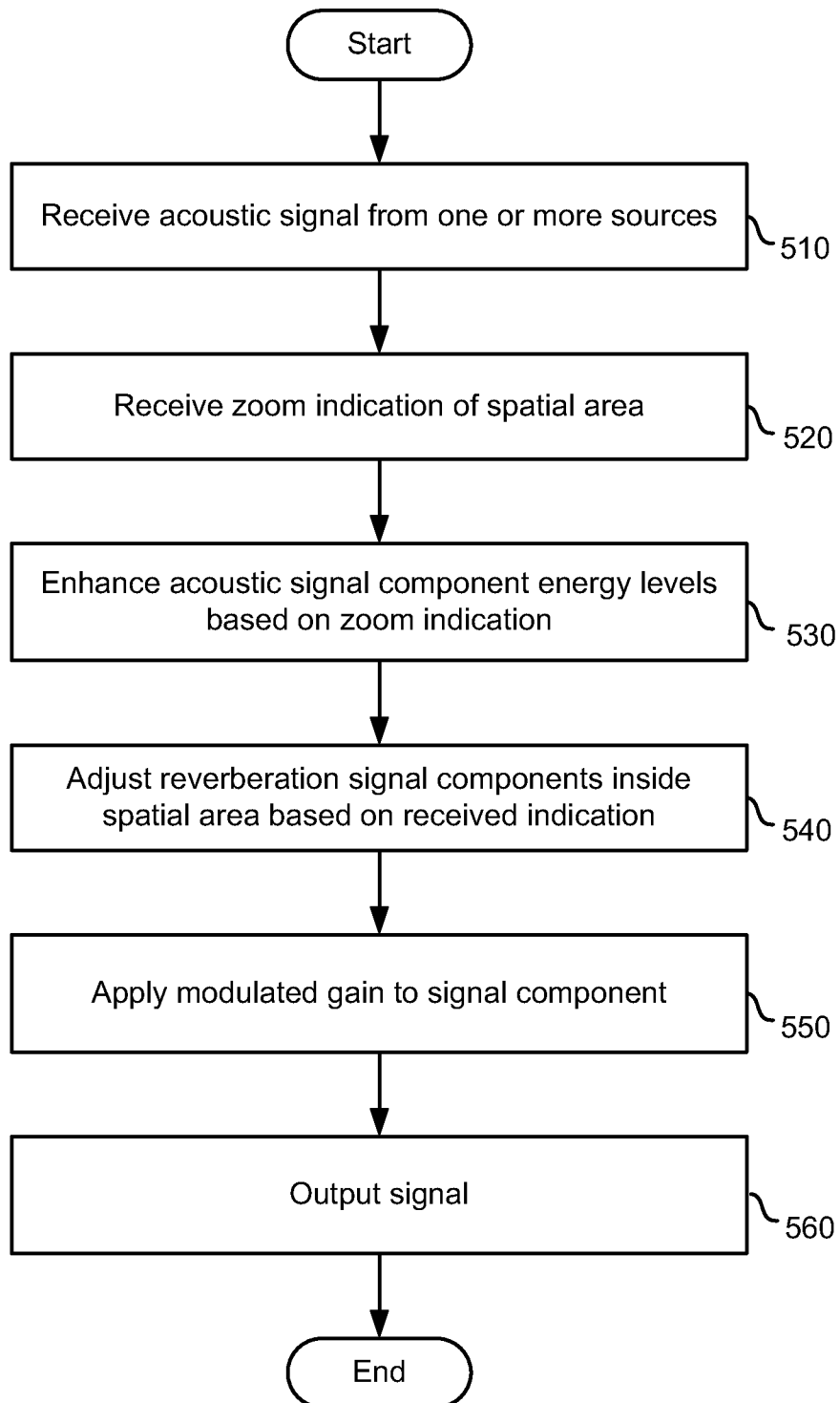
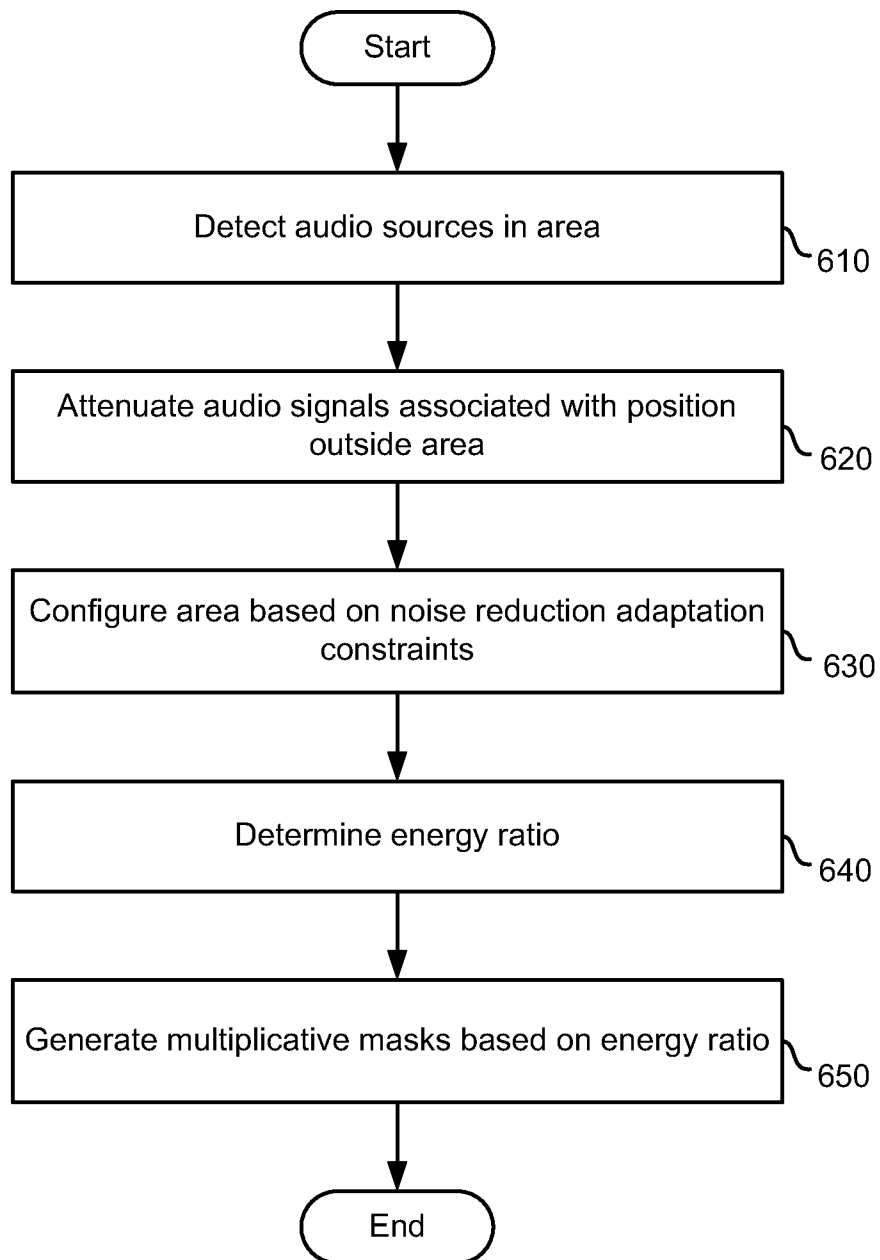
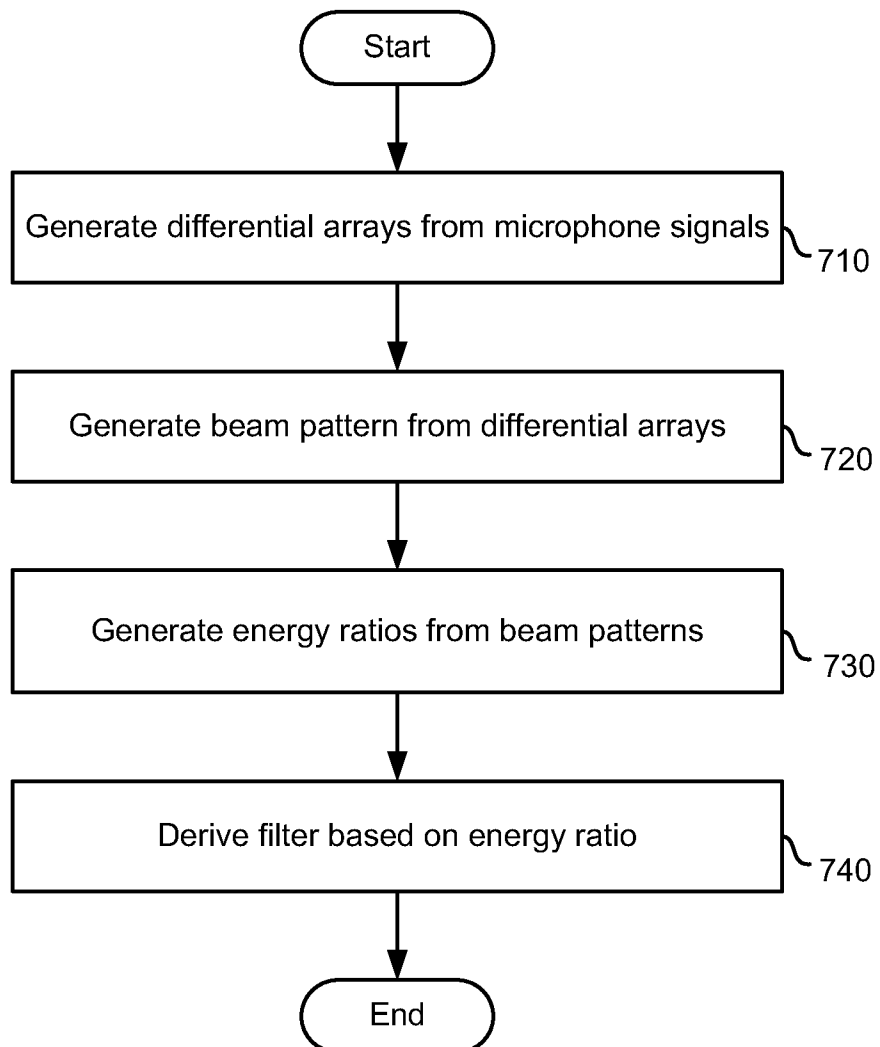


FIG. 3

**FIG. 4**

**FIG. 5**

**FIG. 6**

**FIG. 7**

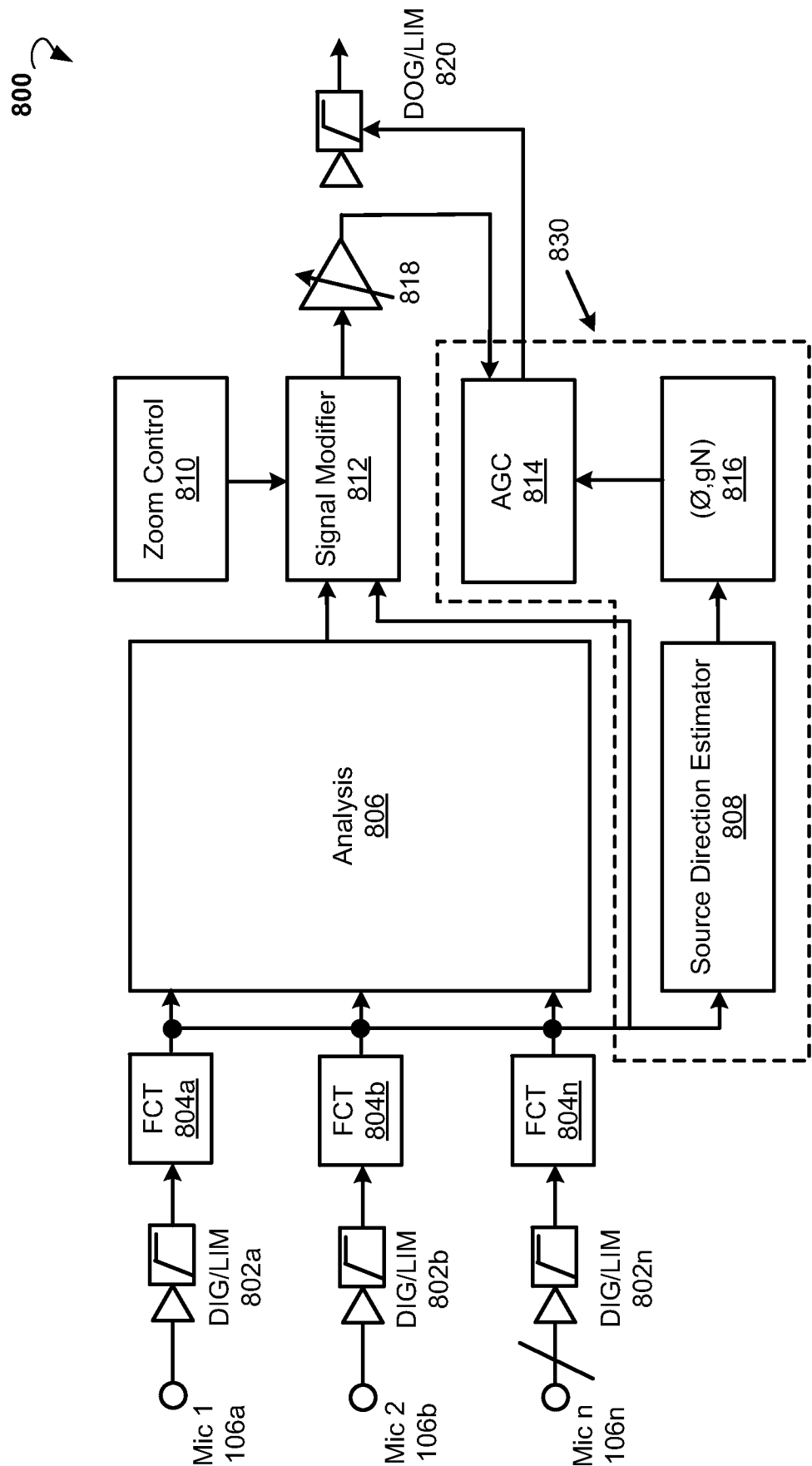
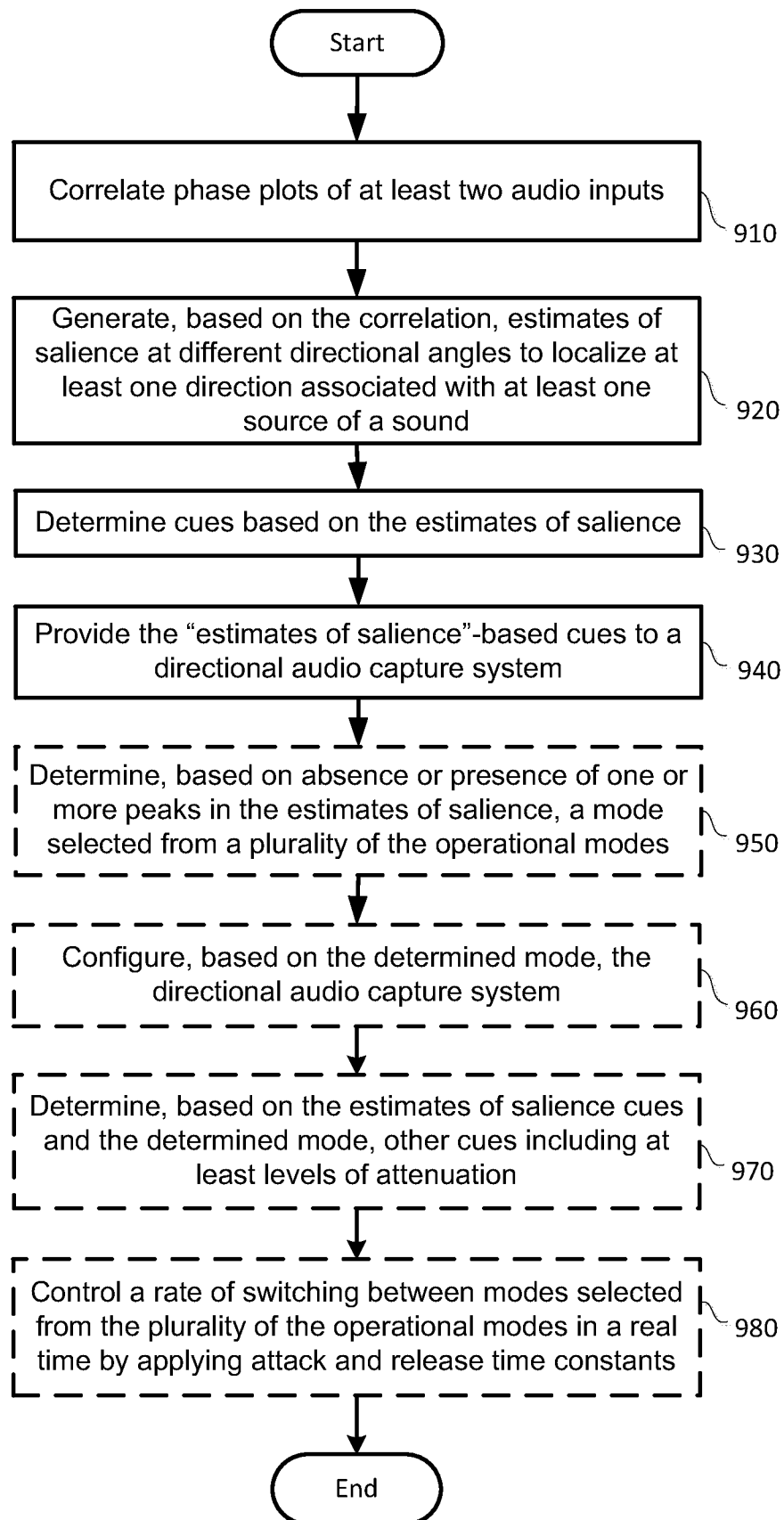
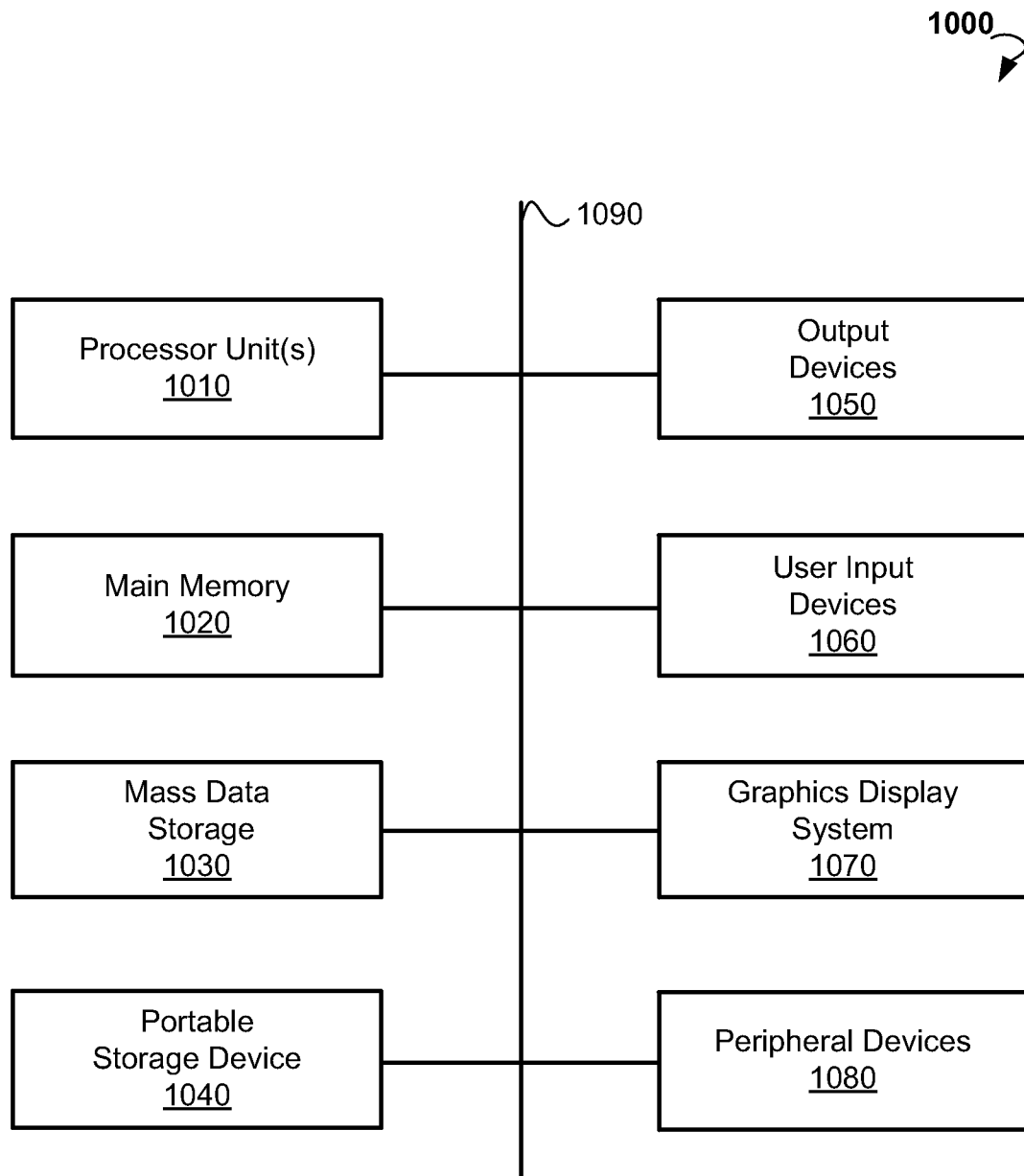


FIG. 8

**FIG. 9**

**FIG. 10**

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2015/063519

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - H04R 3/00 (2015.01)

CPC - H044 3/005 (2015.12)

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC(8) - H04R 3/00 (2015.01)

CPC - H044 1/406, 3/005, 25/407, 2201/401, 2201/403 (2015.12)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC: 381/91, 92, 94.1, 94.2, 94.3, 94.7, 94.8, 95, 312, 313; 704/226, 227, 233, 275 (Keyword delimited)

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

Orbit, Google Patents, Google Scholar, Google

Search terms used: audio, sound, noise, music, directional, angle, orientation, configuration, two, plural, second, three, input, source, salience, zoom, pick, chose, prominent, emphasize, attenuate, amplify, estimate, determine, calculate, processor, controller, memory, storage

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2011/0129095 A1 (AVENDANO et al.) 02 June 2011 (02.06.2011) entire document	1-6, 12, 14-17, 20
Y		7-11, 13, 18, 19
Y	US 2011/0075857 A1 (AOYAGI) 31 March 2011 (31.03.2011) entire document	7-11, 18, 19
Y	US 2012/0257778 A1 (HALL et al.) 11 October 2012 (11.10.2012) entire document	13
A	US 8,194,880 B2 (AVENDANO) 05 June 2012 (05.06.2012) entire document	1-20
A	US 8,233,352 B2 (BEAUCOUP) 31 July 2012 (31.07.2012) entire document	1-20
A	US 2014/0003622 A1 (IKIZYAN et al.) 02 January 2014 (02.01.2014) entire document	1-20

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

* Special categories of cited documents:

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

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

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of 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

26 January 2016

Date of mailing of the international search report

11 FEB 2016

Name and mailing address of the ISA/

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, VA 22313-1450

Facsimile No. 571-273-8300

Authorized officer

Blaine R. Copenheaver

PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774