



- (51) International Patent Classification:
G01S 3/00 (2006.01)
- (21) International Application Number:
PCT/US2015/066012
- (22) International Filing Date:
16 December 2015 (16.12.2015)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
62/093,903 18 December 2014 (18.12.2014) US
- (71) Applicant: ANALOG DEVICES, INC. [US/US]; One Technology Way, Norwood, MA 02062 (US).
- (72) Inventors: TRAA, Johannes; 1977 Orchard Street, Apt. C, Urbana, IL 61807 (US). STEIN, Noah, Daniel; 41 Cameron Avenue, Apt. 1, Somerville, MA 02144 (US). WINGATE, David; 27 Mulberry Lane, Ashland, MA 01721 (US).
- (74) Agent: HARTMANN, Natalya; Patent Capital Group, 2816 Lago Vista Lane, Rockwall, TX 75032 (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).

Declarations under Rule 4.17:

- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))
- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))

Published:

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

(54) Title: SYSTEMS AND METHODS FOR SOURCE LOCALIZATION AND SEPARATION

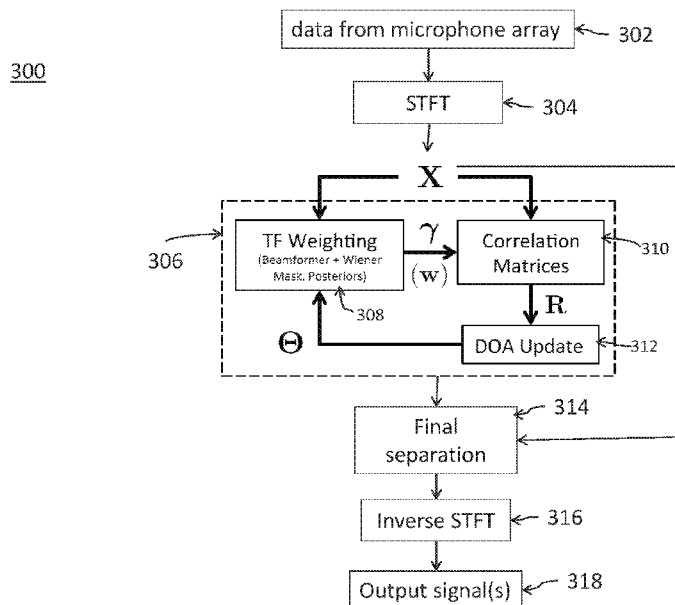
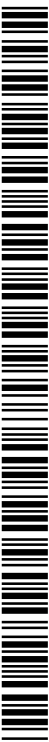


FIGURE 3

(57) Abstract: A method for identifying the direction of arrival of sound waves from first and second acoustic sources is disclosed. The method includes receiving, at a microphone array, acoustic signals including the sound waves from the first and second acoustic sources, converting the received acoustic signals from a time domain to a time-frequency domain, processing the converted acoustic signals to determine an estimated first angle representing the first direction of arrival and an estimated second angle representing the second direction of arrival, and updating the estimated first and second angles, where processing includes localizing, separating and Wiener post-filtering the converted acoustic signals using time-frequency weighting.



SYSTEMS AND METHODS FOR SOURCE LOCALIZATION AND SEPARATION

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of and priority from U.S. Provisional Patent Application Serial No. 62/093,903 filed 18 December 2014 entitled "SYSTEMS AND METHODS FOR SOURCE LOCALIZATION AND SEPARATION," which is incorporated herein by reference in its entirety.

TECHNICAL FIELD OF THE DISCLOSURE

[0002] The present invention relates to the field of signal processing, and in particular to source localization and/or separation.

BACKGROUND

[0003] Use of spoken input for user devices, including smartphones, automobiles, etc., can be challenging due to the fact that, typically, an acoustic environment in which a desired signal from a speaker is acquired also contains undesired signals from other acoustic sources. In such an environment, an acoustic sensor acquires an acoustic signal that has contributions from a plurality of different acoustic sources, where, as used herein, the term "contribution of an acoustic source" refers to at least a portion of an acoustic signal generated by a particular acoustic source, typically the portion being a portion of a particular frequency or a range of frequencies, at a particular time or range of times. When an acoustic source is e.g. a person speaking, there will be multiple contributions, i.e. there will be acoustic signals of different frequencies at different times generated by such a "source."

[0004] In a process generally referred to as "source separation," various digital signal processing techniques are used to recover the original component signals attributable to different sources from a combined signal acquired by the acoustic sensor (i.e. from the acquired signal that has a combination of contributions from different sources). A process of performing source separation without any prior information about

the acoustic signals is often referred to as “blind source separation” (BSS). Source separation can often be improved by processing acoustic signals acquired by multiple acoustic sensors, arranged e.g. in a sensor array, e.g. a microphone array. In such scenarios, each acoustic sensor acquires a corresponding signal that includes contributions from multiple sources and comparison of the signals acquired by different acoustic sensors provides an insight into individual contributions of the different sources.

[0005] In general, the term “source localization” refers to a process of determining spatial position of a particular source within a given environment. Various digital signal processing techniques usually use the term “Direction of Arrival” (DOA) to describe a parameter that indicates direction from which the signal generated by a particular source arrived, thus localizing the source within the environment.

[0006] Sound source localization and separation is used in many applications, including, for example, signal enhancement and noise cancellation for phones or hearing aids, speech recognition, home automation, and voice user interface in the car or home.

[0007] Typically, various source separation techniques use DOA in order to recover signals attributable to one or more of the individual sources. Thus, source localization typically precedes, or may be considered a part of, source separation. For example, many well-known source separation approaches use beamforming, i.e. signal processing techniques used to control the directionality of the reception of a signal, by employing arrays of acoustic sensors that aim to improve directional gain of the sensor array(s) by increasing the gain in the direction of a source of interest (e.g. a speaker) and decreasing the gain in the direction of interferences and noise. Beamforming techniques use information about the DOA of the source, and, therefore, are preceded by localization step where location of the source in the environment is determined or estimated.

[0008] One known approach for finding the DOAs is Steered Response Power (SRP) localization, which searches for peaks in the output power of a family of beamformers as a function of the DOA. In one example, each beamformer in a family of beamformers focuses on a specific direction. SRP localization can be used with a Maximum-Likelihood (ML) formulation. Another known approach computes the

Generalized Cross-Correlation (GCC) function, which can be used with a spectral weighting function such as Phase Transform (PHAT) to enhance the localizer.

[0009] A different known method for finding DOAs uses eigenanalysis of the data correlation matrix. For example, a Multiple Signal Classification (MUSIC) algorithm uses this method to identify signal and noise subspaces and form a MUSIC pseudospectrum that contains peaks at the source DOAs. The MUSIC pseudospectrum plots direction on the x-axis and likelihood of that direction as being the source of a sound on the y-axis, and thus is a function over the space of directions which indicates where sources are likely to be.

[0010] Another known method includes modeling observed data vectors as zero-means Gaussian random variables and using an EM algorithm to learn the sources' covariance parameters. The sources can be separated using multichannel Wiener filtering. According to some implementations, multichannel Wiener filtering can be used separate source signals from background noise. In some implementations, multichannel Wiener filtering can be used to separate speech signals from each other. In one example, in a multichannel case, in which there are multiple channels and multiple source signals, the output of the multichannel Wiener filter includes multiple sources and includes a correlation matrix that describes how the channels are correlated. The multichannel Wiener filter reconstructs source vectors directly.

[0011] The methods discussed above are sequential methods: first the DOA is estimated and the source is localized, and then the signal is separated from other signals and from background noise. One approach for simultaneously localizing and separating various sounds sources uses Bayesian analysis. However, Bayesian analysis uses prior information about the sources, which may not always be available. For example, Bayesian analysis requires prior information about the magnitudes of the sources.

[0012] As the foregoing illustrates, improvements with respect to source localization and separation are desired.

OVERVIEW

[0013] A more effective and efficient method for localizing and separating signals is provided, and involves interpreting the SRP function as a probability distribution and maximizing it as a function of the source DOAs. In one method, a mixture of single-source SRPs (MoSRP) is used. In a second method, an SRP that explicitly models the presence of multiple sources is provided (MultSRP). Some advantages of the second method include simultaneous localization of each of the multiple sources and explicit modeling of interference between sources. Time-Frequency (TF) masking is used to isolate TF bins, described in greater detail below, that correspond to directional signals of interest, thereby merging the localization, separation and Wiener post-filtering steps into one unified approach.

[0014] According to some embodiments, an improved type of Wiener filter may be used for estimating a weight for each of multiple TF bins for each of multiple sources. The weight estimates for each time-frequency bin can be used to determine which bins contain source energy and which bins do not contain source energy. Bins which do not contain source energy may still contain energy, for example, noise. For each time frequency bin, a Wiener filter coefficient is estimated, where the Wiener filter coefficient corresponds to the probability that any of the directional sources are present.

[0015] According to one aspect, a method is provided for identifying a first direction of arrival of sound waves (i.e. acoustic signals) from a first acoustic source and a second direction of arrival of sound waves from a second acoustic source. The method includes receiving, at a microphone array, acoustic signals including a combination of the sound waves from the first and second acoustic sources, converting the received acoustic signals from a time domain to a time-frequency domain, processing the converted acoustic signals to determine an estimated first angle representing the first direction of arrival and an estimated second angle representing the second direction of arrival, and updating the estimated first and second angles. The processing includes localizing, separating and Wiener post-filtering the converted acoustic signals using time-frequency weighting and outputting a time-frequency weighted signal for estimating the first and

second angles. In one example, converting the received acoustic signals from a time domain to a time-frequency domain includes using a short time Fourier transform.

[0016] According to some implementations, the method includes combining the time-frequency weighted signal with the converted acoustic signals to generate a correlation matrix. In some implementations, updating the estimated first and second angles comprises utilizing the correlation matrix and the estimated first and second angles and outputting updated estimated first and second angles.

[0017] According to some implementations, processing the converted acoustic signals to determine the estimated first and second angles includes decomposing the converted acoustic signals to identify signals from each of the first and second acoustic sources by accounting for interference between the first and second acoustic sources in forming the acoustic signals. In some implementations, processing the converted acoustic signals and updating the first and second estimated angles includes iteratively decomposing the converted acoustic signals to simultaneously determine the first and second directions of arrival. In one example, processing the converted acoustic signals includes processing using steered response power localization.

[0018] According to some implementations, the method further includes using an inverse STFT to convert the processed converted acoustic signals back into the time domain and separating the sound waves from the first acoustic source from the sound waves from the second acoustic source.

[0019] As will be appreciated by one skilled in the art, aspects of the present disclosure may be embodied in various manners – e.g. as a method, a system, a computer program product, or a computer-readable storage medium. Accordingly, aspects of the present disclosure may take the form of an entirely hardware embodiment, an entirely software embodiment (including firmware, resident software, micro-code, etc.) or an embodiment combining software and hardware aspects that may all generally be referred to herein as a "circuit," "module" or "system." Functions described in this disclosure may be implemented as an algorithm executed by one or more processing units, e.g. one or more microprocessors, of one or more computers. In various embodiments, different steps and portions of the steps of each of the methods

described herein may be performed by different processing units. Furthermore, aspects of the present disclosure may take the form of a computer program product embodied in one or more computer readable medium(s), preferably non-transitory, having computer readable program code embodied, e.g., stored, thereon. In various embodiments, such a computer program may, for example, be downloaded (updated) to the existing devices and systems (e.g. to the existing radar or sonar receivers or/and their controllers, etc.) or be stored upon manufacturing of these devices and systems.

[0020] Other features and advantages of the disclosure are apparent from the following description, and from the claims.

BRIEF DESCRIPTION OF THE DRAWING

[0021] To provide a more complete understanding of the present disclosure and features and advantages thereof, reference is made to the following description, taken in conjunction with the accompanying figures, wherein like reference numerals represent like parts, in which:

[0022] FIGURE 1 is a diagram illustrating an audio processor receiving signals from multiple sources, according to some embodiments of the disclosure;

[0023] FIGURE 2 is a diagram illustrating a method for identifying a first direction of arrival of sound waves from a first acoustic source and a second direction of arrival of sound waves from a second acoustic source, according to some embodiments of the disclosure;

[0024] FIGURE 3 is one diagram illustrating a method for separating and localizing signals, according to some embodiments of the disclosure;

[0025] FIGURE 4 is a diagram illustrating two data vectors from two sources and the combination of the two data vectors, according to some embodiments of the disclosure;

[0026] FIGURE 5A is a diagram illustrating single-source likelihood over DOAs, according to some embodiments of the disclosure;

[0027] Figure 5B is a diagram illustrating a multi-source SRP likelihood for a data mixture of two sources over a joint space of all DOA pairs, according to some embodiments of the disclosure; and

[0028] Figure 6 is another diagram illustrating a MultSRP method for separating and localizing signals, according to some embodiments of the disclosure.

DESCRIPTION OF EXAMPLE EMBODIMENTS OF THE DISCLOSURE

[0029] FIGURE 1 is a diagram 100 illustrating an audio processor 102 receiving signals from first 104a, second 104b, and third 104n sources, according to some embodiments of the disclosure. The audio processor 102 includes a microphone array 106, a direction finding module 108, a source separating module 110, and an audio processing module 112.

[0030] The microphone array 106 receives (i.e. acquires) a combined sound, referred to in the following as “ambient sound,” including the signals from the first 104a, second 104b and third 104n sources. In other examples, the microphone array 106 receives ambient sound includes signals from more than three sources, and there may be any number of sources present.

[0031] The microphone array 106 may include one or more acoustic sensors, arranged e.g. in a sensor array, each sensor of the array configured to acquire an ambient sound (i.e., each acoustic sensor acquires a corresponding signal). In some embodiments where a plurality of acoustic sensors are employed, the sensors may be provided relatively close to one another, e.g. less than 2 centimeters (cm) apart, preferably less than 1 cm apart. In an embodiment, the sensors may be arranged separated by distances that are much smaller, on the order of e.g. 1 millimeter (mm) or about 300 times smaller than typical sound wavelength, where beamforming techniques, used e.g. for determining DOA of an acoustic signal, do not apply. In other embodiments, the sensors may be provided at larger distances with respect to one another.

[0032] While some embodiments where a plurality of acoustic sensors are employed make a distinction between the signals acquired by different sensors (e.g. for the purpose of determining DOA by e.g. comparing the phases of the different signals),

other embodiments may consider the plurality of signals acquired by an array of acoustic sensors as a single signal, possibly by combining the individual acquired signals into a single signal as is appropriate for a particular implementation. Therefore, in the following, when an “acquired signal” is discussed in a singular form, then, unless otherwise specified, it is to be understood that the signal may comprise several acquired signals acquired by different sensors of the microphone array 106.

[0033] Different source localization and separation techniques presented herein are based on computing time-dependent spectral characteristics X of the signal acquired by the microphone array 106. A characteristic could e.g. be a quantity indicative of a magnitude of the acquired signal. A characteristic is “spectral” in that it is computed for a particular frequency or a range of frequencies. A characteristic is “time-dependent” in that it may have different values at different times.

[0034] In an embodiment, such characteristics may be a Short Time Fourier Transform (STFT), computed as follows. An acquired signal is functionally divided into overlapping blocks, referred to herein as “frames.” For example, frames may be of a duration of 64 milliseconds (ms) and be overlapping by e.g. 48 ms. The portion of the acquired signal within a frame is then multiplied with a window function (i.e. a window function is applied to the frames) to smooth the edges. As is known in signal processing, and in particular in spectral analysis, the term “window function” (also known as tapering or apodization function) refers to a mathematical function that has values equal to or close to zero outside of a particular interval. The values outside the interval do not have to be identically zero, as long as the product of the window multiplied by its argument is square integrable, and, more specifically, that the function goes sufficiently rapidly toward zero. In typical applications, the window functions used are non-negative smooth “bell-shaped” curves, though rectangle, triangle, and other functions can be used. For instance, a function that is constant inside the interval and zero elsewhere is called a “rectangular window,” referring to the shape of its graphical representation. Next, a transformation function, such as e.g. Fast Fourier Transform (FFT), is applied transforming the waveform multiplied by the window function from a time domain to a frequency domain. As a result, a frequency decomposition of a portion of the acquired

signal within each frame is obtained. The frequency decomposition of all of the frames may be arranged in a matrix where frames and frequency are indexed (in the following, frames are described to be indexed by "t" and frequencies are described to be indexed by "f"). Each element of such an array, indexed by (f, t) comprises a complex value resulting from the application of the transformation function and is referred to herein as a "time-frequency (TF) bin" or simply "bin." The term "bin" may be viewed as indicative of the fact that such a matrix may be considered as comprising a plurality of bins into which the signal's energy is distributed.

[0035] Time-frequency bins come into play in BSS algorithms in that separation of a particular acoustic signal of interest (i.e. an acoustic signal generated by a particular source of interest) from the total signal acquired by an acoustic sensor may be achieved by identifying which bins correspond to the signal of interest, i.e. when and at which frequencies the signal of interest is active. Once such bins are identified, the total acquired signal may be masked by zeroing out the undesired time-frequency bins. Such an approach would be called a "hard mask." Applying a so-called "soft mask" is also possible, the soft mask scaling the magnitude of each bin by some amount. Then an inverse transformation function (e.g. inverse STFT) may be applied to obtain the desired separated signal of interest in the time domain. Thus, masking in the frequency domain (i.e. in the domain of the transformation function) corresponds to applying a time-varying frequency-selective filter in the time domain. The desired separated signal of interest may then be selectively processed for various purposes.

[0036] In one example, each source 104a, 104b, 104n has a distinct location, and the signal from each source 104a, 104b, 104n arrives at the microphone array 106 at an angle relative to its source location. Based on this angle, for each signal, the audio processor 102 estimates a direction-of-arrival (DOA). Thus, at the audio processor 102, each source 104a, 104b, 104n has a DOA 114a, 114b, 114n. The first source 104a has a first DOA 114a, the second source 104b has a second DOA 114b, and the third source 104n has a third DOA 114n.

[0037] The microphone array 106 is coupled to the direction finding module 108, and the signals received at the microphone array 106 are transmitted to the direction

finding module 108. The direction finding module 108 estimates the DOAs 114a, 114b, and 114n associated with source signals 104a, 104b, and 104n, as described in greater detail below. The direction finding module 108 is coupled to a separation masking module 110, where the signals corresponding to the various sources 104a, 104b, 104n are separated from each other and from background noise which may be present. The direction finding module 108 and the separation masking module 110 are each also coupled to a further audio processing module 112, where further processing of the acoustic signals occurs. The further audio processing may depend on the application, and may include, for example, enhancing one or more speech signals, and filtering out constant noise or repetitive sounds.

[0038] In traditional array processing, linear filtering algorithms are used for enhancing directional signals. In particular, beamforming is used to constructively add signals received at microphones in the array and suppress noise. There are several different methods for beamforming including Delay-and-Sum (DS) beamforming, Minimum Variance Distortionless Response (MVDR) beamforming, Linearly-Constrained Minimum-Variance (LCMV) beamforming, and Multiple Signal Classification (MUSIC) beamforming.

[0039] In one example, Delay-and-Sum (DS) beamforming involves adding a time delay to the signal recorded from each microphone that cancels out the delay caused by the extra travel time that it took for the signal to reach the microphone (as opposed to microphones that were closer to the signal source). Summing the resulting in-phase signals enhances the signal. This beamforming method can be used to estimate DOA by testing various time delays, since the delay that correlates with the correct DOA will amplify the signal, while incorrect time delays destructively interfere with the signal. The DS beamforming method focuses on the time domain to estimate DOA, and it is inaccurate in noisy environments.

[0040] In another example, Delay-and-Sum (DS) beamforming involves fractional delays in the frequency domain. Generally, when a small microphone array is used, the received signals are processed by measuring the fractional delays in the signals, weighting each channel by a complex coefficient, and adding up the results. According to

one implementation, DS beamforming is used in processing received signals in the single source model described below.

[0041] MVDR beamforming is similar to DS beamforming, but takes into account statistical noise correlations between the channels.

[0042] A Fourier transform can be used to transform the time domain signal into the time-frequency plane by converting time delays between sensors into phase shifts. MVDR beamforming provides good noise suppression by minimizing the output power of the array while not distorting signals from the primary DOA, but it has a power defined by a matrix inversion, and is therefore computationally intensive. The MVDR beamformer solution is:

$$\hat{\mathbf{w}}_{MVDR} = \frac{\mathbf{R}_f^{-1} \mathbf{a}_f(\phi)}{\mathbf{a}_f^H(\phi) \mathbf{R}_f^{-1} \mathbf{a}_f(\phi)} \quad (1)$$

[0043] MVDR and DS beamformers are generalized to the Multi-source case via a multiply-constrained optimization problem, and the solution is the Linearly-Constrained Minimum-Variance (LCMV) beamformer. In particular, in the LCMV beamformer, the weight vector can be used to determine how to weight the channels in the time-frequency plane to preserve energy from desired directions and suppress energy from other directions:

$$\hat{\mathbf{w}}_{LCMV} = \mathbf{R}_f^{-1} \mathbf{A}_f(\Phi) \left[\mathbf{A}_f^H(\Phi) \mathbf{R}_f^{-1} \mathbf{A}_f(\Phi) \right]^{-1} \mathbf{u} \quad (2)$$

[0044] The beamformers discussed above can be used to estimate the coefficients when the source DOA(s) Φ are already known. Thus, in systems using the LCMV, MVDR and DS beamforming methods described above, first the source DOA(s) are determined and then beamforming is performed. The source DOA(s) may be determined using, for example, Steered Response Power Localization as described below.

[0045] The MUSIC beamformer is a subspace method based on an eigenanalysis of the covariance matrix. The MUSIC beamformer requires an eigendecomposition. Additionally, MUSIC is based on the assumption that the subspace that the signals lie in is

orthogonal to the space in which the noise lies. In one example, the MUSIC beamformer decomposes a covariance matrix representing the signal and noise of the received signal.

[0046] Steered Response Power (SRP) Localization is used to estimate source DOA(s) Φ . In some examples, SRP localization is used to estimate DOA's by discretizing the direction space. In particular, source DOAs Φ estimated by SRP Localization can be input in the LCMV beamforming equation (2) above. SRP localization identifies DOAs Φ by searching for peaks in the output power of a single-source beamformer.

[0047] When multiple sources are present, there may be multiple peaks in the SRP. However, in compact microphone arrays or closely spaced microphone arrays, close spacing of the microphone elements makes the steering vectors hard to distinguish, and thus low frequency peaks are poorly localized. Additionally, if the source coefficients are simultaneously large in magnitude, the SRP function is distorted by cross-terms.

[0048] A more accurate and effective approach is to scan all DOA sets Φ using an LCMV beamformer and locate the peak output power. However, this is computationally inefficient and too time-consuming for real-time feedback, since discretizing the DOA search space into D look directions results in D^K Φ 's to be scanned (where K is the number of sources present). Instead, according to one implementation, the multi-source SRP function is modeled as a continuous likelihood function parametrized by Φ and the likelihood function is maximized to identify source DOAs.

[0049] FIGURE 2 is a diagram illustrating a method 200 for identifying a first direction of arrival of sound waves from a first acoustic source and a second direction of arrival of sound waves from a second acoustic source. The method includes, at step 202, receiving, at a microphone array, acoustic signals including the sound waves from the first and second acoustic sources. At step 204, the received acoustic signals, now represented by electrical signals generated by the microphone array, are converted from a time domain to a time-frequency domain. At step 206, the converted acoustic signals are processed to determine an estimated first angle representing the first direction of arrival and an estimated second angle representing the second direction of arrival. Processing includes localizing, separating and Wiener post-filtering the converted acoustic signals using time-frequency weighting and outputting a time-frequency

weighted signal for estimating the first and second angles. At step 208, the estimated first and second angles are updated. According to one feature, the likelihood of the first and second angles is determined integrally as a single unit from the mixed signals received at the microphone array, rather than maximizing the likelihood of each of the first and second angles separately.

[0050] FIGURE 3 is a diagram illustrating a method 300 for separating and localizing signals, according to some embodiments of the disclosure. As shown in Figure 3, the method 300 is an iterative approach in which a probabilistic SRP model is combined with time-frequency masking to perform blind source separation and localization in the presence of non-stationary interference. In particular, the method has an iterative loop 306 including a Time-Frequency (TF) weighting step 308, a correlation matrices step 310, and a direction of arrival (DOA) update step 312.

[0051] The method 300 begins with receiving input acoustic signals x 302 acquired by different microphone elements of the microphone array 106. As described above, each acquired signal 302 may, and typically will, include contributions from multiple sources 104a-104n and a goal of source separation is to distinguish these individual contributions on a per-source basis.

[0052] The acquired input acoustic signals 302 are processed through an STFT 304 to transform the signals from the time domain to the time-frequency plane.

[0053] The output X from the STFT 304 is input to the TF weighting step 308 and to the correlation matrices step 310. The TF weighting step 308 uses TF masking to isolate TF bins that correspond to selected directional signals. In particular, some directional signals are identified as being directional signals of interest, and the corresponding TF bins are isolated. Identifying the directional signal or signals of interest may include separating identified signals, and selecting one (or more) of the separated signals. In one example, the selected signal corresponds to a speech signal, and it may be the speech of a particular speaker.

[0054] In one example, the selected directional signals are identified based on peaks in output power. The TF weighting step 308 receives the output signals from the STFT step 304 as well as a DOA set Θ (DOA matrix) from the DOA update step 312, and

uses these inputs to perform TF weighting as described in greater details in Equations 3-17. Thus, the localization, separation, and Wiener post-filtering steps are merged into the TF weighting step 308.

[0055] The output γ from the TF weighting step 308 is input into the correlation matrices step 310. The correlation matrices step 310 combines the TF weighted input and data output from the STFT 304. The correlation matrices step 310 uses the inputs to derive correlation matrices as described in greater detail below with respect to equations 15 and 16, and outputs an updated correlation matrix R to the DOA update step 312. The DOA update step revises the set of DOA's Θ based on the input correlation matrix R , and outputs the updated DOA's Θ to the TF weighting step 308.

[0056] Following the iterative loop 306 of the method 300 shown in Figure 3, an output set of DOA's Θ indicating the localization results is output from the DOA update step 312 to a final separation step 314. The separation step 314 also receives the STFT processed data \mathcal{X} as input. At the separation step 314, the set of DOA's Θ is used to separate out the signals in the data \mathcal{X} and generates an STFT matrix for each source. The STFT matrices are processed with an inverse STFT at step 316, which transforms each one into a time domain signal. The time domain signals 318 output from step 316 are localized, separated and post-filtered output signals.

[0057] According to various implementations, the method 300 is performed using the following equations.

[0058] According to some implementations a first method for maximizing the SRP as a function of the source DOA's uses an SRP that explicitly models the presence of multiple sources.

[0059] Identifying the DOAs involves maximizing a likelihood function:

$$\mathcal{L}(\theta_1, \theta_2; \mathcal{X}) \quad (3)$$

[0060] where \mathcal{X} is the STFT coefficients of the data from the microphone array, and θ_1 and θ_2 are estimated DOA angles. A Gaussian likelihood for the observed data vectors \mathcal{X}_{f_t} is:

$$\log \mathcal{L}_{f_t}(\theta) = \log \mathcal{N}(\mathcal{X}_{f_t}; \mu_{f_t}, \sigma_f^2 \mathbf{I}) \quad (4)$$

[0061] where the mean μ_{ft} encodes the expected value of \mathbf{x}_{ft} , and σ_f^2 represents the variance of the background noise at frequency f , and \mathbf{I} is the identity matrix, and:

$$\mu_{ft} = \mathbb{E}[\mathbf{x}_{ft}] = \mathbf{A}_f(\Theta) \mathbb{E}[s_{ft}] \quad (5)$$

[0062] for a hypothesized DOA set Θ , where \mathbf{A}_f is the steering matrix including the observed mixing vectors \mathbf{a}_f as elements, and s_{ft} is a vector of complex source coefficients for a time-frequency bin with one component for each source. The expectation $\mathbb{E}[s_{ft}]$ can be approximated with a least squares estimate:

$$\hat{\mathbf{s}}_{ft} = [\mathbf{A}_f^H(\Theta) \mathbf{A}_f(\Theta)]^{-1} \mathbf{A}_f^H(\Theta) \mathbf{x}_{ft} \quad (6)$$

[0063] which is the output of a LCMV beamformer with $\mathbf{R}_f \propto \mathbf{I}$, where H is a Hermitian transpose. In some implementations, there can be regularization within the brackets in equation (6) to make sure that the matrix inverse is well-conditioned. And therefore:

$$\hat{\boldsymbol{\mu}}_{ft} = \mathbf{A}_f(\Theta) \hat{\mathbf{s}}_{ft} = \mathbf{B}_f(\Theta) \mathbf{x}_{ft} \quad (7)$$

[0064] where:

$$\mathbf{B}_f(\Theta) = \mathbf{A}_f(\Theta) [\mathbf{A}_f^H(\Theta) \mathbf{A}_f(\Theta)]^{-1} \mathbf{A}_f^H(\Theta) \quad (8)$$

[0065] Thus, the likelihood of a particular DOA set Θ is:

$$\log \mathcal{L}_{ft}(\Theta) \propto -\frac{1}{2\sigma_f^2} \mathbf{x}_{ft}^H \mathbf{P}_f(\Theta) \mathbf{x}_{ft} \quad (9)$$

[0066] where, in the log domain above, the proportionality sign \propto means equality up to an additive constant (rather than up to a multiplicative constant). This can be aggregated over time t and expanded:

$$\log \mathcal{L}_{ft}(\Theta) \propto -\frac{1}{2\sigma_f^2} \sum_{t=1}^T \|\mathbf{x}_{ft}\|_2^2 - \mathbf{x}_{ft}^H \mathbf{B}_f(\Theta) \mathbf{x}_{ft} \quad (10)$$

[0067] Using the above equations, the DOAs of signals from multiple sources can be efficiently determined more accurately than by previous methods.

[0068] FIGURE 4 is a diagram 400 illustrating a first 402 and second 404 data vectors from first and second sources and the combination 406 of the two data vectors

402 and 404, according to some embodiments of the disclosure. The diagram 400 illustrates the additivity of the first 402 and second 404 data vectors. As illustrated in Figures 5A and 5B, due to interference between the first 402 and second 404 data vectors, a spurious peak in the single source likelihood is present between the true DOAs. If a single-source likelihood is calculated for the superposition of the first 402 and second 404 data vectors, the single source likelihood will indicate the likelihood of a single source at the combination data vector 406 source. This is illustrated in FIGURE 5A, which shows is a diagram 500 illustrating single-source likelihood over DOAs, according to some embodiments of the disclosure. In particular, the diagram 500 shows the single source likelihood, with a peak indicating a DOA around 1.3-1.4 radians. Thus, the single source likelihood equation estimates a single source positioned between the first and second sources, rather than the two separate sources.

[0069] Figure 5B is a diagram 550 illustrating a multi-source SRP likelihood for a data mixture of two sources over a joint space of all DOA pairs, according to some embodiments of the disclosure. The data shown in Figure 5B is derived using equation (10) above, which estimated the first and second sources as having a DOAs at 0.56 radians and at 2.26 radians on the unit circle.

[0070] According to some implementations a second method for maximizing the SRP as a function of the source DOA's uses a mixture of single-source SRP's.

[0071] Maximum likelihood estimation of the source DOAs can be estimated using a gradient ascent on the SRP likelihood shown above in equation 10:

$$\Theta^i \leftarrow \Theta^{i-1} + \eta \Omega \left(\sum_{f=1}^F \sum_{t=1}^T \frac{\partial \log \mathcal{L}_{ft}(\Theta)}{\partial \Theta} \Big|_{\Theta^{i-1}} \right) \quad (11)$$

[0072] where η is the step size, and Ω is a function that normalizes the gradient, which appears in parentheses after the Ω . The gradient indicates which direction corresponds with an improvement in DOA estimates. The step size η is how far to move in the indicated direction. The maximum likelihood can be estimated for both the one-source model and the multiple source model. For the one-source model, the gradient can be

$$\frac{\partial \mathcal{L}_{f,t}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{\sigma_f^2} \frac{2\pi f}{c} \mathbf{m} \operatorname{Im} [\mathbf{a}_f^*(\boldsymbol{\theta}) \odot (\mathbf{R}_f \mathbf{a}_f(\boldsymbol{\theta}))] \quad (12)$$

[0073] where \odot denotes element-wise multiplication, and \mathbf{m} is the matrix of microphone positions (a matrix in which the columns are the positions of the microphones). The single source model can be used when multiple sources are present by modeling the presence of the other sources at each time t with hidden variables z_{ft} that capture which source is active at any selected time. In one example, an Expectation-Maximization (EM) algorithm is used to iterate between estimating z_{ft} 's and DOAs:

$$\log p(\mathbf{X}_f, \mathbf{z}_f | \boldsymbol{\Theta}) \propto \frac{1}{2\sigma_f^2} \sum_{t=1}^T \sum_{k=1}^K [z_{ft} = k] |\mathbf{a}_f^H(\boldsymbol{\theta}_k) \mathbf{x}_{ft}|^2 \quad (13)$$

[0074] the lower bound of the EM algorithm is:

$$\mathcal{Q}_f \propto \frac{1}{2\sigma_f^2} \sum_{k=1}^K \mathbf{a}_f^H(\boldsymbol{\theta}_k) \widehat{\mathbf{R}}_f^{(k)} \mathbf{a}_f(\boldsymbol{\theta}_k) \quad (14)$$

[0075] where

$$\widehat{\mathbf{R}}_f^{(k)} = \frac{\sum_{t=1}^T w_{ft}^{(k)} \mathbf{x}_{ft} \mathbf{x}_{ft}^H}{\sum_{t=1}^T w_{ft}^{(k)}} \quad (15)$$

[0076] are source-specific correlation matrices, and are defined in terms of the posterior probabilities of the z_{ft} 's:

$$w_{ft}^{(k)} = p(z_{ft} = k | \mathbf{x}_{ft}) \propto \frac{|\mathbf{a}_f^H(\boldsymbol{\theta}_k) \mathbf{x}_{ft}|^2}{2\sigma_f^2} \quad (16)$$

[0077] The equations 13-16 show one way to use the single source method of equation 12 for multiple sources. According to other implementations, equation 17 can be used for localization of multiple sources. According to one feature, in the E step, soft TF weights are determined, and in the M step, each source's DOA is optimized. Thus, the EM method alternates between estimating localization (DOA) parameters and estimating separation (TF mask) parameters.

[0078] According to one implementation, the gradient in the multiple source case is:

$$\frac{\partial \log \mathcal{L}_{ft}(\Theta)}{\partial \Theta} = \frac{1}{\sigma_f^2} \frac{2\pi f}{c} \text{mIm} \left[\mathbf{A}_f^* \odot \left(\mathbf{P}_f \mathbf{R}_f \mathbf{A}_f [\mathbf{A}_f^H \mathbf{A}_f]^{-1} \right) \right] \quad (17)$$

[0079] This multiple source case takes cross-talk into account while avoiding the complexity of the EM algorithm.

[0080] The localization accuracy in the presence of ambient noise can be improved using Wiener filtering. This may be done in step 308 of the method 300 shown in Figure 3. In the presence of non-directional interference:

$$\mathbf{x}_{ft} = \mathbf{A}_f(\Phi) \mathbf{s}_{ft} + \mathbf{n}_{ft} + \mathbf{e}_{ft} = \mathbf{b}_{ft} + \mathbf{c}_{ft} \quad (18)$$

[0081] where $\mathbf{b}_{ft} = \mathbf{A}_f(\Phi) \mathbf{s}_{ft}$ and $\mathbf{c}_{ft} = \mathbf{n}_{ft} + \mathbf{e}_{ft}$. According to one example, the MMSE-optimal weighting to recover \mathbf{b}_{ft} is given by the Wiener mask:

$$\gamma_{ft}^{opt} = \frac{\|\mathbf{b}_{ft}\|_2^2}{\|\mathbf{b}_{ft}\|_2^2 + \|\mathbf{c}_{ft}\|_2^2} \quad (19)$$

[0082] Thus, a robust estimate of the correlation matrices is:

$$\hat{\mathbf{R}}_f = \frac{\sum_{t=1}^T \gamma_{ft}^{opt} \mathbf{x}_{ft} \mathbf{x}_{ft}^H}{\sum_{t=1}^T \gamma_{ft}^{opt}} \quad (20)$$

[0083] The weights can be approximated as using equations (21) and (22):

$$\hat{\mathbf{b}}_{ft} = \mathbf{B}_f(\Theta) \mathbf{x}_{ft} \quad (21)$$

$$\hat{\mathbf{c}}_{ft} = \mathbf{x}_{ft} - \hat{\mathbf{b}}_{ft} \quad (22)$$

[0084] According to one feature, interleaving the Wiener masking with DOA optimization improves localization accuracy in the presence of ambient noise. In some implementations, for a mixture of one source models, the correlation matrices shown in

Equation 15 can be estimated by multiplying the posteriors with the Wiener filter weights.

[0085] According to one implementation, the sources can be separated by applying TF masks with weights. In various examples, this may be done in one or more of step 308 and step 314 of the method 300. For example, the following equation can be used:

$$\log w_{ft}^{(k)} \propto \frac{1}{2\nu^2} |\hat{s}_{ft}^{(k)}|^2 \quad (23)$$

[0086] where the source coefficients \hat{s}_{ft} are recovered with LCMV beamforming. The variance is related to the hardness of the mask, such that as the variance moves to zero, the mask becomes binary. The masks can be applied to corresponding components of \hat{s}_{ft} and followed with a Wiener masking step to suppress non-speech interference and reduce the presence of masking artifacts.

[0087] FIGURE 6 is a diagram illustrating a method 600 for separating and localizing signals, according to some embodiments of the disclosure. The method 600 may be considered as a summary, or an alternative representation, of the method 300 described above. Therefore, in the interests of brevity, some steps illustrated in method 600 refer to steps illustrated in method 300 in order to not repeat all of the details of their descriptions.

[0088] The method 600 may be considered as including three main stages: stage 610 that may be referred to as a preprocessing stage, stage 620 that may be referred to as an optimization stage, and stage 630 that may be referred to as a source separation stage.

[0089] As shown in FIGURE 6, the preprocessing stage 610 may include steps 612, 614, 616, and 618. In step 612, acoustic signals are captured by the microphone array 106, as described above with reference to 302. The captured signals 612 may be considered as multiple discrete-time signals $x_m, m = 1, \dots, M$, where m is an integer indicating a particular acoustic sensor of the microphone array 106 comprising M acoustic sensors (i.e. $m = 1; \dots; M$).

[0090] In step 614, STFT is applied to the captured signals x_m . in order to convert the captured signals into the TF domain resulting in complex-values matrices

$$\mathbf{X}_m \in \mathbb{C}^{F \times T}. \quad (24)$$

[0091] The magnitude portion of these matrices may be removed to give

$$\hat{\mathbf{X}} = \mathbf{X} / (\sqrt{M}|\mathbf{X}|) \quad (25)$$

[0092] In step 616, correlation matrices are initialized by estimating correlation matrices for each frequency as:

$$\hat{\mathbf{R}}_f = \frac{1}{T} \sum_{t=1}^T \hat{\mathbf{x}}_{f,t} \hat{\mathbf{x}}_{f,t}^H \quad (26)$$

[0093] In step 618, the DOA parameter matrix

$$\Theta = [\theta^{(1)} \dots \theta^{(k)} \dots \theta^{(K)}] \quad (27)$$

is initialized with Θ_0 where

$$\theta_k \in \mathbb{R}^3 \quad (28)$$

is the unit vector describing the orientation of the kth acoustic source (k being an integer between 1 and n for the acoustic sources 104 illustrated in FIGURE 1) relative to the microphone array 106.

[0094] The initialization of step 618 may be carried out in different manners, including e.g. SRP localization described above.

[0095] As shown in FIGURE 6, the initialized DOA matrix Θ_0 is provided to the optimization stage 620. As shown in FIGURE 6, the optimization stage 620 may include steps 622, 624, 626, and 628 which may be iteratively repeated for a number of iterations I_{max} , in order to improve the estimate of the DOA matrix Θ (i.e. in order to improve DOA estimates for the different acoustic sources 104). The number of iterations I_{max} may be determined by various stopping conditions. For example, in some embodiments, the maximum number of iterations may be pre-defined, while, in other embodiments, iterations may be performed until a certain condition is met, such as e.g. a

pre-specified threshold in the percentage improvement of the likelihood value given by equation (9).

[0096] In step 622, for each frequency, a steering matrix A_f , described above with reference to equation (5) and subsequent equations, is computed as:

$$A_f = \frac{1}{\sqrt{M}} e^{j \frac{2\pi l_f}{c} \mathbf{m}^T \Theta} \quad (29)$$

where l_f is the frequency in Hertz at the f^h frequency band, c is the speed of sound, and

$$\mathbf{m} \in \mathbb{R}^{3 \times M} \quad (30)$$

is a matrix of microphone locations.

[0097] For each frequency, a projector matrix may then be computed as shown above with the equation (8).

[0098] Steering matrices A and projection matrices B may then be, optionally, provided to step 624. In step 624, if Wiener masking described above is used, new correlation matrices are re-estimated as described above with reference to equations (19)-(20). In an embodiment, equations (20) and (19) for re-estimating the new correlation matrices may be re-written as equations (31) and (32) below:

$$\hat{\mathbf{R}}_f = \sum_{n=1}^T \hat{\gamma}_n \tilde{\mathbf{x}}_f \tilde{\mathbf{x}}_f^H \quad (31)$$

$$\hat{\gamma}_n = \frac{\|\mathbf{B} \tilde{\mathbf{x}}_f\|}{\|\tilde{\mathbf{x}}_f\|} \quad (32)$$

[0099] In step 626, a DOA gradient matrix may be computed as

$$\mathbf{G} = \sum_{f=1}^F l_f \mathbf{m} \text{Im} \left[\mathbf{A}_f^* \odot \left((\mathbf{I} - \mathbf{B}_f) \hat{\mathbf{R}}_f \mathbf{A}_f [\mathbf{A}_f^H \mathbf{A}_f]^{-1} \right) \right] \quad (33)$$

[00100] Equation (33) is an exemplary explicit equation for the gradient given in equation (17) above.

[00101] The columns of the gradient matrix given by the equation (33) are normalized as:

$$\mathbf{g}^{(k)} \leftarrow \mathbf{g}^{(k)} / \|\mathbf{g}^{(k)}\|_2 \quad (34)$$

[00102] The gradient matrix G is provided to step 628 where the DOA matrix Θ is adjusted as described with reference to equation (11) above. In particular, the DOA matrix is adjusted as

$$\Theta \leftarrow \Theta + \eta_i G \tag{35}$$

where the step size at the i^{th} iteration is

$$\eta_i = \eta_0 (I_{max} - i) / I_{max} \tag{36}$$

[00103] The columns of the DOA matrix may be normalized as:

$$\theta^{(k)} \leftarrow \theta^{(k)} / \|\theta^{(k)}\|_2 \tag{37}$$

[00104] While equation (33) provides an explicit equation for the gradient given in equation (17) above, step 628 describes the gradient procedure given an appropriate gradient as given in equations (11) and (12) above.

[00105] Step 624 may be performed as a part of 308 and 310 described above, while step 628 corresponds to 312 described above.

[00106] Updated DOA matrix Θ is then provided to the source separation, as illustrated in FIGURE 6 with Θ provided to the separation stage 630 and as illustrated in FIGURE 3 with an arrow from 306 to the final separation step 314.

[00107] As shown in FIGURE 6, the source separation stage 630 may include steps 632 and 634. Following the iterative procedure described above, any number of methods may be used to enhance/separate the directional signals, all of which methods are within the scope of the present disclosure. In one embodiment, in step 632, each source 104 may be isolated by estimating TF masks and applying them to the STFT X. As previously described herein, according to one implementation, the sources can be separated by applying TF masks with weights, which could be done in one or more of step 308 and step 314 of the method 300, using equation (23) provided above using estimates of the source coefficients provided by K LCMV beamformers, each designated to isolate a single source while blocking out, or at least substantially suppressing, the others. In one embodiment, this may be implemented as:

$$\hat{s}_k = [A_k^H A_k]^{-1} A_k^H x_{ft} \tag{38}$$

[00108] The variance $\nu \in [0, \infty)$ controls the hardness of the mask such that as $\nu \rightarrow 0$, the mask becomes binary, assigning each TF bin entirely to a single source.

[00109] In step 634, these masks are applied to any single captured signal (i.e. to any signal captured by one of the acoustic sensors of the microphone array 106) and inverted to the time-domain using inverse STFT, as described above with reference to 316.

[00110] The method 600 is presented for the case of an SRP that explicitly models the presence of multiple sources, i.e. method 600 is a MultSRP method. A method for the mixture of single-source SRPs (MoSRP) would include steps analogous to those illustrated in FIGURE 6 with the main difference residing in the gradients of the two methods, in particular in how the correlation information is used (i.e. the difference between MultSRP and MoSRP is in re-computing the correlation matrices as is done in step 624 described above). For MoSRP, step 624 would involve including posterior probability weights in re-computing the correlation matrices as in equation (15). Gradients for the MoSRP method are given in equation (12).

[00111] The methods for source localization and separation described above may be summarized as follows. In the following summary, third rank tensors are represented with capital letters (e.g. X), while individual elements of a tensor are denoted with X_{ijk} , where "ijk" represents indices corresponding to those most appropriate for the tensor. Sub-matrices of the third rank tensors (i.e. second rank tensors, also referred to as matrices) are denoted, for example, as $X_{::k}$, which indicates that, in this example, only the third index of the third rank tensor X is specified. For sub-matrices, sub-vectors (i.e. first rank tensors derived from the corresponding rank tensors, also referred to as vectors) are similarly denoted as, for example, $X_{:jk}$, indicating that e.g. only the second and third index of the third rank tensor X is specified.

[00112] Source localization refers to determining a DOA of an acoustic signal generated by an acoustic source k of K acoustic sources 104-1 through 104-K, the DOA indicating a DOA of the acoustic signal at a microphone array 106 comprising M microphones. Each of K and M could be an integer equal to or greater than 2. M is typically an integer on the order of 5, but, of course, in various implementations the

value of integer M may be different. K is typically an integer in the range $[2,4]$. Since in a typical deployment scenario it is often not possible to know for sure how many acoustic sources are present, value of K (i.e. the number of acoustic sources being modeled) is estimated/selected based on various considerations that a person of ordinary skill in the art would readily recognize, such as e.g. likely number of acoustic sources, an estimate based on a source-counting algorithm, or prior knowledge.

[00113] In an embodiment, a source localization method may include steps of:

a) determining a time-frequency (TF) tensor (X) of $F \times T \times M$ dimensions, where F is an integer indicating the number of frequency components f and T is an integer indicating the number of time frames t (each of F , T , and M being an integer equal to or greater than 2, where F may be on the order of 500 and T may be on the order of 100), the TF tensor comprising a TF representation, e.g. STFT, of each of M digitized signal streams x , each stream corresponding to a combined acoustic signal captured by one of M microphones of the microphone array (the term “combined” indicating that the captured acoustic signal may include contributions from any combination of one or more of the K acoustic sources), where each element X_{ftm} of the tensor X , f being an integer from a set $\{1, \dots, F\}$, t being an integer from a set $\{1, \dots, T\}$, and m being an integer from a set $\{1, \dots, M\}$, is configured to comprise a complex value indicative of measured magnitude and phase of a portion of a digitized stream x corresponding to a frequency component f at a time frame t for a microphone m ;

b) initializing a DOA tensor (Θ), the DOA tensor being of dimensions $3 \times K$ (i.e. it is a second order tensor, or a matrix) and comprising estimated DOA information for each of the K acoustic sources, where each element $\Theta_{i:k}$ of the DOA tensor (i being an integer from a set $\{1, 2, 3\}$, k being an integer from a set $\{1, \dots, K\}$) is configured to comprise a real value indicative of orientation of a particular acoustic source k with respect to the microphone array (in a 3-dimensional space around the microphone array 106) in dimension i (the columns $\Theta_{:k}$ of Θ are vectors of length 1);

c) computing (equation (26) above) a correlation tensor (R) based on values of the TF tensor, the correlation tensor being of dimensions $M \times M \times F$ and comprising information indicative of correlation of the combined acoustic signals captured by different

microphones of the microphone array, where each element $R_{m_1 m_2 f}$ of the correlation tensor (m_1 and m_2 each being integers from a set $\{1, \dots, M\}$ and f being an integer from a set $\{1, \dots, F\}$) is configured to comprise a complex value indicative of estimated correlation between a portion of the digitized stream x as acquired by microphone m_1 (m_1 being an integer from a set $\{1, \dots, M\}$) and a portion of the digitized stream x as acquired by microphone m_2 (m_2 being an integer from a set $\{1, \dots, M\}$) for a particular frequency component f (f being an integer from a set $\{1, \dots, F\}$);

d) computing (equation (29) above) a steering tensor (A) based on values of the DOA tensor, the steering tensor being of dimensions $M \times K \times F$, where each element A_{mkf} of the steering tensor (m being an integer from a set $\{1, \dots, M\}$, k being an integer from a set $\{1, \dots, K\}$, and f being an integer from a set $\{1, \dots, F\}$) is configured to comprise a complex value indicative of the magnitude and phase response of a microphone m to an acoustic source located at $\Theta_{:,k}$ at a frequency component f ;

e) computing (equation (8) above) a projector tensor (B) based on values of the steering tensor, the projector tensor being of dimensions $M \times M \times F$ and comprising information indicative of which one or more portions of the TF tensor determined in step a) originate from localizable sources (i.e. sources for which it is possible to determine orientation with respect to the microphone array; in other words - directional sources; in other words – sources that may be approximated as point sources for which it is possible to identify their location; e.g. ambient noise coming from all different directions would not be associated with a localizable source because it's not possible to identify or estimate a single direction of arrival of that sound). Each element $B_{m_1 m_2 f}$ of the projector tensor (m_1 and m_2 both being integers from a set $\{1, \dots, M\}$ and f being an integer from a set $\{1, \dots, F\}$) is configured to comprise a complex value indicative of a set (subspace) of data vectors X_{ft} that correspond to signals originating from the estimated orientations in Θ at frequency component f (the product $B_{:,f} * X_{ft}$ results in a vector that approximates the directional components in the signal at time t and frequency f);

f) computing (equation (33) above) a DOA gradient tensor (G) based on values of the steering tensor, values of the projector tensor, and values of the correlation tensor, the DOA gradient tensor being of dimensions $3 \times K$ (i.e. a matrix or a second rank tensor) and

comprising information indicative of a change to the DOA matrix for modifying/improving the estimated DOA information, where each element G_{ik} of the DOA gradient tensor (i being an integer from a set $\{1, 2, 3\}$, k being an integer from a set $\{1, \dots, K\}$) is configured to comprise a real value indicative of an estimated change in the DOA tensor for improving orientation estimates of an acoustic source k (i.e. an estimated change in the DOA matrix Θ that is necessary to improve the source orientation estimates);

g) updating (i.e. re-computing the values of) the DOA tensor based on values of the DOA gradient tensor;

h) iterating steps d)-g) two or more times; and

i) following the iterations, determining the DOA of an acoustic source k based on a column $\Theta_{:,k}$ of the DOA tensor (i.e. a DOA vector for any source k is then obtained from the column $\Theta_{:,k}$ of the DOA matrix).

[00114] In one further embodiment, the source localization method summarized above could further include steps e') and e'') to be iterated together with steps d)-g), steps e') and e'') being as follows:

e') computing a TF weight tensor (W) based on values of the projector tensor B and TF tensor X , the weight tensor being of dimensions $F \times T \times K$, where each element W_{ftk} of the weight tensor is configured to comprise a real value between 0 and 1 indicative of the degree to which acoustic source k is active in the $(f,t)^{th}$ bins of the TF tensor X (i.e. indicating a percentage of energy in the $(f,t)^{th}$ bin for each of M microphones that is attributable to the acoustic signal generated by the acoustic source k), and

e'') re-computing (equation (20)) the correlation tensor R based on values of the TF tensor X and the TF weight tensor W .

[00115] The summary provided above is applicable to both the MultSRP and MoSRP approaches described herein. These approaches begin to differ in how the TF weight tensor is computed in step e'). In the MultSRP method, the TF weight tensor is computed using equation (19), while, in the MoSRP method, the weight tensor is computed using both equations (16) and (19).

[00116] In various embodiments, iterations of steps summarized above may be performed until one or more predefined, or dynamically defined, criteria are met. In an

embodiment, the one or more predefined criteria may include a predefined threshold value indicating improvement, e.g. percentage improvement, of a likelihood value indicating how well the estimated orientations in Θ explain the observed data given the assumed data model (see equation (9)).

Examples

[00117] Example 1 provides a method for determining a direction of arrival (DOA) of an acoustic signal generated by an acoustic source k of K acoustic sources, the DOA indicating a DOA of the acoustic signal at a microphone array including M microphones, each of K and M being an integer equal to or greater than 2, the method including: a) determining a time-frequency (TF) tensor of $F \times T \times M$ dimensions, where F is an integer indicating a number of frequency components f and T is an integer indicating a number of time frames t , the TF tensor including a TF representation of each of M digitized signal streams x , each digitized stream corresponding to a combined acoustic signal captured by one of M microphones of the microphone array; b) initializing a DOA matrix of dimensions $3 \times K$, the DOA matrix including estimated DOA information for each of the K acoustic sources; c) based on values of the TF tensor, computing a correlation tensor of dimensions $M \times M \times F$, the correlation tensor including information indicative of correlation of the combined acoustic signals captured by different microphones of the microphone array; d) based on values of the DOA matrix, computing a steering tensor of dimensions $M \times K \times F$, the steering tensor including information indicative of phase and magnitude response of each microphone of the microphone array to each acoustic source of the K acoustic sources; e) based on values of the steering tensor, computing a projector tensor of dimensions $M \times M \times F$, the projector tensor including information indicative of which one or more portions of the TF tensor determined in step a) originate from localizable sources; f) based on values of the steering tensor, values of the projector tensor, and values of the correlation tensor, computing a DOA gradient matrix of dimensions $3 \times K$, the DOA gradient matrix including information indicative of a change to the DOA matrix for modifying the estimated DOA information; g) updating the DOA matrix based on values of the DOA gradient matrix; h) iterating steps d)-g) two or more times; and i) following

the iterations, determining the DOA of an acoustic source k based on a column $\Theta_{:,k}$ of the DOA matrix.

[00118] Example 2 provides the method according to Example 1, where each element X_{ftm} of the TF tensor is configured to include a complex value indicative of measured magnitude and phase of a portion of a digitized stream x corresponding to a frequency component f at a time frame t for a microphone m .

[00119] Example 3 provides the method according to Examples 1 or 2, where each element Θ_{ik} of the DOA matrix is configured to include a real value indicative of orientation of the acoustic source k with respect to the microphone array in dimension i .

[00120] Example 4 provides the method according to any one of the preceding Examples, where each element $R_{m_1m_2f}$ of the correlation tensor is configured to include a complex value indicative of correlation between a portion of the digitized stream x as acquired by microphone m_1 and a portion of the digitized stream x as acquired by microphone m_2 for a particular frequency component f .

[00121] Example 5 provides the method according to any one of the preceding Examples, where each element A_{mkf} of the steering tensor is configured to include a complex value indicative of a magnitude and a phase response of a microphone m to an acoustic source k at a frequency component f .

[00122] Example 6 provides the method according to any one of the preceding Examples, where each element $B_{m_1m_2f}$ of the projector tensor is configured to include a complex value indicative of a set of data vectors X_{ft} that correspond to localizable signals with steering matrix $A_{:,f}$ at a frequency component f .

[00123] Example 7 provides the method according to any one of the preceding Examples, where each element G_{ik} of the DOA gradient matrix is configured to include a real value indicative of an estimated change in the DOA tensor for improving orientation estimate of the acoustic source k .

[00124] Example 8 provides the method according to any one of the preceding Examples, further including: e') based on values of the projector tensor and values of the TF tensor, computing a TF weight tensor of dimensions $F \times T \times K$, where each element W_{ftk} of the TF weight tensor is configured to include a real value between 0 and 1 indicative of

a degree to which the acoustic source k is active in the $(f,t)^{\text{th}}$ bin, and e'') re-computing the correlation tensor based on the values of the TF tensor and values of the TF weight tensor, where the iterations include iterating steps d-g, e' , and e'' .

[00125] Example 9 provides the method according to Example 8, where computing the TF weight tensor includes using a Wiener mask.

[00126] Example 10 provides the method according to Example 8, where computing the TF weight tensor includes using a Wiener mask and defining source-specific correlation matrices in terms of posterior probabilities using a Wiener mask.

[00127] Example 11 provides the method according to any one of the preceding Examples, where the iterations are performed until one or more predefined criteria are met.

[00128] Example 12 provides a method for identifying a first direction of arrival of sound waves from a first acoustic source and a second direction of arrival of sound waves from a second acoustic source, the method including: receiving, at a microphone array, acoustic signals including the sound waves from the first and second acoustic sources; converting the received acoustic signals from a time domain to a time-frequency domain; processing the converted acoustic signals to determine an estimated first angle representing the first direction of arrival and an estimated second angle representing the second direction of arrival; and updating the estimated first and second angles; where processing includes localizing, separating and Wiener post-filtering the converted acoustic signals using time-frequency weighting and outputting a time-frequency weighted signal for estimating the first and second angles.

[00129] Example 13 provides the method according to Example 12, further including combining the time-frequency weighted signal with the converted acoustic signals to generate a correlation matrix.

[00130] Example 14 provides the method according to Example 13, where updating the estimated first and second angles includes utilizing the correlation matrix and the estimated first and second angles and outputting updated estimated first and second angles.

[00131] Example 15 provides the method according to Example 12, where converting the received acoustic signals from a time domain to a time-frequency domain includes using a short time Fourier transform.

[00132] Example 16 provides the method according to Example 12, where processing the converted acoustic signals to determine the estimated first and second angles includes decomposing the converted acoustic signals to identify signals from each of the first and second acoustic sources by accounting for interference between the first and second acoustic sources in forming the acoustic signals.

[00133] Example 17 provides the method according to Example 12, where processing the converted acoustic signals and updating the first and second estimated angles includes iteratively decomposing the converted acoustic signals to simultaneously determine the first and second directions of arrival.

[00134] Example 18 provides the method according to Example 12, where processing the converted acoustic signals includes processing using steered response power localization.

[00135] Example 19 provides the method according to Example 12, further including using an inverse STFT to convert the processed converted acoustic signals back into the time domain and separating the sound waves from the first acoustic source from the sound waves from the second acoustic source.

[00136] Example 20 provides a system comprising means for implementing the method according to any one of the preceding Examples.

[00137] Example 21 provides a data structure for assisting implementation of the method according to any one of the preceding Examples.

[00138] Example 22 provides a system for determining a DOA of an acoustic signal generated by an acoustic source k of K acoustic sources, the DOA indicating a DOA of the acoustic signal at a microphone array comprising M microphones, each of K and M being an integer equal to or greater than 2, the system including at least one memory element configured to store computer executable instructions, and at least one processor coupled to the at least one memory element and configured, when executing the instructions, to carry out the method according to any one of Examples 1-11.

[00139] Example 23 provides one or more non-transitory tangible media encoding logic that include instructions for execution that, when executed by a processor, are operable to perform operations for determining a DOA of an acoustic signal generated by an acoustic source k of K acoustic sources, the DOA indicating a DOA of the acoustic signal at a microphone array comprising M microphones, each of K and M being an integer equal to or greater than 2, the operations comprising operations of the method according to any one of Examples 1-11.

[00140] Example 24 provides a system for identifying a first direction of arrival of sound waves from a first acoustic source and a second direction of arrival of sound waves from a second acoustic source, the system including at least one memory element configured to store computer executable instructions, and at least one processor coupled to the at least one memory element and configured, when executing the instructions, to carry out the method according to any one of Examples 12-19.

[00141] Example 25 provides one or more non-transitory tangible media encoding logic that include instructions for execution that, when executed by a processor, are operable to perform operations for identifying a first direction of arrival of sound waves from a first acoustic source and a second direction of arrival of sound waves from a second acoustic source, the operations comprising operations of the method according to any one of Examples 12-19.

Variations and implementations

[00142] In the discussions of the embodiments above, components can readily be replaced, substituted, or otherwise modified in order to accommodate particular circuitry needs. Moreover, it should be noted that the use of complementary electronic devices, hardware, software, etc. offer an equally viable option for implementing the teachings of the present disclosure.

[00143] In one example embodiment, any number of electrical circuits used to implement the systems and methods of the FIGURES may be implemented on a board of an associated electronic device. The board can be a general circuit board that can hold various components of the internal electronic system of the electronic device and,

further, provide connectors for other peripherals. More specifically, the board can provide the electrical connections by which the other components of the system can communicate electrically. Any suitable processors (inclusive of digital signal processors, microprocessors, supporting chipsets, etc.), computer-readable non-transitory memory elements, etc. can be suitably coupled to the board based on particular configuration needs, processing demands, computer designs, etc. Other components such as external storage, additional sensors, controllers for audio/video display, and peripheral devices may be attached to the board as plug-in cards, via cables, or integrated into the board itself. In various embodiments, the functionalities described herein may be implemented in emulation form as software or firmware running within one or more configurable (e.g., programmable) elements arranged in a structure that supports these functions. The software or firmware providing the emulation may be provided on non-transitory computer-readable storage medium comprising instructions to allow a processor to carry out those functionalities.

[00144] In another example embodiment, the systems and methods of the FIGURES may be implemented as stand-alone modules (e.g., a device with associated components and circuitry configured to perform a specific application or function) or implemented as plug-in modules into application specific hardware of electronic devices. Note that particular embodiments of the present disclosure may be readily included in a system on chip (SOC) package, either in part, or in whole. An SOC represents an IC that integrates components of a computer or other electronic system into a single chip. It may contain digital, analog, mixed-signal, and often radio frequency functions: all of which may be provided on a single chip substrate. Other embodiments may include a multi-chip-module (MCM), with a plurality of separate ICs located within a single electronic package and configured to interact closely with each other through the electronic package. In various other embodiments, the identification, localization and separation functionalities may be implemented in one or more silicon cores in Application Specific Integrated Circuits (ASICs), Field Programmable Gate Arrays (FPGAs), and other semiconductor chips.

[00145] It is also imperative to note that all of the specifications, dimensions, and relationships outlined herein (e.g., the number of processors, logic operations, etc.) have only been offered for purposes of example and teaching only. Such information may be varied considerably without departing from the spirit of the present disclosure, or the scope of the appended claims or examples. The specifications apply only to one non-limiting example and, accordingly, they should be construed as such. In the foregoing description, example embodiments have been described with reference to particular processor and/or component arrangements. Various modifications and changes may be made to such embodiments without departing from the scope of the appended claims or examples. The description and drawings are, accordingly, to be regarded in an illustrative rather than in a restrictive sense.

[00146] Note that the activities discussed above with reference to the FIGURES are applicable to any integrated circuits that involve signal processing, particularly those that can execute specialized software programs, or algorithms, some of which may be associated with processing digitized real-time data. Certain embodiments can relate to multi-DSP signal processing, floating point processing, signal/control processing, fixed-function processing, microcontroller applications, etc.

[00147] In certain contexts, the features discussed herein can be applicable to medical systems, scientific instrumentation, wireless and wired communications, radar, industrial process control, audio and video equipment, current sensing, instrumentation (which can be highly precise), and other digital-processing-based systems.

[00148] Moreover, certain embodiments discussed above can be provisioned in digital signal processing technologies for medical imaging, patient monitoring, medical instrumentation, and home healthcare. This could include pulmonary monitors, accelerometers, heart rate monitors, pacemakers, etc. Other applications can involve automotive technologies for safety systems (e.g., stability control systems, driver assistance systems, braking systems, infotainment and interior applications of any kind). Furthermore, powertrain systems (for example, in hybrid and electric vehicles) can use high-precision data conversion products in battery monitoring, control systems, reporting controls, maintenance activities, etc.

[00149] In yet other example scenarios, the teachings of the present disclosure can be applicable in the industrial markets that include process control systems that help drive productivity, energy efficiency, and reliability. In consumer applications, the teachings of the signal processing circuits discussed above can be used for image processing, auto focus, and image stabilization (e.g., for digital still cameras, camcorders, etc.). Other consumer applications can include audio and video processors for home theater systems, DVD recorders, and high-definition televisions. Yet other consumer applications can involve advanced touch screen controllers (e.g., for any type of portable media device). Hence, such technologies could readily part of smartphones, tablets, security systems, PCs, gaming technologies, virtual reality, simulation training, etc.

[00150] Note that with the numerous examples provided herein, interaction may be described in terms of two, three, four, or more electrical components. However, this has been done for purposes of clarity and example only. It should be appreciated that the system can be consolidated in any suitable manner. Along similar design alternatives, any of the illustrated components, modules, and elements of the FIGURES may be combined in various possible configurations, all of which are clearly within the broad scope of this Specification. In certain cases, it may be easier to describe one or more of the functionalities of a given set of flows by only referencing a limited number of electrical elements. It should be appreciated that the electrical circuits of the FIGURES and its teachings are readily scalable and can accommodate a large number of components, as well as more complicated/sophisticated arrangements and configurations. Accordingly, the examples provided should not limit the scope or inhibit the broad teachings of the electrical circuits as potentially applied to a myriad of other architectures.

[00151] Note that in this Specification, references to various features (e.g., elements, structures, modules, components, steps, operations, characteristics, etc.) included in “one embodiment”, “example embodiment”, “an embodiment”, “another embodiment”, “some embodiments”, “various embodiments”, “other embodiments”, “alternative embodiment”, and the like are intended to mean that any such features are

included in one or more embodiments of the present disclosure, but may or may not necessarily be combined in the same embodiments.

[00152] It is also important to note that the functions related to acoustic source localization and separation, illustrate only some of the possible localization and separation functions that may be executed by, or within, systems illustrated in the FIGURES. Some of these operations may be deleted or removed where appropriate, or these operations may be modified or changed considerably without departing from the scope of the present disclosure. In addition, the timing of these operations may be altered considerably. The preceding operational flows have been offered for purposes of example and discussion. Substantial flexibility is provided by embodiments described herein in that any suitable arrangements, chronologies, configurations, and timing mechanisms may be provided without departing from the teachings of the present disclosure.

[00153] Numerous other changes, substitutions, variations, alterations, and modifications may be ascertained to one skilled in the art and it is intended that the present disclosure encompass all such changes, substitutions, variations, alterations, and modifications as falling within the scope of the appended claims.

[00154] Although the claims are presented in single dependency format in the style used before the USPTO, it should be understood that any claim can depend on and be combined with any preceding claim of the same type unless that is clearly technically infeasible.

OTHER NOTES, EXAMPLES, AND IMPLEMENTATIONS

[00155] Note that all optional features of the apparatus described above may also be implemented with respect to the method or process described herein and specifics in the examples may be used anywhere in one or more embodiments.

[00156] In a first example, a system is provided (that can include any suitable circuitry, dividers, capacitors, resistors, inductors, ADCs, DFFs, logic gates, software, hardware, links, etc.) that can be part of any type of computer, which can further include

a circuit board coupled to a plurality of electronic components. The system can include means for clocking data from the digital core onto a first data output of a macro using a first clock, the first clock being a macro clock; means for clocking the data from the first data output of the macro into the physical interface using a second clock, the second clock being a physical interface clock; means for clocking a first reset signal from the digital core onto a reset output of the macro using the macro clock, the first reset signal output used as a second reset signal; means for sampling the second reset signal using a third clock, which provides a clock rate greater than the rate of the second clock, to generate a sampled reset signal; and means for resetting the second clock to a predetermined state in the physical interface in response to a transition of the sampled reset signal.

[00157] The 'means for' in these instances (above) can include (but is not limited to) using any suitable component discussed herein, along with any suitable software, circuitry, hub, computer code, logic, algorithms, hardware, controller, interface, link, bus, communication pathway, etc. In a second example, the system includes memory that further comprises machine-readable instructions that when executed cause the system to perform any of the activities discussed above.

WHAT IS CLAIMED IS:

1. A method for determining a direction of arrival (DOA) of an acoustic signal generated by an acoustic source k of K acoustic sources, the DOA indicating a DOA of the acoustic signal at a microphone array comprising M microphones, each of K and M being an integer equal to or greater than 2, the method comprising:

a) determining a time-frequency (TF) tensor of $F \times T \times M$ dimensions, where F is an integer indicating a number of frequency components f and T is an integer indicating a number of time frames t , the TF tensor comprising a TF representation of each of M digitized signal streams x , each digitized stream corresponding to a combined acoustic signal captured by one of M microphones of the microphone array;

b) initializing a DOA matrix of dimensions $3 \times K$, the DOA matrix comprising estimated DOA information for each of the K acoustic sources;

c) based on values of the TF tensor, computing a correlation tensor of dimensions $M \times M \times F$, the correlation tensor comprising information indicative of correlation of the combined acoustic signals captured by different microphones of the microphone array;

d) based on values of the DOA matrix, computing a steering tensor of dimensions $M \times K \times F$, the steering tensor comprising information indicative of phase and magnitude response of each microphone of the microphone array to each acoustic source of the K acoustic sources;

e) based on values of the steering tensor, computing a projector tensor of dimensions $M \times M \times F$, the projector tensor comprising information indicative of which one or more portions of the TF tensor determined in step a) originate from localizable sources;

f) based on values of the steering tensor, values of the projector tensor, and values of the correlation tensor, computing a DOA gradient matrix of dimensions $3 \times K$,

the DOA gradient matrix comprising information indicative of a change to the DOA matrix for modifying the estimated DOA information;

g) updating the DOA matrix based on values of the DOA gradient matrix;

h) iterating steps d)-g) two or more times; and

i) following the iterations, determining the DOA of an acoustic source k based on a column $\Theta_{:k}$ of the DOA matrix.

2. The method according to claim 1, where each element X_{ftm} of the TF tensor is configured to comprise a complex value indicative of measured magnitude and phase of a portion of a digitized stream x corresponding to a frequency component f at a time frame t for a microphone m .
3. The method according to claim 1, where each element Θ_{ik} of the DOA matrix is configured to comprise a real value indicative of orientation of the acoustic source k with respect to the microphone array in dimension i .
4. The method according to claim 1, where each element $R_{m_1m_2f}$ of the correlation tensor is configured to comprise a complex value indicative of correlation between a portion of the digitized stream x as acquired by microphone m_1 and a portion of the digitized stream x as acquired by microphone m_2 for a particular frequency component f .
5. The method according to claim 1, where each element A_{mkf} of the steering tensor is configured to comprise a complex value indicative of a magnitude and a phase response of a microphone m to an acoustic source k at a frequency component f .
6. The method according to claim 1, wherein each element $B_{m_1m_2f}$ of the projector tensor is configured to comprise a complex value indicative of a set of data vectors X_{ft} that correspond to localizable signals with steering matrix $A_{:f}$ at a frequency component f .

7. The method according to claim 1, wherein each element G_{ik} of the DOA gradient matrix is configured to comprise a real value indicative of an estimated change in the DOA tensor for improving orientation estimate of the acoustic source k .
8. The method according to any one of claims 1-7, further comprising:
 - e') based on values of the projector tensor and values of the TF tensor, computing a TF weight tensor of dimensions $F \times T \times K$, where each element W_{ftk} of the TF weight tensor is configured to comprise a real value between 0 and 1 indicative of a degree to which the acoustic source k is active in the $(f,t)^{th}$ bin, and
 - e'') re-computing the correlation tensor based on the values of the TF tensor and values of the TF weight tensor,wherein the iterations comprise iterating steps d-g, e', and e''.
9. The method according to claim 8, wherein computing the TF weight tensor comprises using a Wiener mask.
10. The method according to claim 8, wherein computing the TF weight tensor comprises using a Wiener mask and defining source-specific correlation matrices in terms of posterior probabilities using a Wiener mask.
11. The method according to any one of claims 1-7, wherein the iterations are performed until one or more predefined criteria are met.
12. A method for identifying a first direction of arrival of sound waves from a first acoustic source and a second direction of arrival of sound waves from a second acoustic source, the method comprising:
 - receiving, at a microphone array, acoustic signals including the sound waves from the first and second acoustic sources;
 - converting the received acoustic signals from a time domain to a time-frequency domain;

processing the converted acoustic signals to determine an estimated first angle representing the first direction of arrival and an estimated second angle representing the second direction of arrival; and

updating the estimated first and second angles;

wherein processing includes localizing, separating and Wiener post-filtering the converted acoustic signals using time-frequency weighting and outputting a time-frequency weighted signal for estimating the first and second angles.

13. The method according to claim 12, further comprising combining the time-frequency weighted signal with the converted acoustic signals to generate a correlation matrix.
14. The method according to claim 13, wherein updating the estimated first and second angles comprises utilizing the correlation matrix and the estimated first and second angles and outputting updated estimated first and second angles.
15. The method according to claim 12, wherein converting the received acoustic signals from a time domain to a time-frequency domain includes using a short time Fourier transform.
16. The method according to claim 12, wherein processing the converted acoustic signals to determine the estimated first and second angles includes decomposing the converted acoustic signals to identify signals from each of the first and second acoustic sources by accounting for interference between the first and second acoustic sources in forming the acoustic signals.
17. The method according to claim 12, wherein processing the converted acoustic signals and updating the first and second estimated angles includes iteratively decomposing the converted acoustic signals to simultaneously determine the first and second directions of arrival.

18. The method according to claim 12, wherein processing the converted acoustic signals includes processing using steered response power localization.
19. The method according to claim 12, further comprising using an inverse STFT to convert the processed converted acoustic signals back into the time domain and separating the sound waves from the first acoustic source from the sound waves from the second acoustic source.

100

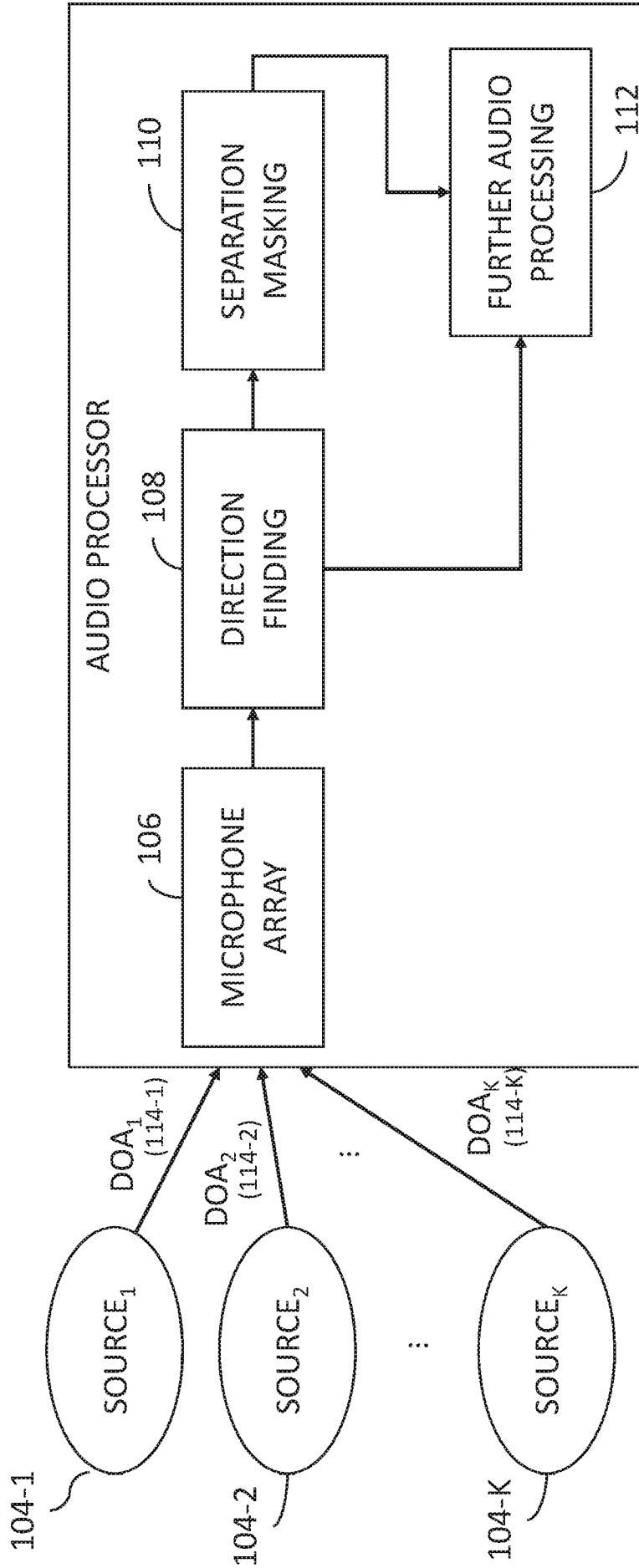


Figure 1

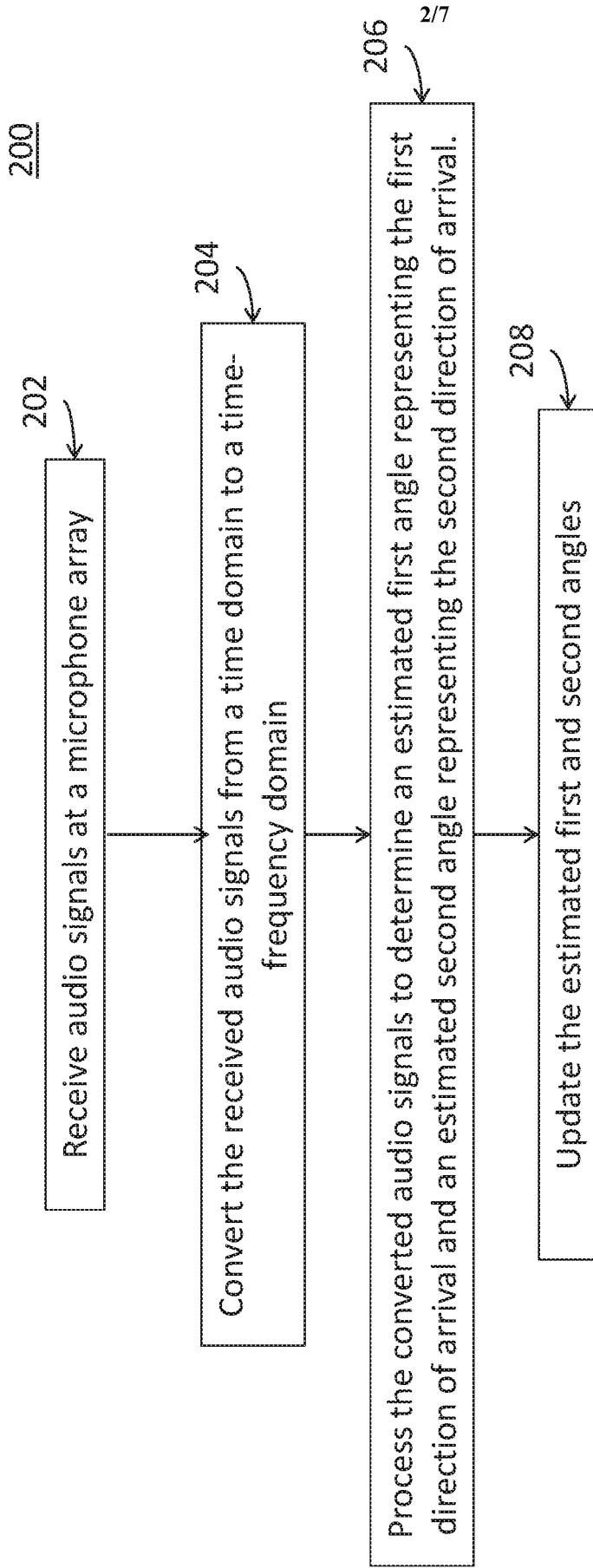


Figure 2

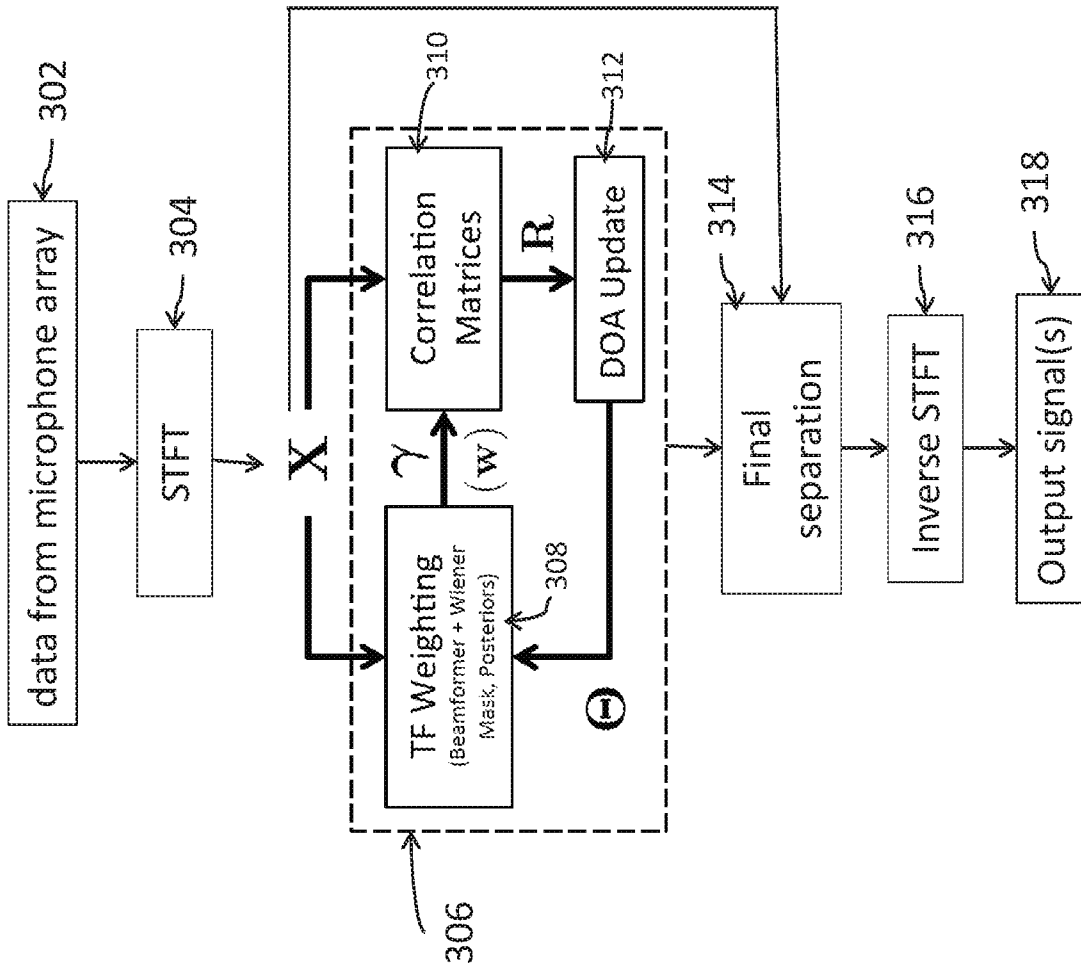


FIGURE 3

400

Additivity of data vectors

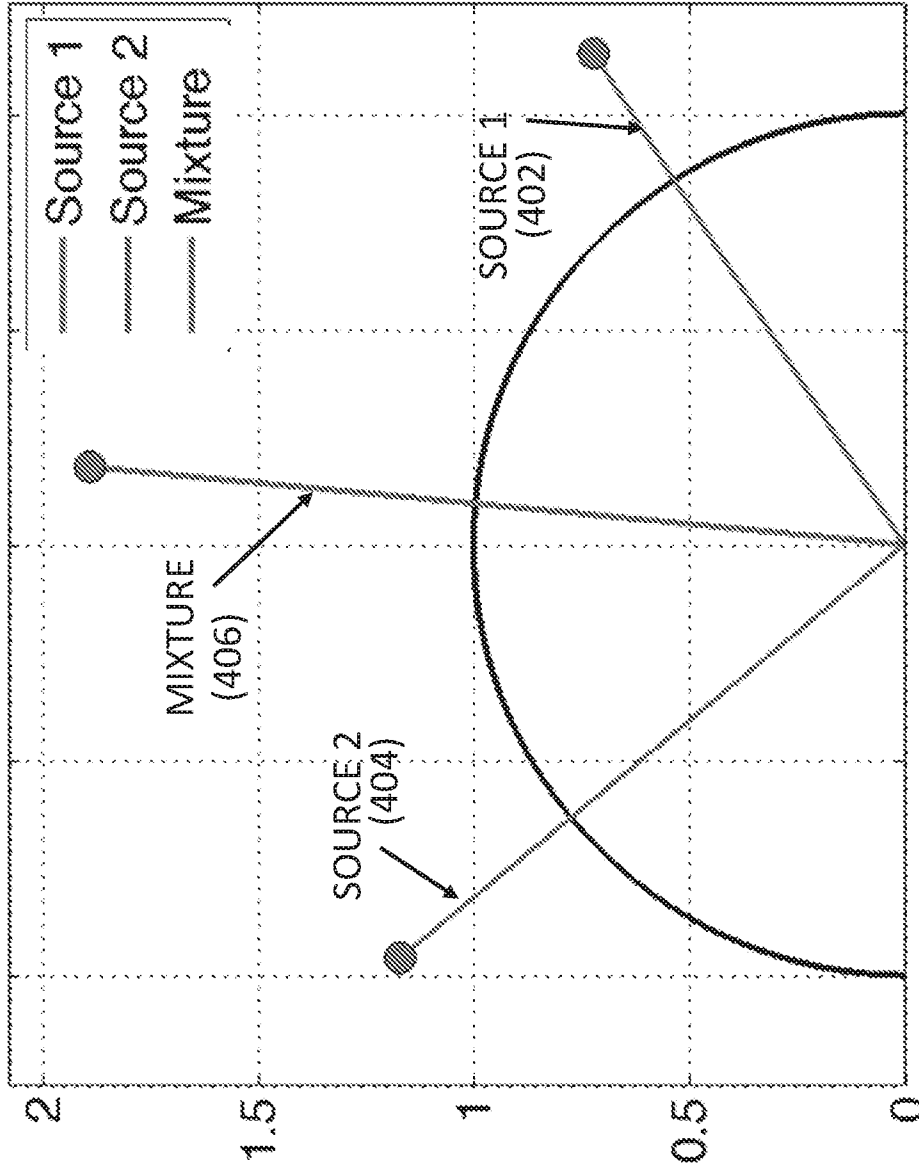


Figure 4

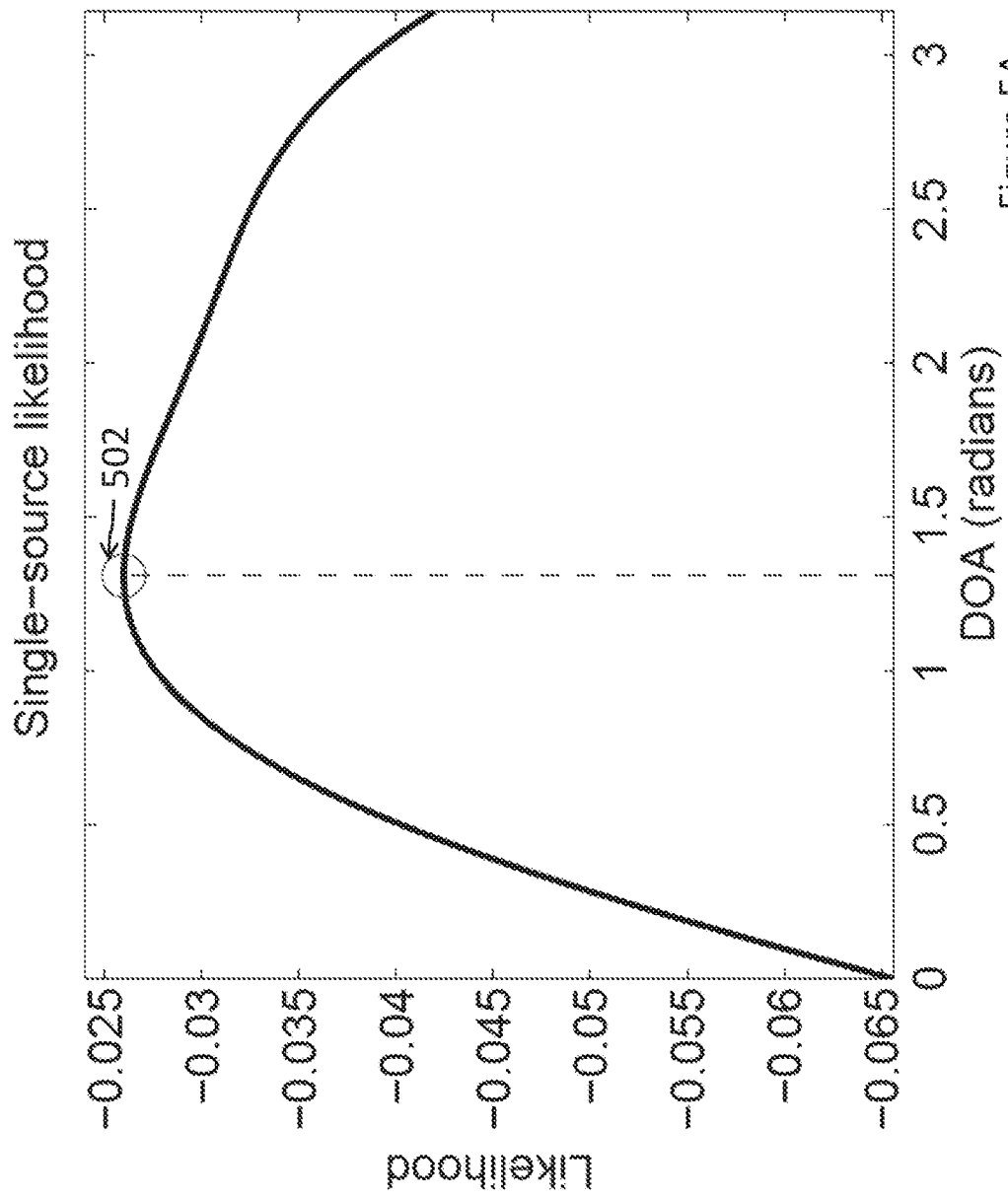


Figure 5A

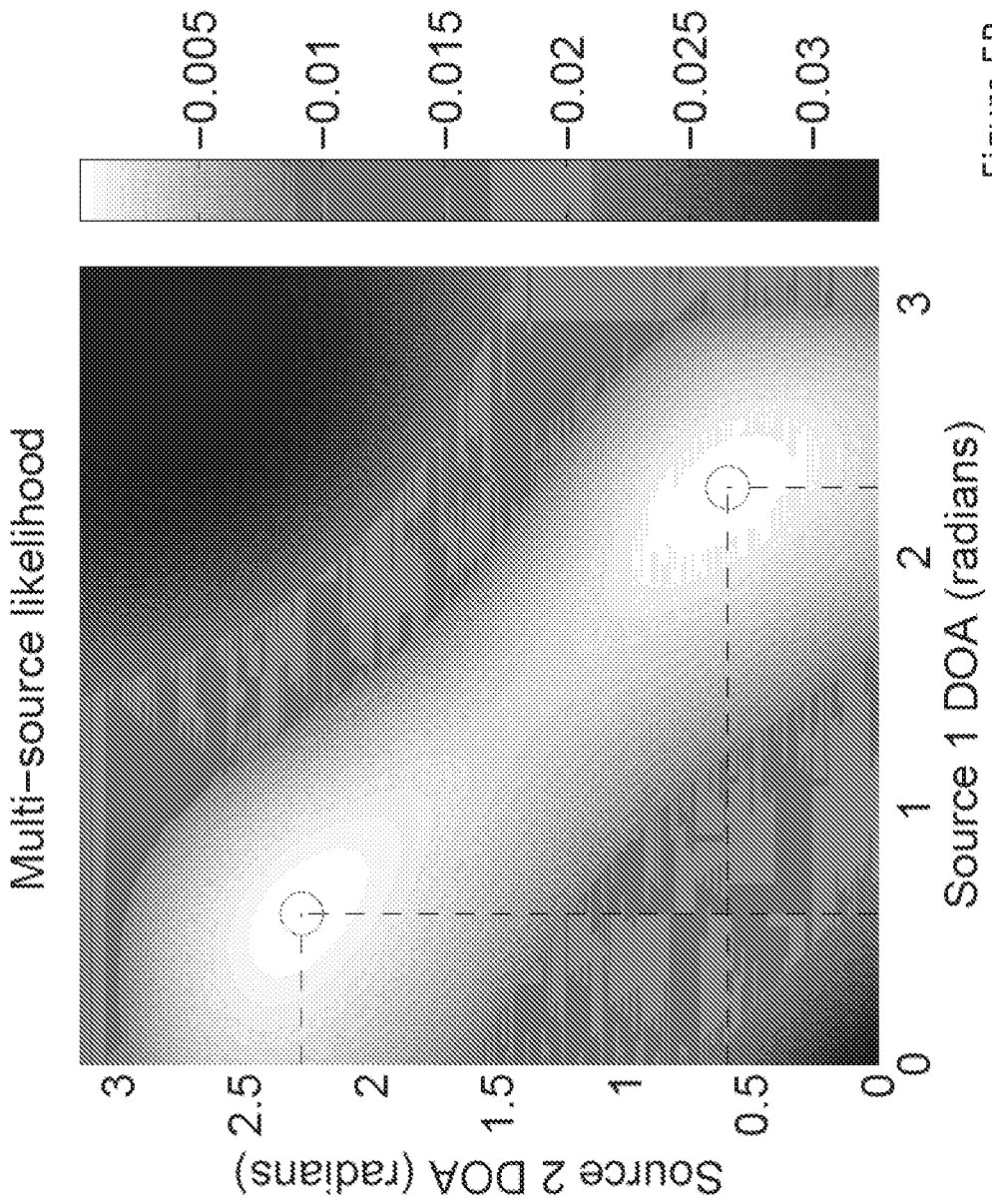


Figure 5B

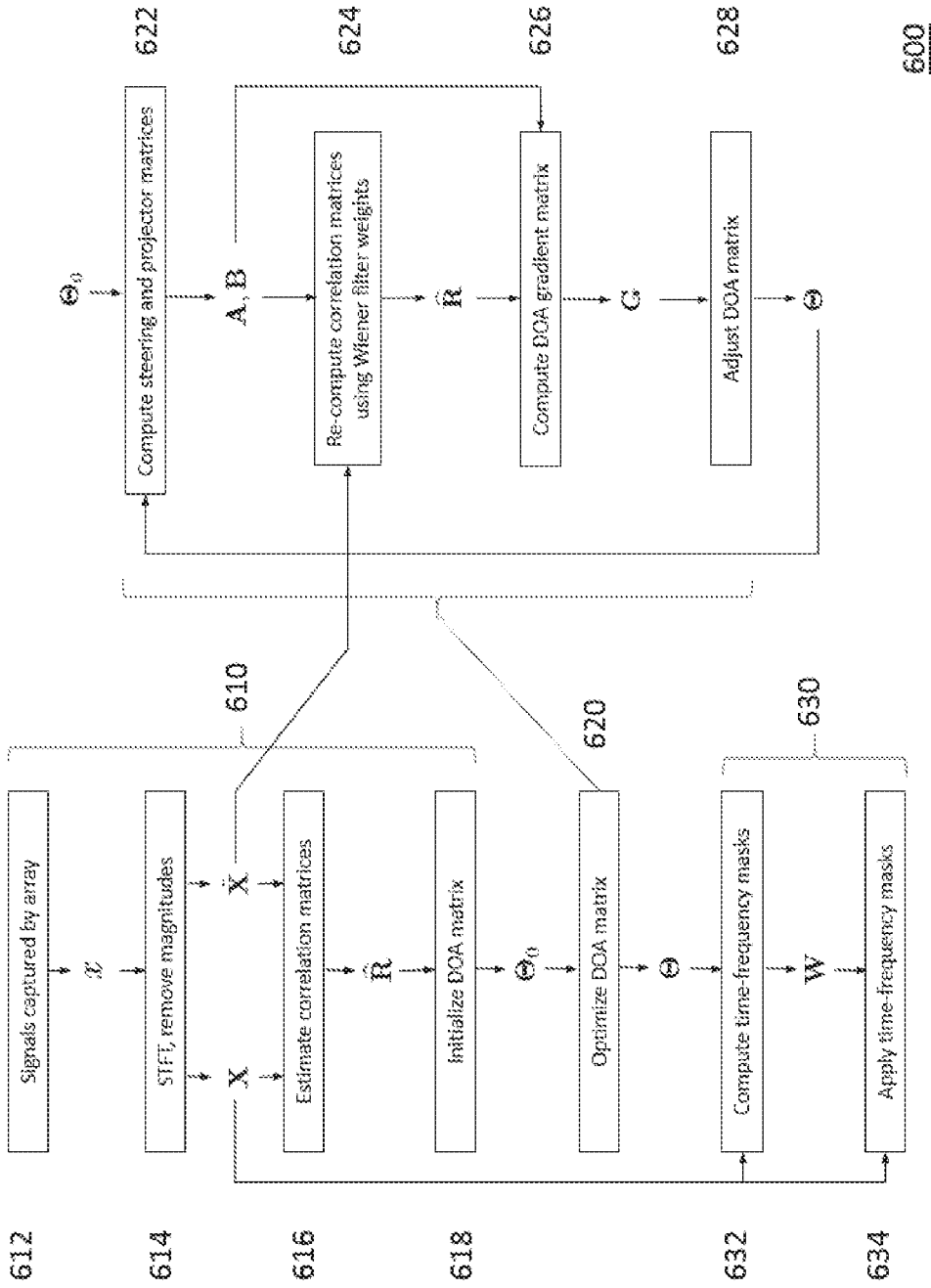


Figure 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2015/066012

A. CLASSIFICATION OF SUBJECT MATTER IPC (2016.01) G01S 3/00		
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 (2016.01) G01S 3/00		
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) Databases consulted: USPTO, THOMSON INNOVATION		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2014192999 A1 ST MICROELECTRONICS SRL [IT] 10 Jul 2014 (2014/07/10) the whole document	1-19
A	US 2014226838 A1 ANALOG DEVICES INC [US] 14 Aug 2014 (2014/08/14) the whole document	1-19
A	US 2014023199 A1 GIESBRECHT DAVID [CA] 23 Jan 2014 (2014/01/23) the whole document	1-19
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> 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 05 Apr 2016		Date of mailing of the international search report 12 Apr 2016
Name and mailing address of the ISA: Israel Patent Office Technology Park, Bldg.5, Malcha, Jerusalem, 9695101, Israel Facsimile No. 972-2-5651616		Authorized officer MOSKOVICH Elad Telephone No. 972-2-5651607

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No. PCT/US2015/066012
--

Patent document cited search report	Publication date	Patent family member(s)	Publication Date
US 2014192999 A1	10 Jul 2014	US 2014192999 A1	10 Jul 2014
		IT VI20130001 A1	09 Jul 2014
US 2014226838 A1	14 Aug 2014	US 2014226838 A1	14 Aug 2014
		CN 104995679 A	21 Oct 2015
		EP 2956938 A1	23 Dec 2015
		KR 20150093801 A	18 Aug 2015
		US 2015086038 A1	26 Mar 2015
		WO 2014127080 A1	21 Aug 2014
		WO 2015048070 A1	02 Apr 2015
		WO 2015100261 A1	02 Jul 2015
		WO 2015157013 A1	15 Oct 2015
US 2014023199 A1	23 Jan 2014	US 2014023199 A1	23 Jan 2014