(12) UK Patent (19) GB (11) 2519315 (13) B

(45) Date of B Publication **16.12.2020**

(54) Title of the Invention: Method and apparatus for identifying actual signal sources among a plurality of signal sources with artefacts detection

(51) INT CL: ***G10L 21/0272*** (2013.01) ***G01S 3/808*** (2006.01)

(21) Application No: **1318302.5**

(22) Date of Filing: **16.10.2013**

(43) Date of A Publication **22.04.2015**

(56) Documents Cited:
WO 2004/012576 A2
WO 1994/026075 A1
US 20140072142 A1
US 20120095753 A1
US 20120069714 A1

(58) Field of Search:
As for published application 2519315 A viz:
INT CL **G01S, G10L, H04R, H04S**
Other: **WPI, EPODOC**
updated as appropriate

Additional Fields
Other: **None**

(72) Inventor(s):
Johann Citerin

(73) Proprietor(s):
Canon Kabushiki Kaisha
(Incorporated in Japan)
3-30-2, Shimomaruko, 3-chome, Ohta-ku, Tokyo,
Japan

(74) Agent and/or Address for Service:
SANTARELLI
49, avenue des Champs-Elysées, Paris 75008,
France (including Overseas Departments and Territories)

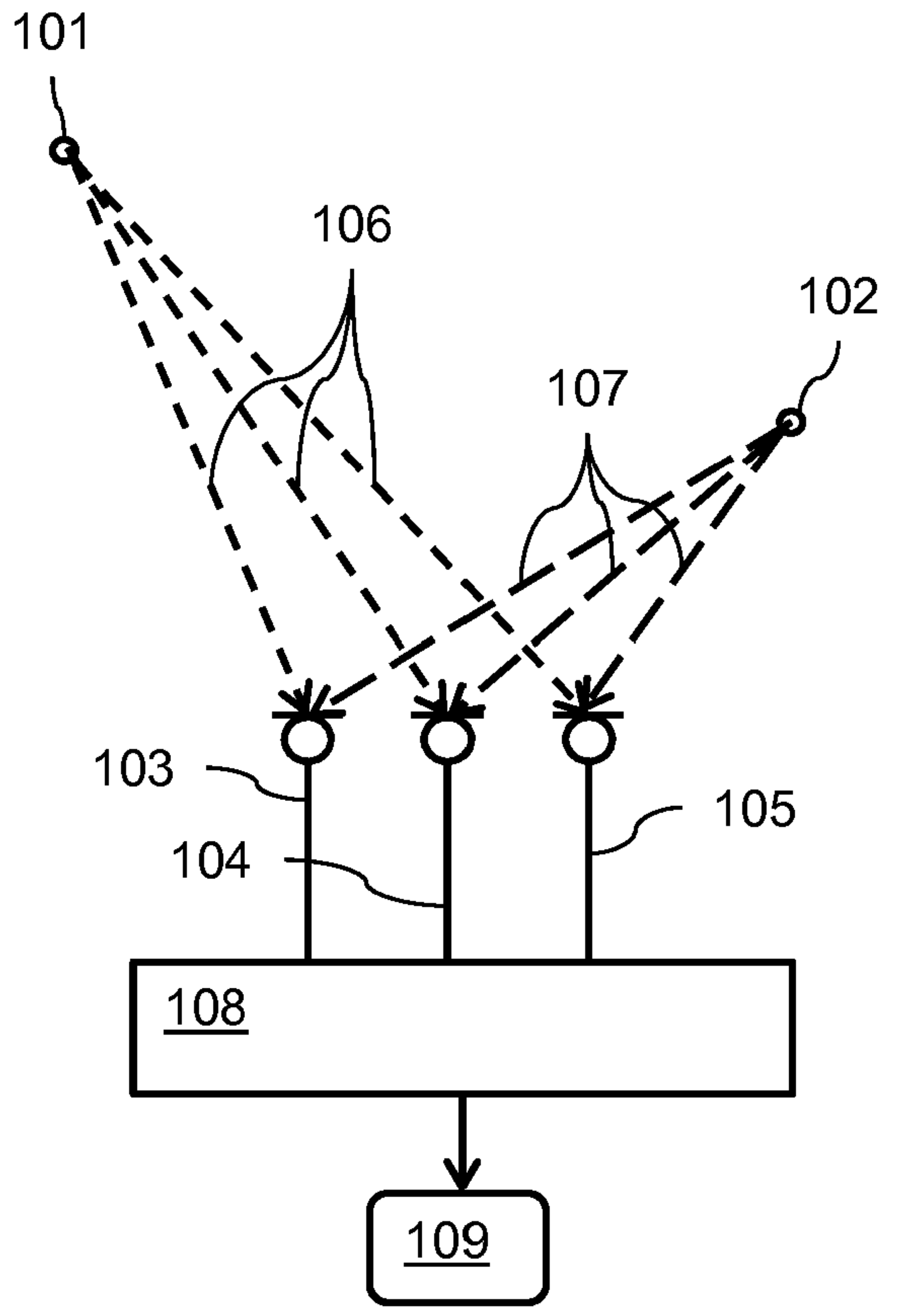
101
106
102
107
103
104
105
108
109

Fig. 1

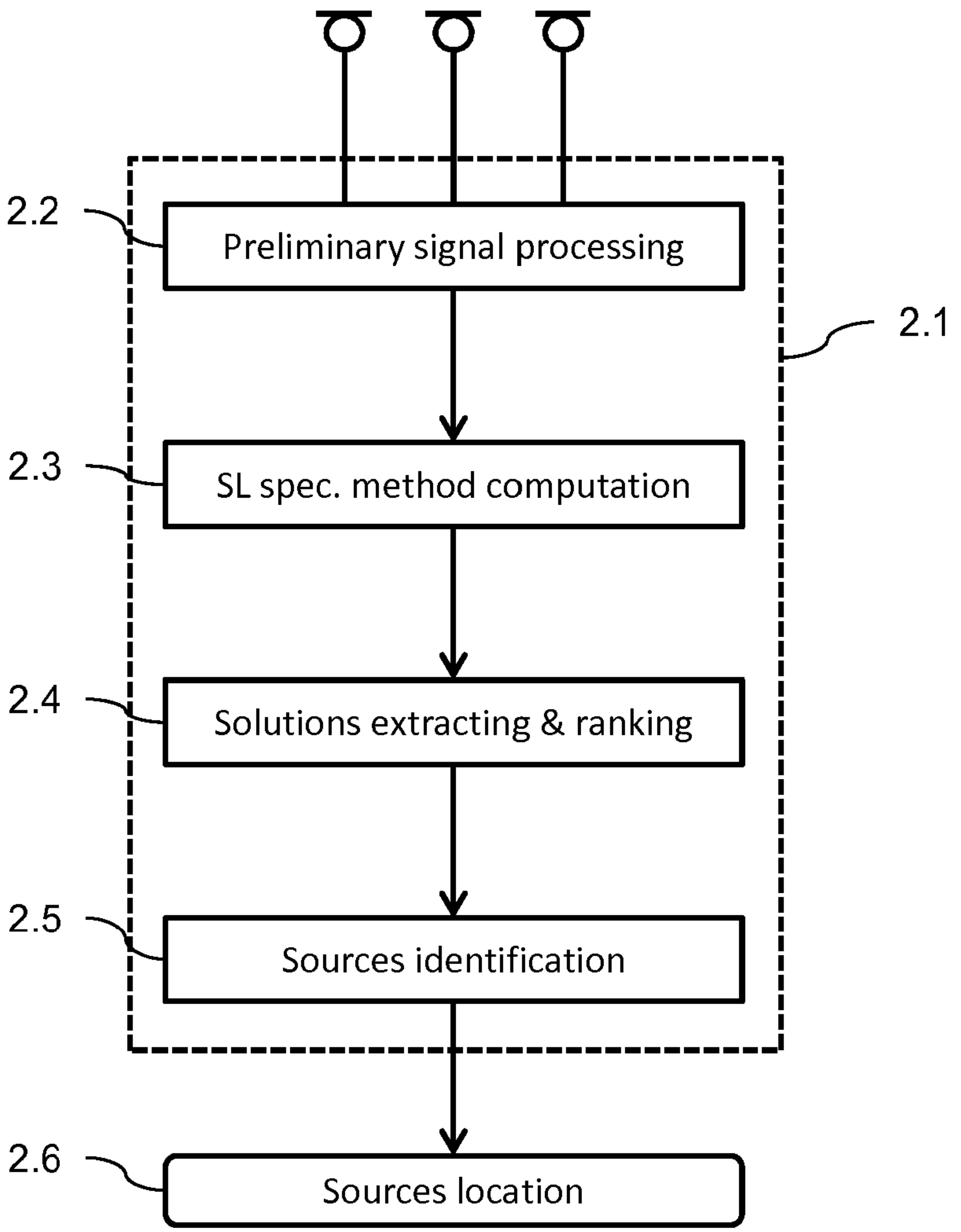
2.2
Preliminary signal processing
2.1
2.3
SL spec. method computation
2.4
Solutions extracting & ranking
2.5
Sources identification
2.6
Sources location

Fig. 2

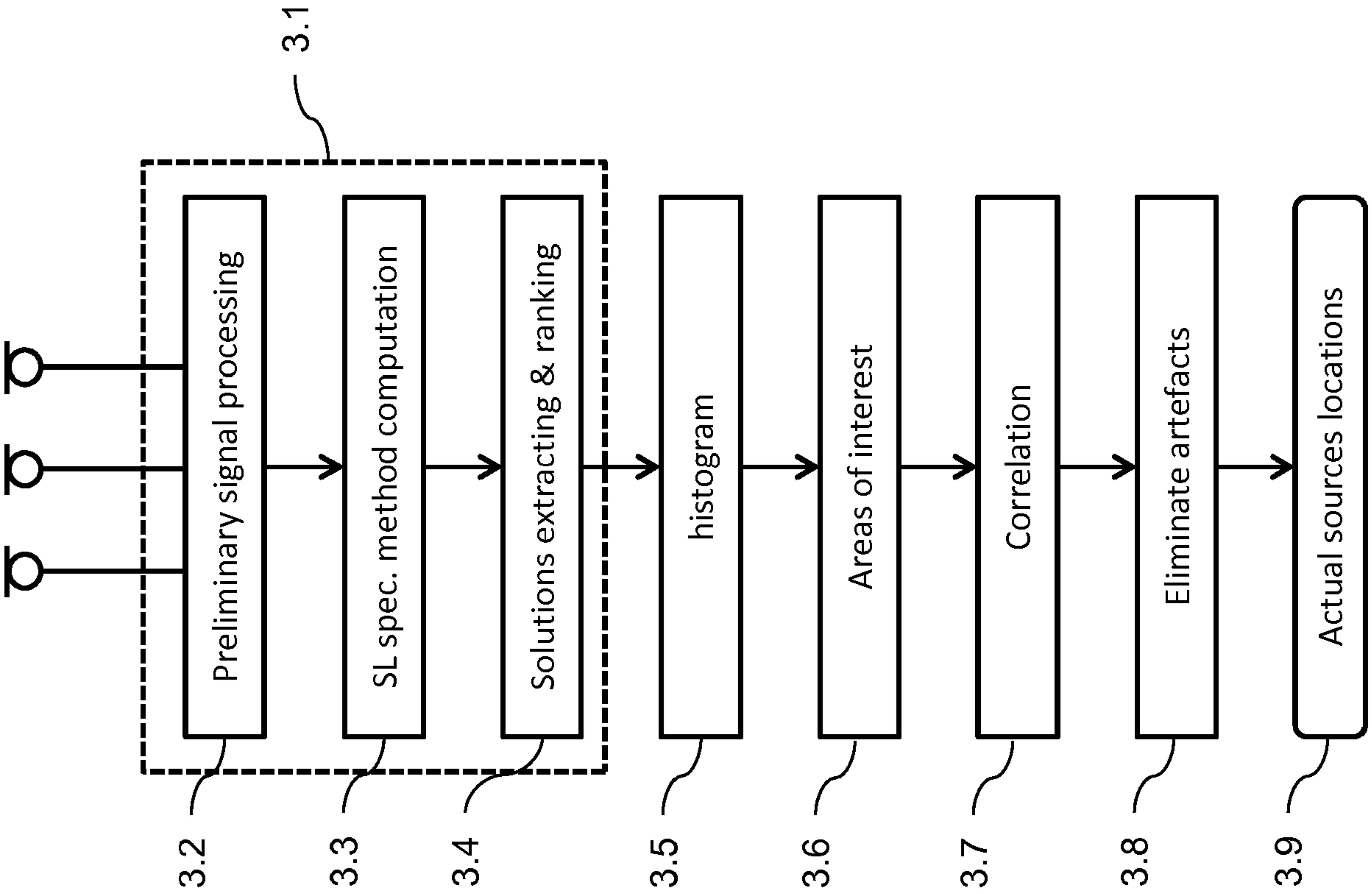
3.1
Preliminary signal processing
SL spec. method computation
Solutions extracting & ranking
histogram
Areas of interest
Correlation
Eliminate artefacts
Actual sources locations
3.2
3.3
3.4
3.5
3.6
3.7
3.8
3.9

Fig. 3

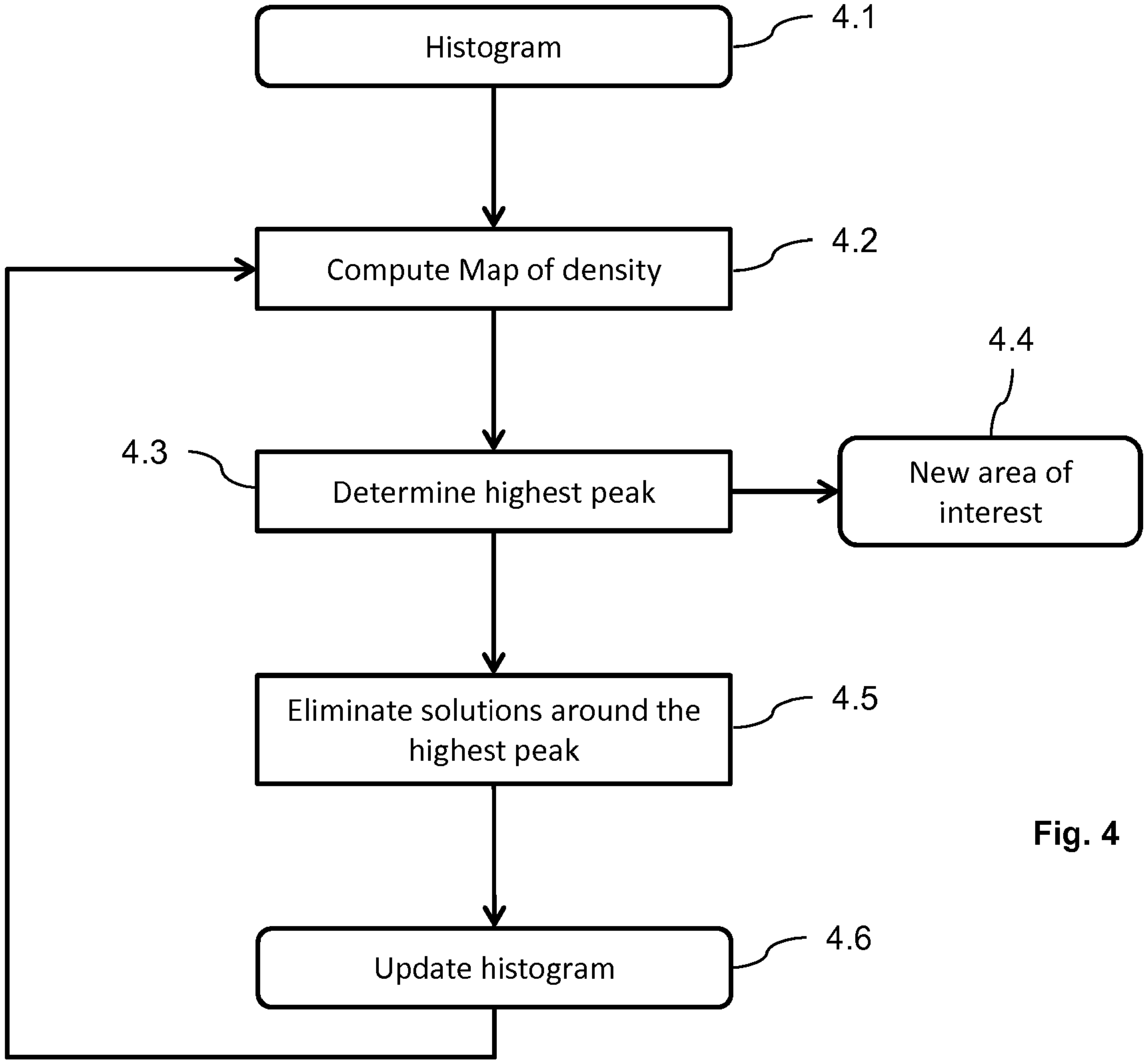
Histogram
4.1
Compute Map of density
4.2
4.3
Determine highest peak
New area of interest
4.4
Eliminate solutions around the highest peak
4.5
Update histogram
4.6

Fig. 4

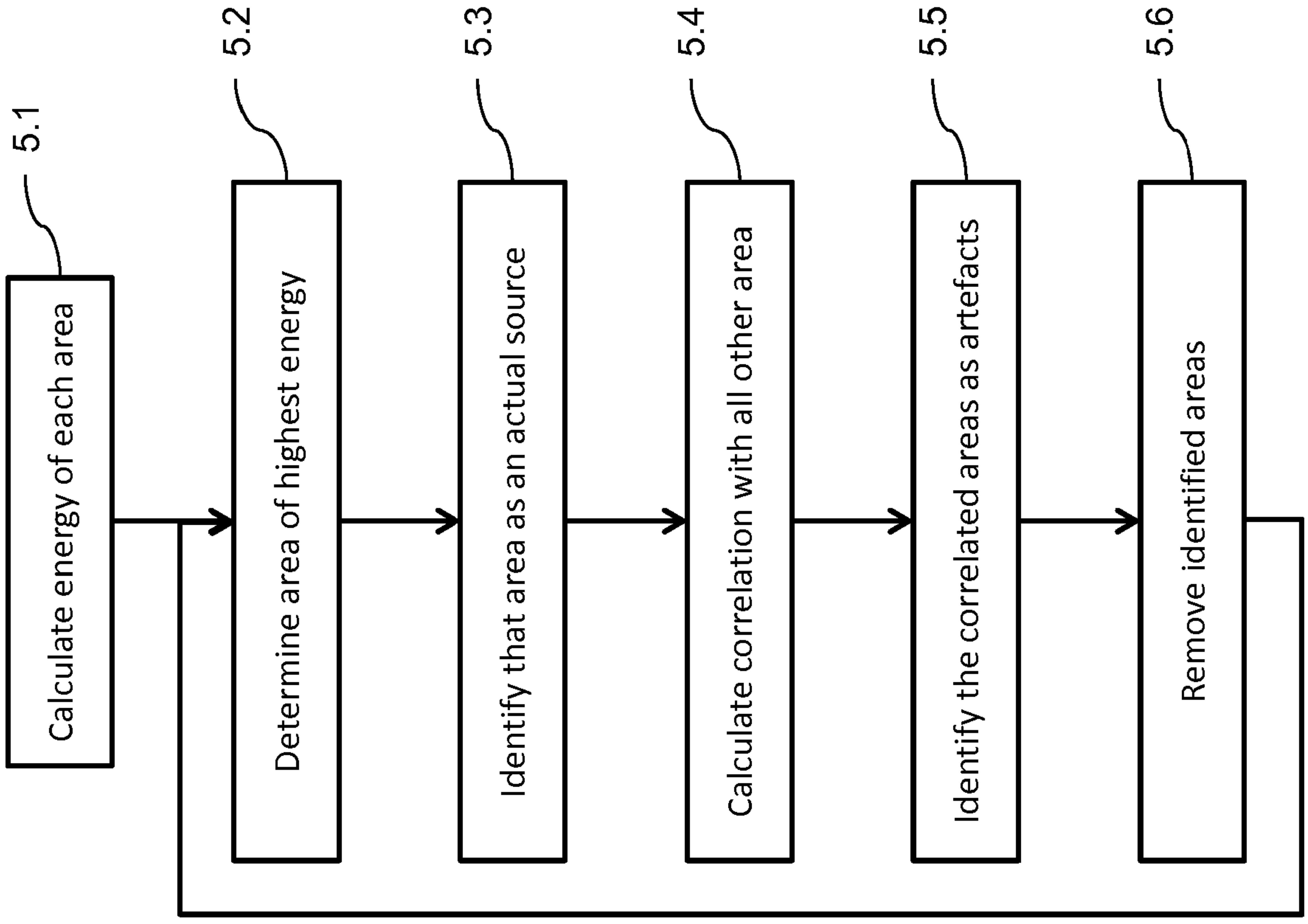
5.1
Calculate energy of each area
5.2
Determine area of highest energy
5.3
Identify that area as an actual source
5.4
Calculate correlation with all other area
5.5
Identify the correlated areas as artefacts
5.6
Remove identified areas

Fig. 5

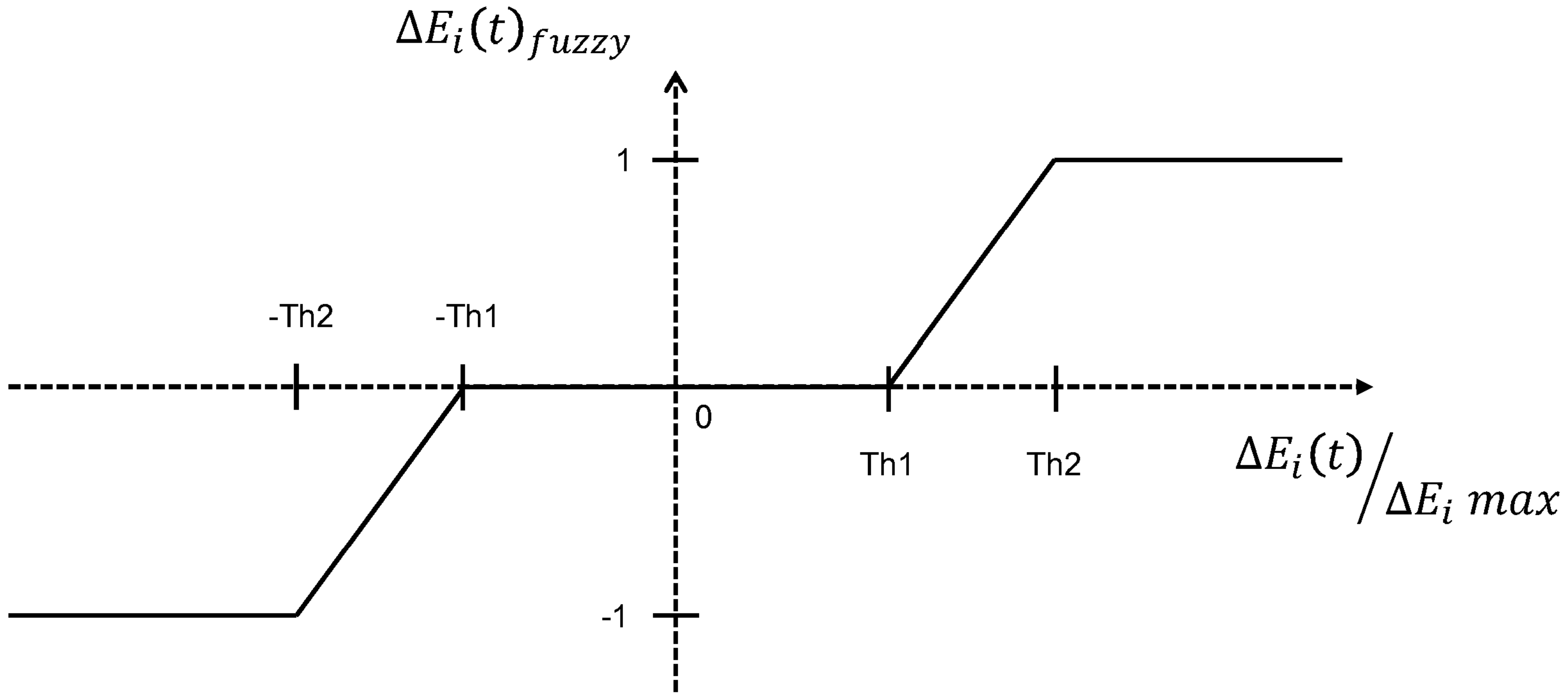

Fig. 6

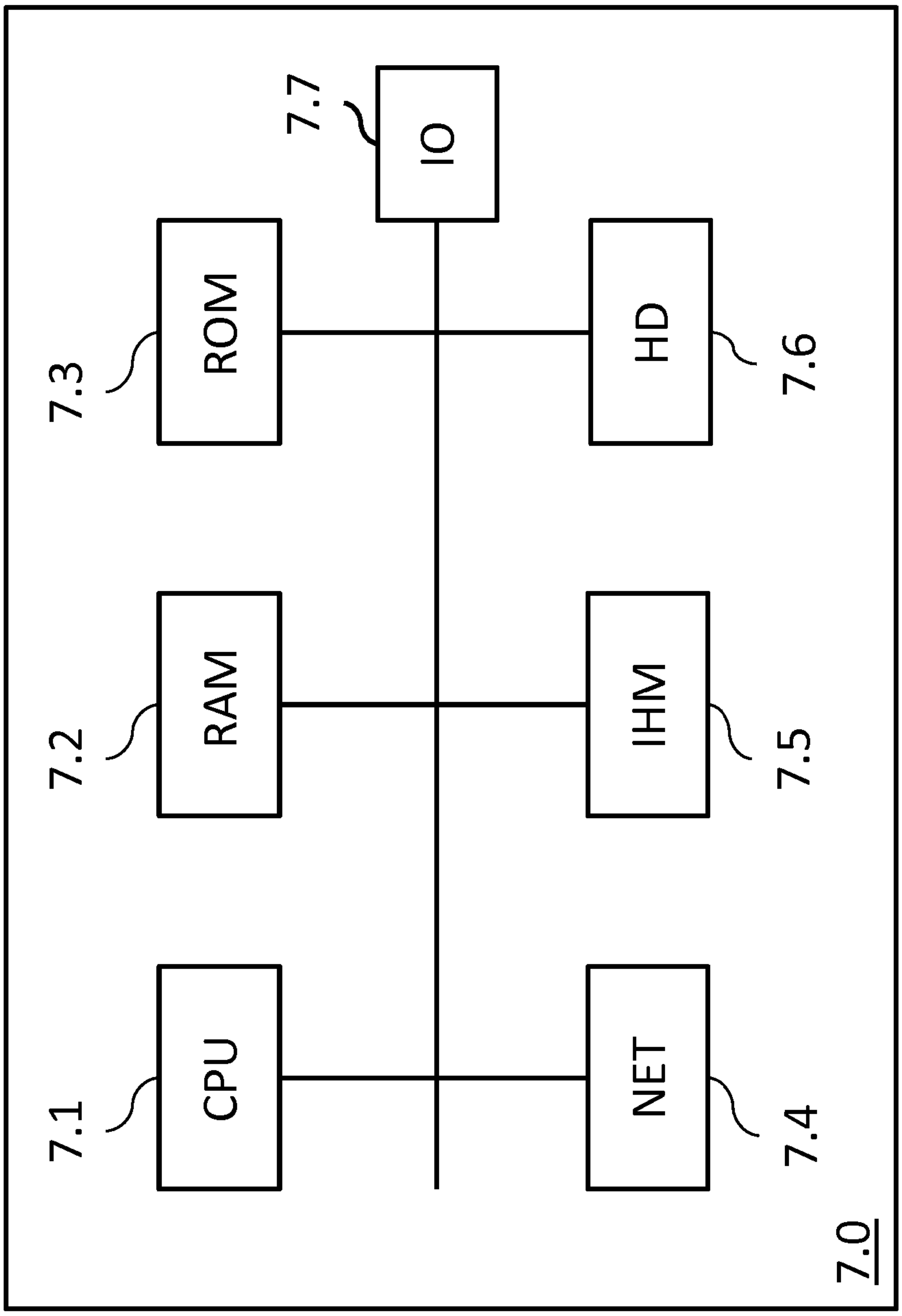

Fig. 7

METHOD AND APPARATUS FOR IDENTIFYING ACTUAL SIGNAL SOURCES AMONG A PLURALITY OF SIGNAL SOURCES WITH ARTEFACTS DETECTION

The present invention concerns a method and a device for sound source separation. More particularly it concerns a method with a detection of artefacts for an improved detection of the signal sources.

Sound source localization aims at isolating signals emitting by sound sources from a source combination of a plurality of audio signals sometimes called a sound mixture. It is a core function in sound sources analysis with applications ranging from Audio-Video surveillance/conferencing, sound enhancement, de-noising to pattern recognition. Classically, in the domain of sound source localization only the direction of the sound source is of interest. A precise localization including the distance of the source is not what is looked for.

A typical sound source localization system is illustrated on **figure 1**. It is considered two audio sources **101** and **102**, each emitting audio signals **106** and **107**. The sound source localization comprises a microphone array **103**, **104** and **105**, composed of at least two microphones. These microphones **103**, **104** and **105** are used to capture a multichannel audio combination of signals. This combination of audio signals is composed by the set of individual channel combination of signals **106** and **107** captured at each microphone of the array. The microphone array is splitting the combination of signals emitted by the sound source into a plurality of channels, each channel consisting in its own representation of the combination of signals. The system also comprises a computing device **108** for executing a sound source localization algorithm. This method is used to estimate the spatial images of the sound sources **109** from the combination of captured audio signals. It is worth noting that the output of the sound source localization algorithm is not the source signals themselves but the estimated spatial images of the sources. These spatial images are defined as the sound signal of each source as recorded at the microphone level. In other words, they are the contributions of each source in the recorded combination of signals.

The microphone array records the multichannel combination of sound sources predominant at different times, located at different positions and exhibiting different spectral contents. The sound source localization method exploits those differences. A first difference is spatial diversity. It stands for the time differences of arrival of a given source at two different microphones. A second difference is spectral diversity. A third difference is temporal diversity, meaning the predominance of some signals emitted by given sources at different moments.

Numerous sound source localization techniques have been proposed, ranging from early multichannel spatial filtering (beamforming) and single-channel spectral modeling techniques to recent techniques jointly exploiting spatial and spectral cues. Angular Spectra methods have been developed in the last decades for cases where beamforming fails: small size/small number of microphones arrays, which is the case for all the current emerging applications. The most well-known angular spectra methods are MUSIC (standing for Multiple Signal Classification), GCC-PHAT (standing for Generalized Cross Correlation with Phase Transform) and SRP-PHAT (standing for Steered Response Power with Phase Transform), and MVDR (standing for Minimum Variance Distortionless Response).

Figure 2 illustrates angular spectra calculation **2.1**. Signals from each microphone, **103**, **104** and **105** on **Figure 1**, minus one reference microphone undergo preliminary signal processing aiming mainly at passing in the frequency space the original temporal signals in a step **2.2**. This is typically done by applying a Fourier transform.

One of the microphones is chosen as a reference microphone.The signals issued from the other microphones are considered relatively to this reference microphone as detailed below. Next, a step **2.3** consists in getting angular spectra of the pre-processed signals. An instant angular spectrum is a value:

$$spec_{inst}(\theta,\varphi)(t)$$

which is relative to a sound activity depending on the direction. In this document, activity of a signal is to be considered as any measure related to the amplitude, or intensity, or energy of the signal, or normalized power, or any value indicative of the strength of sound coming from one direction.

To be more specific, $spec_{inst}$ is not associated to one particular physical quantity. This value is representative of the power of the sound signal regarding the direction. For instance, this value is used in an well-known localization named SRP PHAT, there is for each time-frequency bin (f,t):

$$spec(\theta,\varphi,t,f) = real(X_1 * \frac{\overline{X_2}}{abs\left(X_1 * \overline{X_2}\right)} * e^{-i\Delta_{phase}(\theta,\varphi)})$$

Where

- X_1 et X_2 are transformed (STFT) of the amplitude sound signal (X_1 and X_2 are function to time and frequency), and
- $\Delta_{phase}(\theta,\varphi)$ is the theoretical phase difference (in rad) that would exist between the amplitudes of the sound signal received by each microphone, said sound signal coming from a direction (θ,φ) with a frequency f.

In this case the value *spec* is consistent with a normalized power (i.e without unit). The closer the sound signal is to the direction (θ,φ), the higher is the value taken by *spec*. In other words, *spec* is an abstract value representative of the “sound signal quantity” coming from a given direction (θ,φ).

Direction is given by two angular values (θ,φ) defining a polar coordinate system in three dimensions. It gives a value relative to the level of sound for each direction at instant t. For example, the reference microphone may be the reference of the polar coordinate system in three dimensions defined by (θ,φ).

Signals being decomposed in frequency, the result is typically given as a two dimensional curve along the two angular dimensions θ and φ, for each time-frequency bin (t,f). This result is then subsequently integrated over frequencies as follows:

$$spec_{inst}(\theta,\varphi)(t) = \sum_{f} spec_{tf}(\theta,\varphi)(t,f) \qquad (1)$$

In this equation, (t,f) corresponds to a time-frequency bin value. A (t,f) bin corresponds to an elementary discrete division of time (time frame or time slot) and frequency, after preliminary signal processing, typically a Fourier transform.

The result of the source localization method computation step **2.3** is, for each time-frequency bin, a two dimensional curves along the angles (θ,φ). This two dimensional curve presents some local maximums or peaks. These peaks give candidates for the location of an audio source. The location being given by the coordinates of the peak corresponding to a couple (θ,φ).

In a step **2.4**, these peaks are extracted and sorted according to the value of the peak. In a step **2.5**, knowing the number N of sources to be located, the N peaks with the highest values of angular spectra are identified as corresponding to the N sources. The coordinates of these N peaks give the location of the sources **2.6**.

Actually peaks in the two dimensional curve of angular spectra comes from the actual signal sources, but they may result from a direct or an indirect reception of signal from the sources. Peaks formed by the direct reception of a signal from one source are the one that are looked for. Peaks formed indirectly are artefacts that may advantageously be ignored. These artefacts may be created by reflexion of signals on obstacles, interferences between direct signals or by geometry-related mathematical effects of the microphones array such as symmetries. In the two dimensional curve of angular spectra, there is no way to discriminate peaks due to direct signals from sources and artefacts.

When selecting the highest peak to be used in the source localization, peaks due to artefacts gives false locations because they are not generated by direct signals.

The present invention has been devised to address one or more of the foregoing concerns. It is proposed to consider lowering the false detection of sources by studying the evolution over time of the detected sources. As artefacts leading to false detection are formed by the actual sources, by reflexion or other means, it is assumed that the activity of an artefact is correlated to the activity of the actual source it is formed from. Detected sources are grouped based on the activity correlation. For each group, the detected source with the highest activity in the group is assumed to be formed by a direct signal from an actual source. Other detected sources in the group, which are correlated with this one, are assumed to be artefacts.

According to a first aspect of the invention there is provided a method of identifying actual signal sources among a plurality of signal sources, wherein the method comprises: localizing the plurality of signal sources by a method for localizing signal sources; determining temporal correlations between quantities representative of the activity of the localized signal sources; and identifying actual signal sources among the signal sources as being the signal source with the highest activity within a group of correlated signal sources.

Accordingly, the present invention makes it possible to get rid of most artefacts. Hence, the detection of the actual sources is improved with a lesser level of false detection.

In an embodiment, the localization step comprising the computation of a two dimensional angular spectra curve, the method comprises: determining areas of interest associated to signal sources based on the density of peaks in the two dimensional angular spectra curve.

In an embodiment, the method comprises: determining areas of interest associated to signal sources based on the density of peaks in an histogram representing the accumulation over time of the two dimensional angular spectra curve.

Accordingly, the detection of area of interest is more reliable due to the accumulation of results over time.

In an embodiment, the method comprises: determining an activity parameter for an area of interest based on the accumulation of angular spectra values over said area of interest to be used as the quantity representative of the activity of an area of interest.

In an embodiment, determining temporal correlations between quantities representative of the activity of the localized sources comprises: calculating the plot of the two quantities over time; calculating the linear regression line associated with the plot; and calculating a correlation score defined by the normalization of the average least square error between the points of the plot and the corresponding linear regression line, said localized sources being determined as correlated if the correlation score is greater than a predetermined threshold.

In an embodiment, determining temporal correlations between quantities representative of the activity of the localized sources comprises: calculating a function indicating if a quantity representative of the concentration of the activity of a first signal source in a given area of interest evolves similarly to a second quantity representative of the activity of a second signal source.

In an embodiment, said quantity representative of the concentration of the activity of a signal source is based on the difference of the activity parameter between two subsequent time frames.

In an embodiment, said function is based on the normalization of said difference.

In an embodiment, said function is based on a score defined for positive value of said normalized difference to be 0 below a first threshold, 1 over a second threshold, and a value between 0 and 1 linearly related to said normalized difference for values between the first and the second threshold, the second threshold being greater than the first threshold, said function being calculated symmetrically for negative values.

In an embodiment, said function is the minimum of the absolute value of said normalized difference when they have the same sign, and the opposite of the mean of the two absolute value of said normalized difference when they don't have the same sign.

According to another aspect of the invention there is provided a method for signal source localization, wherein the method comprises: identifying actual signal source localization according to the invention.

According to another aspect of the invention there is provided a device for identifying actual signal sources among a plurality of signal sources, wherein the device comprises: a localizing module for localizing the plurality of signal sources by a method for localizing signal sources; a determining module for determining temporal correlations between quantities representative of the activity of the localized signal sources; and an identifying module for identifying actual signal sources among the signal sources as being the signal source with the highest activity within a group of correlated signal sources.

In an embodiment, the localizing module comprising the computation of a two dimensional angular spectra curve, the device comprises: a determining module for determining areas of interest associated to signal sources based on the density of peaks in the two dimensional angular spectra curve.

In an embodiment, the device comprises: a determining module for determining areas of interest associated to signal sources based on the density of peaks in an histogram representing the accumulation over time of the two dimensional angular spectra curve.

In an embodiment, the device comprises: a determining module for determining an activity parameter for an area of interest based on the accumulation of angular spectra values over said area of interest to be used as the quantity representative of the activity of an area of interest.

In an embodiment, the determining module for determining temporal correlations between quantities representative of the activity of the localized sources comprises: a calculating module for calculating the plot of the two quantities over time; a calculating module for calculating the linear regression line associated with the plot; and a calculating module for calculating a correlation score defined by the normalization of the average least square error between the points of the plot and the corresponding linear regression line, said localized sources being determined as correlated if the correlation score is greater than a predetermined threshold.

In an embodiment, the determining module for determining temporal correlations between quantities representative of the activity of the localized sources comprises: a calculating module for calculating a function indicating if a quantity representative of the concentration of the activity of a first signal source in a given area of interest evolves similarly to a second quantity representative of the activity of a second signal source.

In an embodiment, said quantity representative of the concentration of the activity of a signal source is based on the difference of the activity parameter between two subsequent time frames.

In an embodiment, said function is based on the normalization of said difference.

In an embodiment, said function is based on a score defined for positive value of said normalized difference to be 0 below a first threshold, 1 over a second threshold, and a value between 0 and 1 linearly related to said normalized difference for values between the first and the second threshold, the second threshold being greater than the first threshold, said function being calculated symmetrically for negative values.

In an embodiment, said function is the minimum of the absolute value of said normalized difference when they have the same sign, and the opposite of the mean of the two absolute value of said normalized difference when they don't have the same sign.

In an embodiment, the device comprises: a device for identifying actual signal source localization according to the invention.

According to another aspect of the invention there is provided a computer program product for a programmable apparatus, the computer program product comprising a sequence of instructions for implementing a method according to the invention, when loaded into and executed by the programmable apparatus.

According to another aspect of the invention there is provided a computer-readable storage medium storing instructions of a computer program for implementing a method according to the invention.

At least parts of the methods according to the invention may be computer implemented. Accordingly, the present invention 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". Furthermore, the present invention may take the form of a computer program product embodied in any tangible medium of expression having computer usable program code embodied in the medium.

Since the present invention can be implemented in software, the present invention can be embodied as computer readable code for provision to a programmable apparatus on any suitable carrier medium. A tangible carrier medium may comprise a storage medium such as a floppy disk, a CD-ROM, a hard disk drive, a magnetic tape device or a solid state memory device and the like. A transient carrier medium may include a signal such as an electrical signal, an electronic signal, an optical signal, an acoustic signal, a magnetic signal or an electromagnetic signal, e.g. a microwave or RF signal.

Embodiments of the invention will now be described, by way of example only, and with reference to the following drawings in which:

Figure 1 illustrates a typical sound source separation system;

Figure 2 illustrates angular spectra calculation;

Figure 3 illustrates main steps of a process for sound sources localization with artefacts detection in an embodiment of the invention;

Figure 4 illustrates the main steps of a process for determining the areas of interest according to an embodiment of the invention;

Figure 5 illustrates the main steps of a process for calculating the correlations and identifying the areas of interest corresponding to the actual sources according to an embodiment of the invention;

Figure 6 illustrates the score defined for estimating the variation of the energy associated to an area of interest according to an embodiment of the invention; and

Figure 7 is a schematic block diagram of a computing device for implementation of one or more embodiments of the invention.

Considering the angular spectra calculation and the extraction of peaks, the solutions obtained with the angular spectra methods are not completely stable over time. This is the case even when considering fixed sources that do not move. The received signal is subject to intrinsic fluctuations due to noise and differences in the interactions between signals. The locations of the different peaks, whatever their actual origin, actual sources or artefacts, vary over time around some average locations. This is detrimental to the efficiency of subsequent signal processing that it is wanted to be applied to it, including artefacts removal which is one of the objectives here. One solution to get rid of this variability is to consider histograms representing the accumulation of the results over time. Other solutions may also be considered, like summing instantaneous angular spectra over time, or taking for each spatial coordinate the maximum of angular spectra over time.

Some additional details may be useful to describe an example of the use of histograms for the foregoing issue. For each time frame, an angular spectra two dimensional curve is computed. It is referred to as “instant” angular spectrum at time frame *t*. It is possible to use a localization algorithm on this instant angular spectrum at time frame *t*, hence exhibiting the *n* most probable sources locations at time frame *t*. Those sources locations are referred to as “instant” “solutions” at time frame *t*. There are usually several solutions for each frame *t*. All those solutions may be accumulated in a two dimensional histogram over time, by summing the number of solutions obtained for each time frame.

An advantage of doing so is keeping track of the solutions over time, hence reducing the effects of variability. Another advantage of doing so is to reduce the effects of the noise. Pure noise does not create peaks of its own as it is assumed to be evenly spread over time. It only increases the fluctuations associated with real peaks whether they are formed by actual sources or strong artefacts. Hence, noise-related solutions tend to be more sparsely distributed than sources-related solutions, which tend to translate into smaller and more isolated peaks on the histogram.

It is defined areas of interest as the high density areas of peaks in the histogram. They appear by contrast with areas of low density of peak or areas with no peaks at all. It is assumed that they correspond to actual sources or strong artefacts. The areas of interest are therefore candidates for the location of the actual sources.

Determining the areas of interest must not necessarily be done on histograms. While the use of histograms is a preferred embodiment as it leads to a more robust estimation, in some embodiment, the areas of interest may be determined directly on the two dimensional angular spectra curve. The highest peak will be considered as the areas of interest. It will be described in detail the embodiment based on using histograms, but this example is not limitative.

Once the areas of interest are determined, either obtained from histograms or not, actual sources need to be discriminate from the artefacts. Areas of interest are never due to only fluctuations in the received signal. They are assumed to correspond to either an actual source or a strong artefact. The strong artefacts are physically related to the actual sources. They may be due to reflections of the signal on walls or ground. They may also be due to effects related to the geometry of the microphones array such as symmetries. Geometry indeed makes solutions arise in some “phantom locations”, as an echo of the actual sources locations. It is a mathematical effect due to the way the spec is calculated. Whatever the actual physical phenomenon yielding the artefact, an artefact is always associated with an actual source.

Therefore, it is possible to identify artefacts by looking for possible correlation between the activities of the different areas of interest. Actual sources are usually independent. Areas of interest with a correlated activity are likely to be associated with a single actual source. It is safe to say that isolating a group of areas of interest with a correlated activity allows identifying an actual source and the different artefacts associated with this actual source. As the signal generated by a reflection or geometry symmetry in the microphone is typically weaker than the original signal directly received from the actual source, it is assumed that the area of interest corresponding to the highest peak in the angular spectra curve corresponds to the actual source. The remaining areas of interest correlated with this area of interest are identified as artefacts generated by this actual source. Accordingly, by identifying the areas of interest and dividing them into groups of areas of interest having a correlated activity, it is possible to identify the actual source and discriminate them from the areas of interest generated by artefacts.

Figure 3 illustrates main steps of a process for sound sources localization with artefacts detection in an embodiment of the invention.

First, a step **3.1** aims at computing the angular spectra two dimensional curve from the received signals. It consists typically in a step **3.2** of preliminary signal processing, a step **3.3** of angular spectra computation and a step **3.4** of extracting and sorting the peaks in the angular spectra two dimensional curve as described in relation to **Figure 2**.

In a step **3.5**, advantageously, a two dimensional histogram over time is built by accumulating the solutions resulting from the previous step, namely the angular spectra two dimensional curve. The accumulation may be done on a predefined number of time frames or continue during all the process. This step is optional as the areas of interest may be computed directly on a single angular spectra two dimensional curve.

In a step **3.6**, the areas of interest are determined from the histogram. A possible process to do so is detailed below in relation to **Figure 4**.

In a step **3.7**, the correlations between the different areas of interest found in the previous step are determined. Based on these correlations, the actual sources are identified as well as the correlated artefacts. A possible process to do so is detailed below in relation to **Figure 5**.

In a step **3.8**, based on this identification of the actual sources and artefacts, the latters are eliminated to focus on the actual sources.

The artefacts being eliminated, the location of the area of interest associated with the actual sources give the location of these actual sources **3.9**. For example, the centre of the area of interest may be chosen for determining the two angles defining this location.

Figure 4 illustrates the main steps of a process for determining the areas of interest according to an embodiment of the invention.

The process advantageously takes the histogram **4.1** as input. As already explained, in some embodiments, it may take the angular spectra two dimensional curve as well.

In a step **4.2**, a map of density of solutions is computed. A solution corresponds to a peak in either the histogram or the angular spectra two dimensional curve. The density of solutions is defined, for a given position defined by a couple of angles (θ, φ), by the density of solutions in the neighborhood of the position.

One way of calculating this density is the following: a radius R is chosen; the value z of the density at a point of coordinates (θ_1, φ_1) is the number of solutions N in a circle of radius R around the position (θ_1, φ_1):

$$z(\theta_1, \varphi_1) = N$$

An alternate solution is to weight N by the distance to the position (θ_1, φ_1), for example by using a formula like:

$$z(\theta_1, \varphi_1) = \sum_{i=1}^{N} \frac{1}{d_i}$$

where: i is the index of one solution of histogram inside radius R around (θ_1, φ_1), d_i is the distance between (θ_i, φ_i) and (θ_1, φ_1) and N is the total number of solutions of histogram inside radius R around (θ_1, φ_1).

In a step **4.3**, the highest peak in this map of density is searched. This highest peak gives a new area of interest **4.4** centred on this peak.

In a step **4.5**, the solutions, meaning the peaks of the histogram, are eliminated in the new defined area of interest. The goal is to suppress this new area of interest from the histogram to search for the next one. For example, a radius R may be defined and all peaks of the histogram in the area defined by the centre of the area and this radius R is supressed from the histogram. This process yields to an updated histogram **4.6**. From this updated histogram, a new or updated map of density is computed by coming back to step **4.2**. A new area of interest is then computed by iterating on these steps. The process is stopped when a sufficient number of areas of interest have been defined to find all the actual sources. For example, when a predefined number of areas is obtained or when the highest peak in the density map gets lower than a predefined threshold.

Figure 5 illustrates the main steps of a process for calculating the correlations and identifying the areas of interest corresponding to the actual sources according to an embodiment of the invention. This process corresponds to a possible implementation of step **3.7** of **Figure 3**.

It is assumed that the actual sources are independent from each other. Therefore their activities are not correlated. It is also assumed that the artefacts of highest activity depend on the actual sources. Therefore, their activities are correlated to the actual source they depend on. Finally, it is assumed that an actual source activity is always higher than the activity of an artefact which depends on it. This is based on the fact that an original signal is always louder than its reverberation, reflection, or artificial echoes.

The activity of a signal is analogous to its energy, for example the sound energy for an audio signal. More particularly, $E_i(t)$ (activity parameter) is another quantity parameter which indicates that the quantity representative of the activity of an area of interest is more or less concentrated. It may be measured for an area of interest as the received energy or any other variable or metric associated with the sound energy coming from a particular area of interest, for example sound pressure.

Regarding the calculation of the energy of the signal for a given area of interest, microphones record the full sound related to all directions and there is no access to the actual acoustic energy coming from a particular direction. It is needed to determine a metric that is related to this energy and which may be calculated from the available data. This metric should be an equivalent of the acoustic activity in a particular direction and a particular time. Several such metrics may be used, in the detailed embodiment of the invention, the following one is used:

$$E_i(t) = \sum_{(\theta,\varphi)\in[\theta_i\pm\Delta\theta,\varphi_i\pm\Delta\varphi]} \sum_f spec(\theta,\varphi,t,f)$$

Where $spec$ is the angular spectrum already described and calculated and i is the index of the current area of interest which the metric is to be calculated. (θ_i,φ_i) is the angular position **4.4** corresponding to the area of interest i as obtained for example in step **4.3** of **Figure 4**. $\Delta\theta$ and $\Delta\varphi$ are used to define the spatial extension of the area of interest. For example, if considering areas defined by a circle of radius R, $\Delta\theta = \Delta\varphi = R$. This formula has the supplementary advantage to reduce the random noise, since the intrinsic fluctuations effects are partially cancelled out in the averaging over the whole area of interest.

In a step **5.1**, the energy of each area of interest is calculated, for example with the method described above. Next, the area of interest having the highest energy is determined in a step **5.2**. If this area of interest was an artefact it would be correlated to another area of interest having a higher energy than its own energy. Being the area with the highest energy, this area is necessarily related to an actual source and not an artefact. It is identified accordingly as an actual source in a step **5.3**.

In a step **5.4**, the activity of the last identified actual source is tested against all the remaining areas of interest in order to identify all the areas of interest with a correlated energy. A possible method to do so is to consider the energy as calculated in step **5.1** over the time. If the “energy” is defined by an activity parameter called $Ei(t)$ related to the “energy” of an area of interest i over the time, it is possible to test the correlation of two areas i and j by calculating the $Ei(t)$ versus $Ej(t)$ plot. Next the associated linear regression line is calculated. A correlation score may be defined by the normalization of the average least square error between the points of the plot and the linear regression line. The two areas are determined as correlated if the correlation score is greater than a predetermined threshold.

In a step **5.5** all the area of interest determined as being correlated with the actual source are identified as being artefacts. The identified actual source and all this artefacts are removed in a step **5.6**.

The process loops back to step **5.2** for the identification of the next actual source and its artefacts until all the area of interest have been identified.

Alternatively, the correlation computation step **5.4** may be implemented using fuzzy logic. In many situations, the actual sources and the artefacts may not be fully correlated due to some interferences, noise and secondary sources. In such situation, the correlation computation may be improved by the use of fuzzy logic. Such a correlation computation is described now in detail.

This solution aims at determining a score to the predicate: “When $Ei(t)$ increases a lot, $Ej(t)$ increases a lot”. This predicate is assumed to be an equivalent to “The area of interest i is correlated to the area of interest j”. It is worth noting that different actual formulae, following fuzzy logic principle, may be used to express this predicate. The actual formula given here is only an example of such a formula used in the described embodiment.

First a score or function $score1_{fuzzy_i}(t)$ is defined for the predicate “$Ei(t)$ increases a lot”.

$\Delta E_i(t)$ is defined as the variation of the energy at time t:

$$\Delta E_i(t) = \big(E_i(t) - E_i(t-1)\big)$$

$\Delta E_i\ max$ is defined the maximum of the absolute value of this variation over a period of time :

$$\Delta E_i\ max = max_t(|\Delta E_i(t)|)$$

The ratio ${}^{\Delta E_i(t)}\!/_{\Delta E_i\ max}$ is therefore a value related to the level of variation of the energy at time frame t. It corresponds to a normalization of $\Delta E_i(t)$. This value is positive if the energy increases and negative if the energy decreases.

For positive value of the ratio ${}^{\Delta E_i(t)}\!/_{\Delta E_i\ max}$ the score $score1_{fuzzy_i}(t)$ is defined to be 0 below a first threshold Th1, 1 over a second threshold Th2, and a value between 0 and 1 linearly related to

$\Delta E_i(t) / \Delta E_i\,max$ for values between Th1 and Th2. The second threshold Th2 is greater than the first threshold Th1. It means that below the first threshold Th1, the evolution of the energy is assumed to be low, over the second threshold Th2, the evolution of the energy is assumed to be high, and in between, the evolution of the energy is assumed to be intermediate and related to the calculated score $score1_{fuzzy_i}(t)$. For negative values, the score is calculated symmetrically as illustrated on **Figure 6**.

Next the fuzzy score of the predicate “When $Ei(t)$ increases a lot, $Ej(t)$ increases a lot” should be given a fuzzy score as well. It is called $score2_{fuzzy_{ij}}(t)$.

This score may be defined as follow:

$$\begin{cases} \text{if } \left(\text{sign}\left(score1_{fuzzy_i}(t)\right) = sign\left(score1_{fuzzy_j}(t)\right)\right) \\ \text{then } score2_{fuzzy_{ij}}(t) = min\left(score1_{fuzzy_i}(t)), abs(score1_{fuzzy_j}(t))\right) \\ \text{if } \left(\text{sign}\left(score1_{fuzzy_i}(t)\right) \neq sign\left(score1_{fuzzy_j}(t)\right)\right) \\ \text{then } score2_{fuzzy_{ij}}(t) = -\dfrac{abs(score1_{fuzzy_i}(t)) + abs(score1_{fuzzy_j}(t))}{2} \end{cases}$$

A global score $score_{fuzzy\,total}$ for the predicate over the time may then be defined as:

$$score_{fuzzy\,total} = mean\left(score2_{fuzzy_{ij}}(t)\right)$$

$$= \frac{1}{N_{large_{ij}}} \sum_{t \in N_{large_{ij}}} score2_{fuzzy_{ij}}(t)$$

Where $N_{large_{ij}}$ is a subset of all frames, containing all frames t where:

$$\left(score1_{fuzzy_i}(t) \neq 0\right)\ OR\ \left(score1_{fuzzy_j}(t) \neq 0\right)$$

This global score $score_{fuzzy\,total}$ may advantageously be used for determination of the correlation computation step **5.4** in **Figure 5**. If the global score is over a predetermined threshold, the two areas of interest are assumed to be correlated. The described fuzzy logic score to establish the correlation is only an example and other metrics may be used to achieve the same goal.

The method described in the foregoing makes it possible to discriminate between actual signal sources and artefacts issued from these actual sources. By focusing on actual sources, the results of the localization are improved. As actual sources may be discriminated, it is no longer necessary to know the exact number of actual sources before the localization. It is also possible to locate reflexion paths identified as such which may be of interest in some applications. The computation burden added to classical sound source localization is low as the added computations needs are negligible as compared to the angular spectra computation.

Figure 7 is a schematic block diagram of a computing device **7.0** for implementation of one or more embodiments of the invention. The computing device **7.0** may be a device such as a micro-computer, a workstation or a light portable device. The computing device **7.0** comprises a communication bus connected to:

- a central processing unit **7.1**, such as a microprocessor, denoted CPU;

- a random access memory **7.2**, denoted RAM, for storing the executable code of the method of embodiments of the invention as well as the registers adapted to record variables and parameters necessary for implementing the method for encoding or decoding at least part of an image according to embodiments of the invention, the memory capacity thereof can be expanded by an optional RAM connected to an expansion port for example;

- a read only memory **7.3**, denoted ROM, for storing computer programs for implementing embodiments of the invention;

- a network interface **7.4** is typically connected to a communication network over which digital data to be processed are transmitted or received. The network interface **7.4** can be a single network interface, or composed of a set of different network interfaces (for instance wired and wireless interfaces, or different kinds of wired or wireless interfaces). Data packets are written to the network interface for transmission or are read from the network interface for reception under the control of the software application running in the CPU **7.1**;

- a user interface **7.5** may be used for receiving inputs from a user or to display information to a user;

- a hard disk **7.6** denoted HD may be provided as a mass storage device;

- an I/O module **7.7** may be used for receiving/sending data from/to external devices such as a video source or display.

The executable code may be stored either in read only memory **7.3**, on the hard disk **7.6** or on a removable digital medium such as for example a disk. According to a variant, the executable code of the programs can be received by means of a communication network, via the network interface **7.4**, in order to be stored in one of the storage means of the communication device **7.0**, such as the hard disk **7.6**, before being executed.

The central processing unit **7.1** is adapted to control and direct the execution of the instructions or portions of software code of the program or programs according to embodiments of the invention, which instructions are stored in one of the aforementioned storage means. After powering on, the CPU **7.1** is capable of executing instructions from main RAM memory **7.2** relating to a software application after those instructions have been loaded from the program ROM **7.3** or the hard-disc (HD) **7.6** for example. Such a software application, when executed by the CPU **7.1**, causes the steps of the flowcharts shown in **Figures 2** to **5** to be performed.

Any step of the algorithm shown in **Figure 2** to **5** may be implemented in software by execution of a set of instructions or program by a programmable computing machine, such as a PC (“Personal Computer”), a DSP (“Digital Signal Processor”) or a microcontroller; or else implemented in hardware by a machine or a dedicated component, such as an FPGA (“Field-Programmable Gate Array”) or an ASIC (“Application-Specific Integrated Circuit”).

Although the present invention has been described hereinabove with reference to specific embodiments, the present invention is not limited to the specific embodiments, and modifications will be apparent to a skilled person in the art which lie within the scope of the present invention.

Many further modifications and variations will suggest themselves to those versed in the art upon making reference to the foregoing illustrative embodiments, which are given by way of example only and which are not intended to limit the scope of the invention, that being determined solely by the appended claims. In particular the different features from different embodiments may be interchanged, where appropriate.

In the claims, the word “comprising” does not exclude other elements or steps, and the indefinite article “a” or “an” does not exclude a plurality. The mere fact that different features are recited in mutually different dependent claims does not indicate that a combination of these features cannot be advantageously used.

CLAIMS

1. A method of identifying actual sound sources among a plurality of sound sources, wherein the method comprises:
    - localizing the plurality of sound sources by a method for localizing sound sources;
    - determining temporal correlations between quantities representative of the activity of the localized sound sources; and
    - identifying actual sound sources among the sound sources as being the sound source with the highest activity within a group of correlated sound sources.

2. A method according to claim 1, wherein the localization step comprising the computation of a two dimensional angular spectra curve, the method comprises:
    - determining areas of interest associated to sound sources based on the density of peaks in the two dimensional angular spectra curve.

3. A method according to claim 2, wherein the method comprises:
    - determining areas of interest associated to sound sources based on the density of peaks in an histogram representing the accumulation over time of the two dimensional angular spectra curve.

4. A method according to claim 2 or 3, wherein the method comprises:
    - determining an activity parameter for an area of interest based on the accumulation of angular spectra values over said area of interest to be used as the quantity representative of the activity of an area of interest.

5. A method according to any one claim 1 to 4, wherein determining temporal correlations between quantities representative of the activity of the localized sources comprises:
   - calculating the plot of the two quantities over time;
   - calculating the linear regression line associated with the plot; and
   - calculating a correlation score defined by the normalization of the average least square error between the points of the plot and the corresponding linear regression line, said localized sources being determined as correlated if the correlation score is greater than a predetermined threshold.

6. A method according to claim 4, wherein determining temporal correlations between quantities representative of the activity of the localized sources comprises:
   - calculating a function indicating if a quantity representative of the concentration of the activity of a first sound source in a given area of interest evolves similarly to a second quantity representative of the activity of a second sound source.

7. A method according to claim 6, wherein said quantity representative of the concentration of the activity of a sound source is based on the difference of the activity parameter between two subsequent time frames.

8. A method according to claim 7, wherein said function is based on the normalization of said difference.

9. A method according to claim 8, wherein said function is based on a score defined for positive value of said normalized difference to be 0 below a first threshold, 1 over a second threshold, and a value between 0 and 1 linearly related to said normalized difference for values between the first and the second threshold, the second

threshold being greater than the first threshold, said function being calculated symmetrically for negative values.

10. A method according to claim 9, wherein said function is the minimum of the absolute value of said normalized difference when they have the same sign, and the opposite of the mean of the two absolute value of said normalized difference when they don’t have the same sign.

11. A method for sound source localization, wherein the method comprises:
    - identifying actual sound source localization according to any one claim 1 to 10.

12. A device for identifying actual sound sources among a plurality of sound sources, wherein the device comprises:
    - a localizing module for localizing the plurality of sound sources by a method for localizing sound sources;
    - a determining module for determining temporal correlations between quantities representative of the activity of the localized sound sources; and
    - an identifying module for identifying actual sound sources among the sound sources as being the sound source with the highest activity within a group of correlated sound sources.

13. A device according to claim 12, wherein the localizing module comprising the computation of a two dimensional angular spectra curve, the device comprises:
    - a determining module for determining areas of interest associated to sound sources based on the density of peaks in the two dimensional angular spectra curve.

14. A device according to claim 13, wherein the device comprises:

- a determining module for determining areas of interest associated to sound sources based on the density of peaks in an histogram representing the accumulation over time of the two dimensional angular spectra curve.

15. A device according to claim 13 or 14, wherein the device comprises:
    - a determining module for determining an activity parameter for an area of interest based on the accumulation of angular spectra values over said area of interest to be used as the quantity representative of the activity of an area of interest.

16. A device according to any one claim 12 to 15, wherein the determining module for determining temporal correlations between quantities representative of the activity of the localized sources comprises:
    - a calculating module for calculating the plot of the two quantities over time;
    - a calculating module for calculating the linear regression line associated with the plot; and
    - a calculating module for calculating a correlation score defined by the normalization of the average least square error between the points of the plot and the corresponding linear regression line, said localized sources being determined as correlated if the correlation score is greater than a predetermined threshold.

17. A device according to claim 15, wherein the determining module for determining temporal correlations between quantities representative of the activity of the localized sources comprises:
    - a calculating module for calculating a function indicating if a quantity representative of the concentration of the activity of a first sound source in a given area of interest evolves similarly to a second quantity representative of the activity of a second sound source.

18. A device according to claim 17, wherein said quantity representative of the concentration of the activity of a sound source is based on the difference of the activity parameter between two subsequent time frames.

19. A device according to claim 18, wherein said function is based on the normalization of said difference.

20. A device according to claim 19, wherein said function is based on a score defined for positive value of said normalized difference to be 0 below a first threshold, 1 over a second threshold, and a value between 0 and 1 linearly related to said normalized difference for values between the first and the second threshold, the second threshold being greater than the first threshold, said function being calculated symmetrically for negative values.

21. A device according to claim 20, wherein said function is the minimum of the absolute value of said normalized difference when they have the same sign, and the opposite of the mean of the two absolute value of said normalized difference when they don’t have the same sign.

22. A device for sound source localization, wherein the device comprises:

- a device for identifying actual sound source localization according to any one claim 12 to 21.

23. A computer program product for a programmable apparatus, the computer program product comprising a sequence of instructions for implementing a method according to any one of claims 1 to 11, when loaded into and executed by the programmable apparatus.

24. A computer-readable storage medium storing instructions of a computer program for implementing a method according to any one of claims 1 to 11.

25. A method of identifying actual sound sources among a plurality of sound sources substantially as hereinbefore described with reference to, and as shown in Figure 3 to 5.

14 10 15