Hindawi Publishing Corporation
The Scientific World Journal
Volume 2014, Article ID 582397, 13 pages
http://dx.doi.org/10.1155/2014/582397

Research Article

Incident Signal Power Comparison for Localization of Concurrent Multiple Acoustic Sources

Daniele Salvati[1] and Sergio Canazza[2]

[1] *Department of Mathematics and Computer Science, University of Udine, 33100 Udine, Italy*
[2] *Department of Information Engineering, University of Padova, 35131 Padova, Italy*

Correspondence should be addressed to Sergio Canazza; canazza@dei.unipd.it

Received 5 August 2013; Accepted 2 January 2014; Published 20 February 2014

Academic Editors: S. Bourennane and J. Marot

In this paper, a method to solve the localization of concurrent multiple acoustic sources in large open spaces is presented. The problem of the multisource localization in far-field conditions is to correctly associate the direction of arrival (DOA) estimated by a network array system to the same source. The use of systems implementing a Bayesian filter is a traditional approach to address the problem of localization in multisource acoustic scenario. However, in a real noisy open space the acoustic sources are often discontinuous with numerous short-duration events and thus the filtering methods may have difficulty to track the multiple sources. Incident signal power comparison (ISPC) is proposed to compute DOAs association. ISPC is based on identifying the incident signal power (ISP) of the sources on a microphone array using beamforming methods and comparing the ISP between different arrays using spectral distance (SD) measurement techniques. This method solves the ambiguities, due to the presence of simultaneous sources, by identifying sounds through a minimization of an error criterion on SD measures of DOA combinations. The experimental results were conducted in an outdoor real noisy environment and the ISPC performance is reported using different beamforming techniques and SD functions.

1. Introduction

The sensory capacity to analyze acoustic space is a very important function of an auditory system. The need for the development of an understanding of the sound environment has attracted many researchers over the past twenty years to build sensory systems that are capable of locating acoustic sources in space. Acoustic source localization (ASL) is an important task in a growing number of applications. Fields of application in which identification of the location of acoustic sources is desired include audio surveillance, teleconferencing systems, hands-free acquisition in car, system monitoring, human-machine interaction, musical control interfaces, videogames, virtual reality systems, voice recognition, fault analysis of machinery, autonomous robots, processors for digital hearing aids, high-quality recording, multiparty telecommunications, dictation systems, and acoustic scene analysis. The aim of an ASL system is to estimate the position of sound sources in space by analyzing the sound field with a microphone array, a set of microphones arranged to capture the spatial information of sound.

Several application areas that may potentially provide advantages in using the acoustic location have led to the development of many signal processing algorithms, which mostly consider the specific acoustic environment, the signal properties, and the localization goal.

ASL can be performed by two basic methods: indirect and direct. The indirect approach is used to estimate source positions by implementing the following two steps: in the first one, a set of time difference of arrivals (TDOAs) are estimated using measurements across various combinations of microphones, and in the second one, when the position of the sensors and the speed of sound are known, the source positions can be estimated using geometric considerations and approximate estimators: closed-formed estimators based on a least squares solution [1–7] (for an overview on closed-form estimators, see [8]) and iterative maximum likelihood estimators [9–15]. The direct approach involves the search space by constructing a spatial energy map and estimating, for each possible point of interest, the values that maximize a specific likelihood function that provides a coherent value from the entire system of arrays. The position of the sources can be estimated directly and spatial likelihood functions can be defined [16–20].

In near-field conditions, since the sources radiate the sound in spherical waves, a hyperboloid describes all of the possible points of an acoustic source that generates the same TDOA to an array of two microphones. Indirect methods aim at estimating TDOAs for microphone pairs, typically using the generalized cross-correlation (GCC) [21] and the adaptive eigenvalue decomposition (AED) [22] based on the blind system identification, which focuses on the impulse responses between the source and the microphones. The extension of the AED in the case of multiple microphones was proposed in [23], and it is efficiently performed with a normalized multichannel frequency-domain least mean square algorithm [24, 25]. However, the steered response power (SRP) is a direct method based on maximizing the power output of a beamformer. Beamforming is a combination of the delayed signals from each microphone in a manner in which an expected pattern of radiation is preferentially observed. In general, the SRP is computed in frequency-domain using the fast Fourier transformer on a signal portion, calculating the response power on each frequency bin, and subsequently fusing these estimates to obtain the final result. The conventional SRP is performed with the delay and sum beamformer [26]; it consists of the synchronization of signals that steer the array in a certain direction, and it sums the signals to estimate the power of the spatial filter. The SRP phase transform (SRP-PHAT) [18] is a widely used filtered beamforming. PHAT filter [21] places equal importance on each frequency by dividing the spectrum by its magnitude. It normalizes the amplitude of the spectral density using only the phase information with the advantage to improve performance in case of moderate noise and reverberation. SRP-PHAT is deeply used due to the fact that it can be efficiently computed by coherent summing the GCC-PHAT from all of the microphone pairs for each possible point of interest. The high-resolution SRP has been developed to improve the performance of the spatial filter, and the adaptive beamformer is called the minimum variance distortionless response (MVDR) due to Capon [27]. The multiple signal classification (MUSIC) algorithm is based on an eigen subspace decomposition method [28, 29], and the estimation of signal parameters via rotational invariance techniques (ESPRIT) is based on subspace decomposition exploiting the rotational invariance [30–32].

In far-field conditions, we are no longer able to detect the spherical wavefront in relationship with the distance of source from an array and the size of the array, and the wavefront is approximate to a plane. In this condition, with an array of microphones, we are able to estimate only the direction of arrival (DOA) of the source but not its distance from the array. In the far-field case the hyperboloid, which is the locus of points that generates the same TDOA to a microphone pair, can be approximated with the cone whose vertex is located at the midpoint of the array. Thus, we need a network of arrays to perform the localization of a source (at least two arrays for two-dimensional space). Hence, the position estimation is computed by intersection of lines and by an approximated solution for overdetermined systems using the linear least squares method. In the case of an array containing M microphones ($M > 2$) the DOA estimation can be computed with the indirect method of multichannel cross-correlation coefficient (MCCC) [33, 34], which is based on TDOAs estimation using GCC, and on the use of the spatial prediction error to measure the correlation among multiple signals. It has the advantage of using the redundant information between microphones to estimate the DOA in a more robust manner under a reverberant and noisy condition. The family of SRP direct methods with an array is used to estimate the DOAs of sources by picking the values corresponding to the principal peaks of the steered response power of a beamforming.

Recently, more sophisticated algorithms have been proposed for time delay estimation that use minimum entropy [35, 36] and blind source separation [37–39]. In [39], the authors demonstrate that the broadband independent component analysis methods are more robust against high background noise levels compared with the conventional GCC-PHAT approach.

Both indirect and direct methods have been tested in many single source scenarios; however, in multiple sources cases, they require new considerations. Several works address the problem of multiple sources using a Bayesian approach based on the tracking of the sources and using Kalman filter [40–47] and Particle filter [19, 48–53]. Some studies consider an approach without tracking in reverberant environments [39, 54–57].

In a real open space, the traditional techniques based on Bayesian filters (Kalman and Particle filters) are difficult to apply for localization of concurrent multiple acoustic sources, because sources are often discontinuous with numerous short-duration events and the spatial resolution may be poor in some areas of analysis. Note that in practical applications the localization in a open space needs a reduced number of arrays, due to limited space for installing it and not to invade the monitoring spaces in an excessive way. Besides, methods based on movement tracking can fail in some specific situations: during the initialization phase of the filter, in the presence of sources with unpredictable trajectory (e.g., in the case of rapid changes of the velocity vector), and when two sources have intersecting trajectories.

As a solution to this problem, we present the approach based on the incident signal power comparison (ISPC). A preliminary work was proposed in [58, 59]. This paper describes a detailed step-by-step ISPC algorithm introducing a diagonal loading (DL) [60, 61] for MVDR beamforming, which gives more stable ISP estimation, and reporting new experimental results in a real scenario. ISPC is designed for a distributed array system, and it is based on source extraction and on a verification of similarity among sound sources. The first step consists of source extraction using beamforming techniques and estimation of the incident signal power (ISP) of every source captured on the array. The second step involves the comparison of the ISP spectrum from different arrays using a spectral distance (SD) measure. The ISP spectrum permits identification of sounds so that the spectrum power distance minimizes an error criterion. Therefore, the identification of the correct combination of DOAs is estimated by identifying the minor value of SD measures.

The location in a free-field outdoor environment can be employed for audio surveillance, sound monitoring, and analysis of acoustic scenes. In particular, Section 5 describes a prototype system for multiple source localization in a public space for monitoring a large area with a joint audio-video system, in which the positional estimates by acoustic analysis are used to steer a video-camera consequently.

The paper is organized as follows. After presenting the signal model in Section 2, the multiple sources localization problem is described in Section 3. In Section 4 the ISPC algorithm is presented. Finally, Section 5 illustrates experimental results obtained in a real-world scenario.

2. Signal Model

We assume N acoustic sources and R arrays, each composed of M microphones, and consider the omnidirectional characteristics of both the sources and the microphones. We will refer to the model of discrete-time obtained by performing a sampling operation on the continuous-time signal $x(t)$ with a uniform sampling period T_s. A discrete-time signal is expressed by

$$x\left(kT_S\right) = x\left(\frac{k}{f_s}\right) \quad k = 0, 1, \dots, \tag{1}$$

where k is the sample time index and f_s is the sampling frequency. As usual, we will allow the sample period T_s to remain implicit and refer to it simply as $x(k)$.

The free-field discrete-time signal received by the mth microphone of the rth array can be modeled as

$$x_{rm}\left(k\right) = \sum_{n=1}^{N} \alpha_{rnm} s_n\left(k - k_{rn} - \tau_{rnm}\right) + v_{rm}\left(k\right), \tag{2}$$

where α_{rnm} is the attenuation of the sound propagation (inversely proportional to the distance from source n to microphone m of array r), $s_n(k)$ are the unknown uncorrelated source signals, k_{rn} is the propagation time from the unknown source n to the reference sensor of array r, τ_{rnm} is the TDOA of the signal n between the mth microphone and the reference of the rth array, and $v_{rm}(k)$ is the additive noise signal at the sensor m of array r, assumed to be uncorrelated with not only all of the source signals but also with the noise observed at the other sensors.

In far-field case the relationship between TDOA and DOA can be solved easily with geometrical considerations. Therefore, for a generic pair of microphones with TDOA τ_{rn}, DOA estimate is obtained as

$$\theta_{rn} = \arcsin\left(\frac{\tau_{rn} c}{d}\right), \tag{3}$$

where c is the speed of sound and d the distance between microphones.

The vector $\boldsymbol{\Theta}_n$ for each source n, considering the signal model (2), is defined by

$$\boldsymbol{\Theta}_n = \left[\theta_{1n}, \theta_{2n}, \dots, \theta_{Rn}\right]^T \tag{4}$$

which contains the DOAs of the acoustic source n by each array. In the case of N sources and R arrays, we can write the matrix $R \times N$, which contains all DOAs of distributed array network as

$$\boldsymbol{\Theta} = \left[\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2, \dots, \boldsymbol{\Theta}_N\right] = \begin{bmatrix} \theta_{11} & \theta_{12} & \cdots & \theta_{1N} \\ \theta_{21} & \theta_{22} & \cdots & \theta_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{R1} & \theta_{R2} & \cdots & \theta_{RN} \end{bmatrix}. \tag{5}$$

The estimated DOAs angles, obtained for each array r, are written with the following vector:

$$\underline{\widehat{\boldsymbol{\Theta}}}_r = \left[\widehat{\theta}_{r1}, \widehat{\theta}_{r2}, \dots, \widehat{\theta}_{rN}\right], \tag{6}$$

where we consider the DOA values in ascending order ($\widehat{\theta}_{r1} < \widehat{\theta}_{r2} < \widehat{\theta}_{r3}$, etc.). Next, the estimated sorted DOAs matrix $\underline{\widehat{\boldsymbol{\Theta}}}$ is defined as

$$\underline{\widehat{\boldsymbol{\Theta}}} = \begin{bmatrix} \underline{\widehat{\boldsymbol{\Theta}}}_1 \\ \underline{\widehat{\boldsymbol{\Theta}}}_2 \\ \vdots \\ \underline{\widehat{\boldsymbol{\Theta}}}_R \end{bmatrix} = \begin{bmatrix} \widehat{\theta}_{11} & \widehat{\theta}_{12} & \cdots & \widehat{\theta}_{1N} \\ \widehat{\theta}_{21} & \widehat{\theta}_{22} & \cdots & \widehat{\theta}_{2N} \\ \vdots & \vdots & \ddots & \ddots \\ \widehat{\theta}_{R1} & \widehat{\theta}_{R2} & \cdots & \widehat{\theta}_{RN} \end{bmatrix}. \tag{7}$$

The position of the source n can be calculated by combining the DOAs estimated by the R arrays for that source.

3. Multiple Sources Localization

The multiple sources localization problem is to correctly assign the R DOAs values to the source n. In some applications, situations arise for which we cannot assign unambiguously TDOAs or DOAs to the same source. The example in Figure 1 shows the case of two sources with a configuration of two arrays for the 2D location. As we can see, the combination of incorrect DOAs leads to an incorrect position estimation. The two DOAs calculated by the two arrays can be combined following two different configurations: (1) $\widehat{\theta}_{11} - \widehat{\theta}_{21}, \widehat{\theta}_{12} - \widehat{\theta}_{22}$; (2) $\widehat{\theta}_{12} - \widehat{\theta}_{21}$, $\widehat{\theta}_{11} - \widehat{\theta}_{22}$. The first configuration implies the correct localization of the sound sources, whereas the second leads to an incorrect localization of both the sources.

In general, the goal is to get the matrix $\boldsymbol{\Theta}$ to properly order the values of (7). Considering θ_{rn} as the nth DOA of array r, the assignment of the correct value of the DOA for the unknown sources can be ambiguous; namely the exact position of the elements in the matrix of (6) cannot be uniquely determined:

$$\widehat{\theta}_{rn} \longrightarrow \theta_{rn}. \tag{8}$$

The possible combinations of the DOAs of matrix (7) are $O = (N!)^{(R-1)}$.

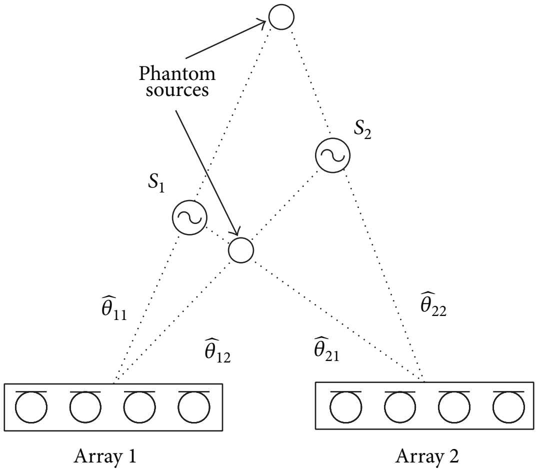

Figure 1: The problem of multiple sources localization.

4. Incident Signal Power Comparison (ISPC)

Incident signal power comparison (ISPC) combines the DOAs from different arrays by considering the similarity criterion among acoustic sources. To check for this similarity, we can estimate for each array the ISP referring to all estimated DOAs using beamforming techniques. Once the ISPs are obtained, we can define an efficient error criterion for comparing the different possible combinations of the DOAs using a SD measure of ISPs pair between different arrays.

DOA estimation is a crucial step of ASL systems. In a free-field environment for far-field cases, it can be calculated by means of MCCC and SRP methods. After obtaining an estimation of the sorted DOAs matrix (the matrix $\widehat{\underline{\boldsymbol{\Theta}}}$ of (7)), the steps of the ISPC algorithm are (1) source extraction using beamforming techniques and estimation of ISPs for each DOA, (2) ISPC using SD measurement between ISPs of different array, (3) calculation of all DOAs combinations, and (4) verification of the most consistent target combinations minimizing an error criterion on SD measurements. Finally, the localization of multiple sources can be computed by considering the estimated DOAs combination. Figure 2 illustrates the ISPC steps.

4.1. Incident Signal Power Estimation. The ISP is the power spectral density of the beamformer output that is steered to a specified direction. The SRP is based on maximizing the power output of a beamformer. Beamforming is a multichannel signal processing techniques that enhance the acoustic signals coming from a specific steered position, while reducing the signals coming from other directions. In the frequency domain, the output of a generic beamformer of rth array in matrix notation can be written as

$$Y_r(f) = \mathbf{W}^H(f, \theta_{rn})\, \mathbf{X}_r(f), \tag{9}$$

where $\mathbf{X}_r(f) = [X_{r1}(f), X_{r2}(f), \ldots, X_{rM}(f)]^T$, $Y_r(f)$ and $X_{rm}(f)$ are the discrete Fourier transform of the signals, $\mathbf{W}(f, \theta_{rn}) = [W_1(f, \theta_{rn}), W_2(f, \theta_{rn}), \ldots, W_M(f, \theta_{rn})]^T$ is the vector of the beamformer weights for steering and filtering the data on the direction θ_{rn}, f is the frequency bin index, and

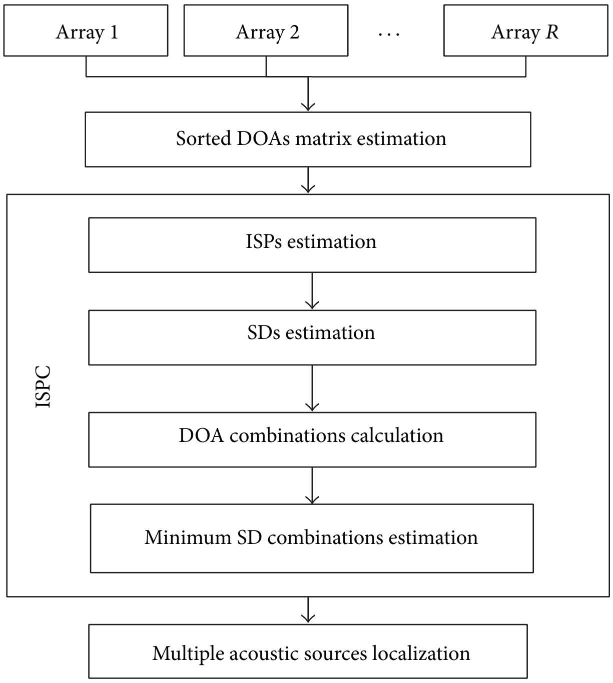
Array 1
Array 2
···
Array R
Sorted DOAs matrix estimation
ISPC
ISPs estimation
SDs estimation
DOA combinations calculation
Minimum SD combinations estimation
Multiple acoustic sources localization

Figure 2: The steps for the ISPC algorithm.

the superscript H represents the Hermitian (complex conjugate) transpose.

The ISP of the beamformer output for a generic frequency f is given by

$$\begin{aligned} \mathrm{ISP}(f) &= E\left\{\left|Y_r(f)\right|^2\right\} \\ &= \mathbf{W}^H(f, \theta_{rn})\, E\left\{\mathbf{X}_r(f)\, \mathbf{X}_r^H(f)\right\} \mathbf{W}(f, \theta_{rn}) \\ &= \mathbf{W}^H(f, \theta_{rn})\, \boldsymbol{\Phi}_r(f)\, \mathbf{W}(f, \theta_{rn}), \end{aligned} \tag{10}$$

where $\boldsymbol{\Phi}_r(f)$ is the cross-spectral density matrix, which is square $M \times M$ and symmetric, and $E\{\cdot\}$ denotes mathematical expectation. We consider a power spectrum calculated with $f = F_{\min}, F_{\min}+1, \ldots, F_{\max}$, where $F_{\min}$ and $F_{\max}$ are the index values of a specific frequency range (FR), which defines the range interesting for the optimal performance of the ISPC algorithm. Note that beamformer pattern function is frequency dependent; then the main lobe narrows with increasing frequency and spatial aliasing can occur (for a comprehensive dissertation, refer to [62]).

Several beamforming techniques exist (a review can be found in [63]); however, the spatial filter methods that are used for comparing ISPC experimental results are the SRP based on delay and sum beamforming, the SRP with the Dolph-Chebyshev window (SRP-DC), and the MVDR with DL.

Hence, the ISP corresponding to delay and sum SRP can be written from (10) as

$$\mathrm{ISP}_{rn}^{\mathrm{SRP}}(f) = \mathbf{A}^H(f, \theta_{rn})\, \boldsymbol{\Phi}_r(f)\, \mathbf{A}(f, \theta_{rn}), \tag{11}$$

where $\mathbf{A}(f,\theta_{rn})$ is the steering vector corresponding to direction θ_{rn}. For a uniform linear array with microphone distance d, the steering vector takes the form

$$\begin{aligned}\mathbf{A}(f,\theta_{rn}) = \Big[1, e^{j2\pi(f-1)df_s \sin\theta_{rn}/c}, \dots, \\ e^{(j2\pi(f-1)df_s \sin\theta_{rn}/c)(M-1)}\Big]^T.\end{aligned} \tag{12}$$

The SRP-DC is obtained from (11) introducing the Dolph-Chebyshev window $\mathbf{w}$:

$$\mathrm{ISP}_{rn}^{\text{SRP-DC}}(f) = \left[\mathbf{w} \odot \mathbf{A}(f,\theta_{rn})\right]^H \boldsymbol{\Phi}_r(f)\left[\mathbf{w} \odot \mathbf{A}(f,\theta_{rn})\right], \tag{13}$$

where $\odot$ denotes element-wise multiplication.

The adaptive MVDR beamforming [27] is based on minimization problem of the following equation

$$\begin{aligned}&\underset{\mathbf{W}(f,\theta_{rn})}{\operatorname{argmin}}\, \mathbf{W}^H(f,\theta_{rn})\boldsymbol{\Phi}_r(f)\mathbf{W}(f,\theta_{rn}) \\ &\text{subject to} \quad \mathbf{W}^H(f,\theta_{rn})\mathbf{A}(f,\theta_{rn}) = 1.\end{aligned} \tag{14}$$

The aim of the MVDR filter is to minimize the noise and sources coming from different directions, while keeping a fixed gain on the desired direction. Solving (14) using the method of Lagrange multipliers, we can write

$$\mathbf{W}_{\text{MVDR}}(f,\theta_{rn}) = \frac{\boldsymbol{\Phi}_r^{-1}(f)\mathbf{A}(f,\theta_{rn})}{\mathbf{A}^H(f,\theta_{rn})\boldsymbol{\Phi}_r^{-1}(f)\mathbf{A}(f,\theta_{rn})}. \tag{15}$$

In practical applications, the inverse of the cross-spectral density matrix can be calculated using the Moore-Penrose pseudoinverse, defined as

$$\boldsymbol{\Gamma}^+ = \mathbf{V}\mathbf{S}^{-1}\mathbf{U}^H, \tag{16}$$

where $\boldsymbol{\Gamma} = \mathbf{U}\mathbf{S}\mathbf{V}^H$ is the singular value decomposition of the matrix $\boldsymbol{\Gamma}$. Moreover, if the cross-spectral density matrix is ill-conditioned, the spatial spectrum may not exist. Therefore, a DL [60, 61] method is adopted to calculate the inverse matrix in a stable way. The ISP with MVDR filter and DL becomes

$$\mathrm{ISP}_{rn}^{\text{MVDR}}(f) = \frac{1}{\mathbf{A}^H(f,\theta_{rn})\left(\boldsymbol{\Phi}_r(f) + \mu\mathbf{I}\right)^+\mathbf{A}(f,\theta_{rn})}, \tag{17}$$

where $\mathbf{I}$ is the identity matrix and μ is the loading level:

$$\mu = \frac{1}{L}\operatorname{tr}\left\{\boldsymbol{\Phi}_r(f)\right\}\Delta, \tag{18}$$

where $\operatorname{tr}\{\cdot\}$ denotes the trace of the squared matrix and Δ is the normalized loading constant. Typically, the values are $\Delta = 0.1$, $\Delta = 1$, $\Delta = 10$ [64].

Therefore, we can define the matrix $\mathbf{P}$ containing all the ISPs related to the matrix (7):

$$\begin{aligned}\mathbf{P} = \big[&\mathbf{P}_{11}, \mathbf{P}_{12}, \dots, \mathbf{P}_{1N}, \mathbf{P}_{21}, \mathbf{P}_{22}, \dots, \mathbf{P}_{2N}, \\ &\mathbf{P}_{R1}, \mathbf{P}_{R2}, \dots, \mathbf{P}_{RN}\big]\end{aligned} \tag{19}$$

which has a dimension of $(F_{\max} - F_{\min}) \times RN$, where the total number of ISPs is $I = NR$ and $\mathbf{P}_{rn} = [\mathrm{ISP}_{rn}(F_{\min}), \mathrm{ISP}_{rn}(F_{\min}+1), \dots, \mathrm{ISP}_{rn}(F_{\max})]^T$.

4.2. Spectral Distance Estimation. To compare the ISPs of different arrays, spectral distance (SD) functions are used. Distance measures produce measurements of the dissimilarity of two sound spectra. We define the SD estimation between the ISP_{rn} and the ISP_{ij} of two DOAs of different arrays as

$$E_{rnij} = \frac{1}{L}\sum_{f=F_{\min}}^{F_{\max}} \left|\mathcal{S}\left\{\mathrm{ISP}_{rn}(f), \mathrm{ISP}_{ij}(f)\right\}\right|, \tag{20}$$

where $L = (F_{\max} - F_{\min} + 1)$, r and i are the index labels of the array, $r \neq i$, n and j are the index labels for the sorted DOAs of array, and $\mathcal{S}\{\mathrm{ISP}_{rn}(f), \mathrm{ISP}_{ij}(f)\}$ is the SD function to measure the dissimilarity of spectra. We consider the four most common SD functions to verify how our system performance varies as a function of different equations. A classic spectral estimation method is linear prediction (LP) [65], for which we insert a negative one to standardize the minimum to zero as all functions

$$E_{rnij}^{\text{LP}} = \frac{1}{L}\sum_{f=F_{\min}}^{F_{\max}} \left|\frac{\mathrm{ISP}_{rn}(f)}{\mathrm{ISP}_{ij}(f)} - 1\right|. \tag{21}$$

The other functions are the Itakura-Saito (IS) distance measure [66]

$$E_{rnij}^{\text{IS}} = \frac{1}{L}\sum_{f=F_{\min}}^{F_{\max}} \left|\frac{\mathrm{ISP}_{rn}(f)}{\mathrm{ISP}_{ij}(f)} - \log\frac{\mathrm{ISP}_{rn}(f)}{\mathrm{ISP}_{ij}(f)} - 1\right|, \tag{22}$$

the root mean square (RMS) log [67]

$$E_{rnij}^{\text{RMS}} = \frac{1}{L}\sum_{f=F_{\min}}^{F_{\max}} \left(\log\frac{\mathrm{ISP}_{rn}(f)}{\mathrm{ISP}_{ij}(f)}\right)^2, \tag{23}$$

and the COSH measure [68]

$$\begin{aligned}E_{rnij}^{\text{COSH}} = \frac{1}{L}\sum_{f=F_{\min}}^{F_{\max}} \Bigg|&\frac{\mathrm{ISP}_{rn}(f)}{\mathrm{ISP}_{ij}(f)} - \log\frac{\mathrm{ISP}_{rn}(f)}{\mathrm{ISP}_{ij}(f)} \\ &+ \frac{\mathrm{ISP}_{ij}(f)}{\mathrm{ISP}_{rn}(f)} - \log\frac{\mathrm{ISP}_{ij}(f)}{\mathrm{ISP}_{rn}(f)} - 2\Bigg|.\end{aligned} \tag{24}$$

The total number of SD measures between all the ISPs pair of different arrays is $Q = N^2R(R-1)/2$.

4.3. DOA Combinations Calculation. Let us represent the sorted matrix of the DOAs using the graph theory to better understand the DOAs combinations calculation and the verification of the most consistent target combination minimizing an error criterion. Then, we can express the matrix (7) and all of its combinations as being composed of nodes and edges, connecting pairs of vertices. An example of three arrays and three sources is shown in Figure 3. Each row of the graph contains the sorted DOAs of an array: $\underline{\widehat{\boldsymbol{\Theta}}}_1 = [\widehat{\theta}_{11}, \widehat{\theta}_{12}, \widehat{\theta}_{13}]^T$, $\underline{\widehat{\boldsymbol{\Theta}}}_2 = [\widehat{\theta}_{21}, \widehat{\theta}_{22}, \widehat{\theta}_{23}]^T$, and $\underline{\widehat{\boldsymbol{\Theta}}}_3 = [\widehat{\theta}_{31}, \widehat{\theta}_{32}, \dots, \widehat{\theta}_{3N}]^T$. Each DOA is a node of graph and the edges represent the possible connections between nodes with the values E_{rnij}, which is the estimated SD between the ISPs on array r of DOA i and on array n of DOA j. The combination of incorrect

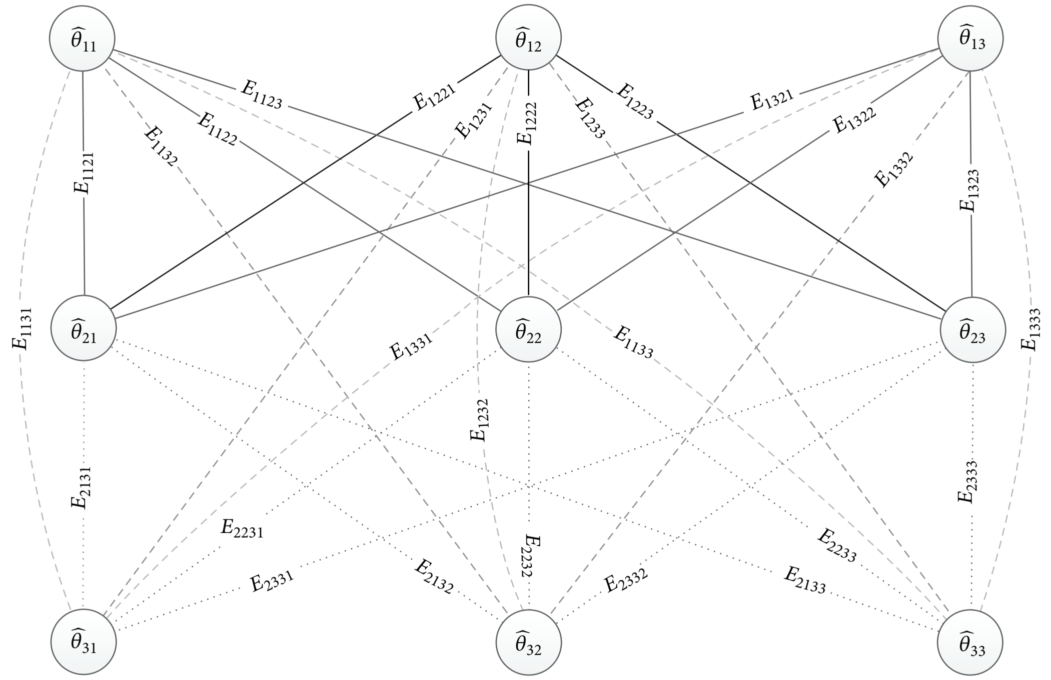

Figure 3: Graphic representation of DOAs and SDs estimations.

DOAs leads to an incorrect position estimation (see Figure 1). Thus, if we represent a combination of DOAs as a sum of values of the edges that connect the nodes, we expect that the minimum value of different sums corresponds to the correct combination. To calculate the possible combinations of DOAs between the arrays, it is helpful to introduce a matrix labeling of DOAs (7):

$$\mathbf{B} = \begin{bmatrix} B_{11} & B_{12} & \dots & B_{1N} \\ B_{21} & B_{22} & \dots & B_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ B_{R1} & B_{R2} & \dots & B_{RN} \end{bmatrix} \tag{25}$$

in which the generic element is expressed as

$$B_{rn} = (r-1)\,N + n \tag{26}$$

with $r = 1, 2, \dots, R$ and $n = 1, 2, \dots, N$. The matrix label $\mathbf{B}$ associates the position of the DOAs referring to the sorted matrix $\widehat{\boldsymbol{\Theta}}$. Estimating the minimum error of an SD combination, we can obtain the matrix $\widehat{\boldsymbol{\Theta}}$ with the correct position of the DOAs, in which each column contains the DOAs of the source n.

Furthermore, we can represent the graph representation of DOAs and the SDs as the adjacency matrix $\boldsymbol{\Lambda}$, which is an $RN \times RN$ matrix of SD values. The entry in row ($B_{rn} = 1, \dots, RN$) and column ($B_{ij} = 1, \dots, RN$) is defined as an SD estimation E_{rnij} if there is an edge connecting vertex B_{rn} and vertex B_{ij} in the graph, or it is defined as zero otherwise. The relationships between DOAs and SDs can be expressed by the following equation of the adjacency matrix element:

$$\Lambda_{B_{rn}B_{ij}} = E_{rnij}. \tag{27}$$

The symmetric adjacency matrix results in the following equation:

$$\boldsymbol{\Lambda} = \begin{bmatrix} 0 & \dots & 0 & E_{1121} & \dots & E_{112N} & \dots & E_{11R1} & \dots & E_{11RN} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & E_{1N21} & \dots & E_{1N2N} & \dots & E_{1NR1} & \dots & E_{1NRN} \\ E_{2111} & \dots & E_{211N} & 0 & \dots & 0 & \dots & E_{21R1} & \dots & E_{21RN} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ E_{2N11} & \dots & E_{2N1N} & 0 & \dots & 0 & \dots & E_{2NR1} & \dots & E_{2NRN} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ E_{R111} & \dots & E_{R11N} & E_{R121} & \dots & E_{1121} & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ E_{RN11} & \dots & E_{RN1N} & E_{RN21} & \dots & E_{RN2N} & \dots & 0 & \dots & 0 \end{bmatrix}. \tag{28}$$

These SD values are weights of the edges of the graph. An example of three arrays and three sources is presented in Figure 3; in this example, we have 27 total SD comparisons (3 for each source).

To calculate all possible combinations of DOAs, we can work on the label matrix **B**. Considering that the first row of **B** related to the first array remains unchanged, we can compute the combinations in two steps. In the first step, the permutations of the N labels of **B** for each $R-1$ row (for each array) are calculated. The number of permutations for each row is $U = N!$. Thus, we define the permutation matrix T as

$$\begin{aligned}\mathbf{T} = [&\mathbf{T}_1, \mathbf{T}_{21}, \mathbf{T}_{22}, \dots, \mathbf{T}_{2U}, \mathbf{T}_{31}, \mathbf{T}_{32}, \dots, \mathbf{T}_{3U}, \dots, \\ &\mathbf{T}_{R1}, \mathbf{T}_{R2}, \dots, \mathbf{T}_{RU}],\end{aligned} \tag{29}$$

where

$$\mathbf{T}_1 = [B_{11}, B_{12}, \dots, B_{1N}]^T, \tag{30}$$

$$\mathbf{T}_{ru} = \mathscr{P}_u \left\{[B_{r1}, B_{r2}, \dots, B_{rN}]^T\right\}, \tag{31}$$

where $\mathbf{T}_{ru}$ is the vector of uth permutation $\mathscr{P}_u$ $(u = 1, \dots, U)$ of rth array $(r = 2, 3, \dots, R)$, which contains the N DOAs label of row r. The matrix **T** has a dimension of $N \times U(R - 1) + 1$. In the second step, the combinations of column indices of matrix **T** with value from 2 to RU give the $U^{(R-1)} = N!^{(R-1)} = O$ possible combinations. We consider the combinations of $R-1$ groups, each one composed by U elements (the permutation), assuming that one member (the index column of matrix **T**) from each of the $R-1$ groups is used in each combination and assuming that the order is not a distinguishing factor. We define a matrix **Z** of dimension $O \times (R-1)$, which stores the combinations of groups of column indices of matrix **T**:

$$\mathbf{Z} = \begin{bmatrix} Z_{11} & Z_{12} & \dots & Z_{1(R-1)} \\ Z_{21} & Z_{22} & \dots & Z_{2(R-1)} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{O1} & Z_{O2} & \dots & Z_{O(R-1)} \end{bmatrix}. \tag{32}$$

The generic element Z_{or} with $o = 1, 2, \dots, O$ and $r = 1, 2, \dots, R-1$ can be calculated with the following equations:

$$\begin{aligned} Z_{(o+i-1)r} &= U(r-1) + 2, \quad i = 1, 2, \dots, U_1, \\ Z_{(o+U_1+1)r} &= \begin{cases} Z_{(o+U_1+1)r}, & \text{if } Z_{(U_1+1)r} > U_2, \\ Z_{(o+U_1+1)r} + 1, & \text{otherwise,} \end{cases} \end{aligned} \tag{33}$$

where $U_1 = U^{(r-1)}$ and $U_2 = U(r-1) + U + 1$.

Hence, a combination label matrix **C** of $I \times O$ dimension is used to store the DOA label of all combinations:

$$\mathbf{C} = [\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_O], \tag{34}$$

where $\mathbf{C}_o$ is the vector, which contains the I DOA labels of combination o:

$$\begin{aligned} \mathbf{C}_o = [&B_{11}, T_{Z_{o1}1}, T_{Z_{o2}1} \dots, T_{Z_{o(R-1)}1}, \\ &B_{12}, T_{Z_{o1}2}, T_{Z_{o2}2}, \dots, T_{Z_{o(R-1)}2}, \dots, \\ &B_{1N}, T_{Z_{o1}N}, T_{Z_{o2}N}, \dots, T_{Z_{o(R-1)}N}]^T \\ = [&C_{1o}, C_{2o}, \dots, C_{Io}]^T. \end{aligned} \tag{35}$$

4.4. Minimum SD Measure Estimator. For each source, identified by R nodes (the arrays), we have $R(R-1)/2$ edges; then, the number of edges for a combination of DOAs is $G = NR(R-1)/2 = Q/N$. The values of matrix **C** are used to calculate the SD estimation of all combinations. Thus, we can define the SD estimation of the generic combination o as the sum of the weights of all the edges:

$$D_o = \sum_{n=1}^{N} \sum_{r=1}^{R-1} \sum_{i=1}^{R-r} \Lambda_{C_{o_1 o} C_{o_2 o}}, \tag{36}$$

where $o = 1, 2, \dots, O$, $o_1 = (n-1)R + r$ and $o_2 = (n-1)R + r + i$. Accordingly, we define the SD vector of all combinations

$$\mathbf{D} = [D_1, D_2, \dots, D_O]^T. \tag{37}$$

Finally, the index of the minimum value of the vector **D** identifies the target combination as

$$\hat{o} = \underset{o}{\operatorname{argmin}}\, D_o, \tag{38}$$

and the DOAs matrix $\widehat{\boldsymbol{\Theta}}$ is estimated by ordering the label matrix **B** with the combination $\mathbf{C}_{\hat{o}}$.

4.5. Overall Procedure. The processing steps of the full ISPC algorithm are summarized in Algorithm 1. After the DOAs estimation and creation of the matrix $\underline{\widehat{\boldsymbol{\Theta}}}$ defined by (7) the ISPC algorithm is applied if multiple sources are detected. In practice, the matrix does not always present all the DOA values. In these cases, the missing value of array r can be represented with a zero value in the label matrix **B** (25). The overall procedure is composed by the following steps: (1) building of the label matrix **B** (25) and calculation of ISPs and the matrix **P** (19); (2) estimation of the SD measurements between all ISP pairs of arrays and creation of the adjacency matrix $\boldsymbol{\Lambda}$ (28); (3) calculation of the permutations matrix **T** (29) and the all DOA combination matrix **C** (34); (4) calculation of the vector **D** (37) that contains the SD estimation for each DOAs combination and finding the minimum value of **D** (38), for using the index value $\hat{o}$ in the matrix **C** to properly order the matrix $\underline{\widehat{\boldsymbol{\Theta}}}$ and estimate the matrix $\widehat{\boldsymbol{\Theta}}$.

5. Experimental Results

The experimental results were conducted in an outdoor real noisy environment and the ISPC performance is reported using different beamforming techniques (SRP, SRP-DC, MVDR) and SD estimations (LP, IS, RMS, COSH). A prototype system for two-dimensional localization has been

```
Require: $N > 1$
  $I = NR$  {ISPs}
  $O = N!^{(R-1)}$  {DOA combinations}
  $Q = N^2 R(R-1)/2$  {SDs}
  $U = N!$  {DOA label permutations for an array}
  for $r = 1$ to $R$ do
    for $n = 1$ to $N$ do
      Calculate (26)
      Calculate (10) with (11) (13) (17)
    end for
  end for
  Calculate (25)
  Calculate (19)
  for $r = 1$ to $R - 1$ do
    for $n = r + 1$ to $R$ do
      for $i = 1$ to $N$ do
        for $j = 1$ to $N$ do
          Calculate (20) with (21) (22) (23) (24)
          Calculate (27)
        end for
      end for
    end for
  end for
  Calculate (28)
  for $r = 2$ to $R$ do
    for $u = 1$ to $U$ do
      Calculate (31)
    end for
  end for
  Calculate (29)
  for $r = 1$ to $R - 1$ do
    while $o < O + 1$ do
      Calculate (33)
    end while
  end for
Calculate (32)
Calculate (34)
Calculate (36)
Calculate (38)
```

Algorithm 1: ISPC.

installed on the roof of the building that houses the Computer Science Department in Udine University (Figure 4).

5.1. System Setup. The acoustic localization prototype includes two linear arrays, each composed of four omnidirectional microphones. Very small sized arrays are used because a real application of such systems would require that the public spaces are not invaded in an excessive way; therefore, there might not be enough space to install the arrays. The arrays are located 11.4 m apart at a height of 12.1 m above the plane. The sample rate of the digital system is 48 kHz, and the microphone distance is 25 cm. The system consists of two parallel processing lines, corresponding to the Array 1 and Array 2 (Figure 5).

The first processing step is the DOAs estimation. SRP-PHAT is used for the DOAs estimation. The values corresponding to the principal N peaks of the SRP-PHAT function (in practice, those peaks which are above a given threshold) allow the DOAs estimation of the N acoustic sources. The assumed DOA range is -90° $+90^\circ$, where zero is in front of the array and the microphone reference is the first from left.

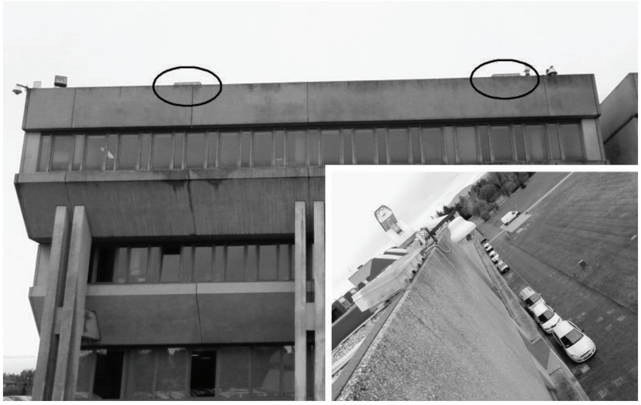

Figure 4: The prototype installed on the roof of the University building. The two arrays are encircled.

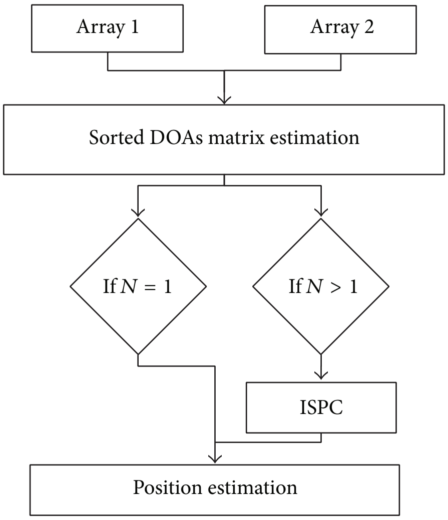

Figure 5: The block diagram of the processor showing the data flow of all of the tasks of the experimental prototype.

In the second step, the two-dimensional coordinates of the source can be estimated by combining the DOAs at the arrays. If more than one source is identified, a beamformer and an SD comparison provide a guide to solve the problem of associating the DOAs of the Array 1 with those of the Array 2. The calculation of the two-dimensional position of the source is a simple triangulation problem. However, we must consider that the two arrays are not coincidental with the plane of interest but are placed at a certain height. We must consider that the possible points identified by the DOA are located on a cone surface whose vertex is placed in the array and whose axis is the straight line joining the two arrays. Every array represents a cone: the intersection of the two cones is represented by a circumference. The intersection point between the circumference and the plane of interest is the estimation of the source distance from arrays (see Figure 6). Hence, we consider d_a to be the distance of the arrays, h to be the height

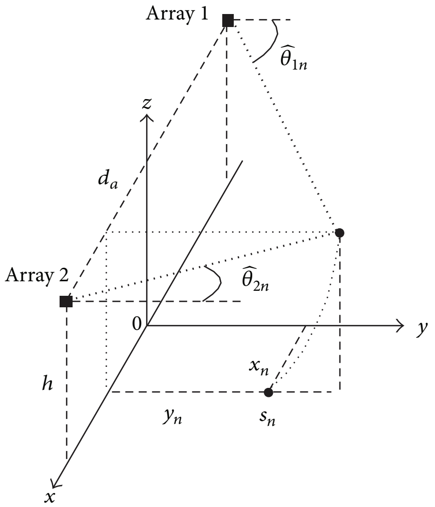

FIGURE 6: The two-dimensional position of the source of the experimental prototype.

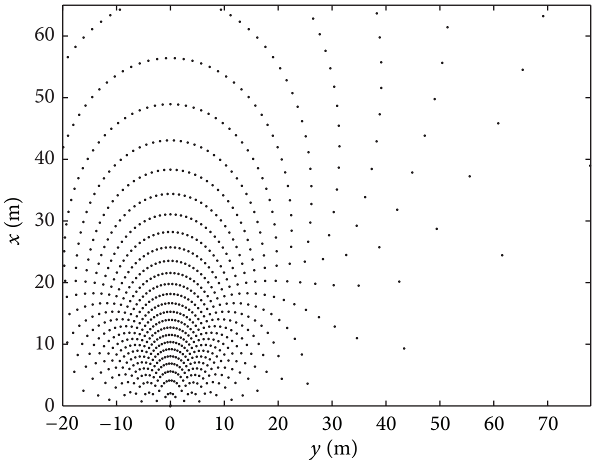

FIGURE 7: The x-y sample space position of the plane of interest.

of arrays above the plane of interest, and $\widehat{\theta}_{1n}$ and $\widehat{\theta}_{2n}$ to be the DOA estimated on the Array 1 and Array 2, and we obtain

$$x_n = \frac{d_a}{2}\left(\frac{\tan\widehat{\theta}_{2n} + \tan\widehat{\theta}_{1n}}{\tan\widehat{\theta}_{2n} - \tan\widehat{\theta}_{1n}}\right),$$
$$y_n = \sqrt{\left(\frac{d_a}{\tan\widehat{\theta}_{2n} - \tan\widehat{\theta}_{1n}}\right)^2 - h^2}. \tag{39}$$

The spatial resolution of the system depends on the distance between the microphones, the distance between the arrays, and the sample frequency of digital system. Figure 7 shows the possible xy coordinates of the considered area of analysis. The zero of the xy axes reference is located in the middle of the distance between the two arrays. The spatial resolution tends to decrease with an increasing distance from the arrays and an increasing angle from the axis perpendicular to the array.

TABLE 1: Position reporting the source label for each test and the SPL of the sources.

Test label	s_1 (Voice)	s_2 (Hammer)	s_3 (Motor car)	s_4 (Honk)
	70 dB(A)	100 dB(A)	68 dB(A)	88 dB(A)
p_1	1	2	3	—
p_2	1	3	4	—
p_3	1	4	5	—
p_4	1	5	6	—
p_5	1	7	8	—
p_6	1	8	9	—
p_7	1	9	10	—
p_8	1	10	11	—
p_9	2	3	1	—
p_{10}	3	4	1	—
p_{11}	4	5	1	—
p_{12}	5	6	1	—
p_{13}	7	7	1	—
p_{14}	8	8	1	—
p_{15}	9	10	1	—
p_{16}	10	11	1	—
p_{17}	10	12	—	13
p_{18}	10	13	—	14
p_{19}	10	14	—	15
p_{20}	10	15	—	19
p_{21}	6	16	—	17
p_{22}	6	17	—	18
p_{23}	6	18	—	19
p_{24}	6	19	—	20
p_{25}	12	13	—	10
p_{26}	13	14	—	10
p_{27}	14	15	—	10
p_{28}	15	18	—	10
p_{29}	16	17	—	6
p_{30}	17	18	—	6
p_{31}	18	19	—	6
p_{32}	19	20	—	6

5.2. Experiment Setup. Experiments were conducted that consider the area of analysis of 60 × 90 m shown in Figure 8, that is, the parking lots of the University. Twenty zones of acoustic source positioning are considered. They are labeled with a number as we can see in Figure 8. The sources used are a human voice (s_1), a hammer striking an iron bar (s_2), a motor car (s_3), and a honk car (s_4). The hammer striking an iron bar and the honk car are short-duration event sounds.

Two types of experiments were performed. The first type used sounds with different spectral content, named Test 1. The second type, instead, used sounds with similar spectral content, named Test 2. Test 1 is composed of thirty-two parts ($p_1, p_2, \ldots, p_{32}$), each one with three sources placed in different positions (see Table 1). In various parts of Test 1, the sources were positioned at increasing distances along the y axis and the x axis. Table 1 also reported the sound pressure level (SPL) of each source. The environmental noise was in

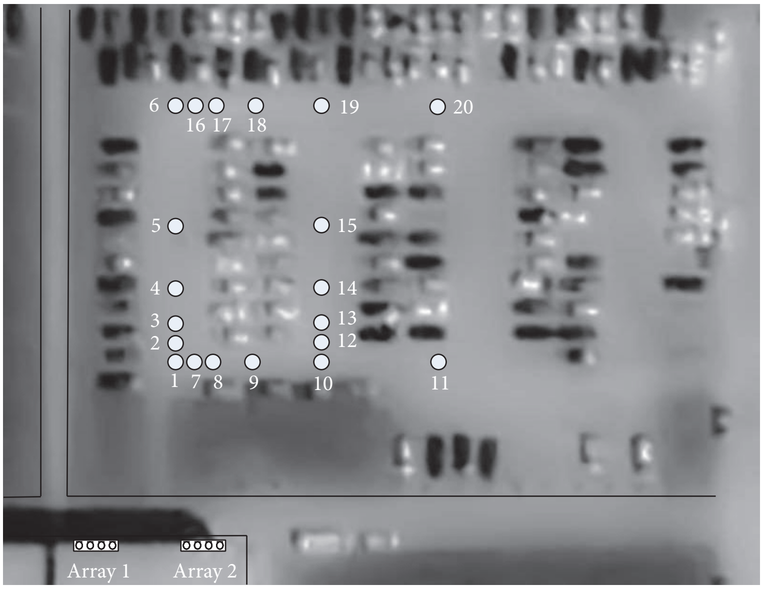

Figure 8: Map of the study area indicating the position of arrays (Array 1 and Array 2) and sources (the twenty labeling numbers: 1, 2, . . . , 20).

a range of 40–50 dB(A). In Test 2, two car sounds were used. The test was performed by placing two car sources in 1 and 7, as shown in Figure 8.

5.3. System Localization Evaluation. An evaluation of the system localization using a single source for each position was computed. Table 2 shows the real xy coordinates of the source points and the root mean square (RMS) error of the estimation using SRP-PHAT method. We can see that the estimation error increases in distant areas and when the angle of incidence on the array is large.

5.4. ISPC Evaluation. Tables 3 and 4 summarize the results for Test 1 and Test 2, respectively, comparing the localization success rate (as a percentage) with different beamforming algorithms (SRP, SRP-DC, and MVDR) and SD functions (LP, IS, RMS, and COSH).

The localization success rate is the ratio between the number of correct combinations and the number of ambiguities (NOA). NOA is the number of frames in which we have ambiguity to properly associate the DOAs to the sources; that is, the associations are incorrect in practice. The audio signal frame was divided into 17.5 ms overlapping and a Hann-windowed with a length of 140 ms. The parking area, where the tests were conducted, is a public area. Thus, we must consider that there are other sources in the acoustic scene: other sounds of cars that are moving in the parking area and in the nearby streets.

Table 3 summarizes the results of all thirty-two tests (Test 1). The number of NOA is 750, and the three frequency ranges (FR) for the SD estimation are 20–675 Hz, 20–2000 Hz, and 20–8000 Hz. The frequency value of 675 Hz takes into account the spatial aliasing limit, which, in our case, is $f = c/(2d) = 337/(2 \cdot 0.25) = 675$ Hz. The phenomena of spatial aliasing implies that the main lobe of the beamformer has a set of identical copies, called grating lobes. The appearance of grating lobes is a function of both microphone spacing and incident frequency. When fully visible, a grating lobe is equal in amplitude to the main lobe of the array. This fact reduces the array response, and, therefore, by defining the spatial sampling requirement and removing the grating lobes, we obtain a greater efficiency in the ISPC.

Table 2: Position referring to Figure 8 and the RMS errors of the localization estimation using SRP-PHAT method.

Source label	x (m)	y (m)	RMS error
1	1.5	20	1.1
2	1.5	23	1.7
3	1.5	26	2.2
4	1.5	32	1.9
5	1.5	38	2.5
6	1.5	52	4.1
7	4.5	20	0.9
8	7.5	20	4.2
9	10.5	20	8.6
10	20	20	18.2
11	30	20	8.8
12	20	23	9.2
13	20	26	15.8
14	20	32	15.7
15	20	38	4.8
16	4.5	52	4.2
17	7.5	52	8.3
18	10.5	52	17.5
19	20	52	20.2
20	30	52	23.6

Table 4 depicts the results of Test 2 with an FR of 20–675 Hz and a NAM of 100. We can note that the accuracy decreases, especially with regard to the RMS and COSH

Table 3: Results of Test 1 reporting the summary of the thirty-two tests ($p_1, p_2, \ldots, p_{32}$).

(Hz)	Localization success rate (%)											
FR	SRP-LP	SRP-IS	SRP-RMS	SRP-COSH	SRP-DC-LP	SRP-DC-IS	SRP-DC-RMS	SRP-DC-COSH	MVDR-LP	MVDR-IS	MVDR-RMS	MVDR-COSH
20–675	28.4	38.2	85.2	79.1	23.4	37.8	86.4	73.7	43.6	68.6	90.1	83.4
20–2000	42.1	45.2	72.3	63.5	43.8	42.4	71.2	61.1	45.4	63.1	77.6	72.1
20–8000	53.2	52.4	69.5	58.5	51.9	49.4	68.3	58.7	48.4	56.0	65.5	61.8

Table 4: Results of Test 2 using two car sounds.

(Hz)	Localization success rate (%)											
FR	SRP-LP	SRP-IS	SRP-RMS	SRP-COSH	SRP-DC-LP	SRP-DC-IS	SRP-DC-RMS	SRP-DC-COSH	MVDR-LP	MVDR-IS	MVDR-RMS	MVDR-COSH
20–675	50.2	53.8	57.5	47.1	52.4	51.3	59.4	55.7	45.4	62.0	52.0	53.5

functions, and this result highlights the limitation of the proposed approach in the case of spectrally similar sources.

The best results for Test 1 were obtained with the RMS log SD function and FR = [20–675] Hz. MVDR has the greatest capacity for location with 90.1% of successful DOAs association.

6. Conclusions

The novel incident signal power comparison algorithm is used to solve the ambiguous problem of correctly linking the DOAs from different arrays to the same source in a far-field condition with concurrent sources. Experimental results have shown that this approach can be a solution for a multisource localization that requires a frame-to-frame analysis, that is, in those cases in which the traditional filtering approach can not be applied. An evaluation of the system in a real scenario is reported, installing a hardware/software prototype on the roof of the University building and analyzing the results comparing three types of beamforming and four functions for the SD estimation. The interest in locating in a far-field outdoor context may be attractive for audio surveillance, sound monitoring, and the analysis of acoustic scenes. The ISPC is successfully used in a joint audio-video system for monitoring a large area. The best performances are obtained with RMS SD measure on frequency range between 20 Hz and the spatial aliasing frequency limit. We achieved a success rate of 90.1% using MVDR beamforming. We showed the limitation of the proposed algorithm in case of sources that have a similar spectral content.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

The authors would like to thank Rudolf Rabenstein (Professor, University of Erlangen-Nurnberg, Germany) for his valuable advice and helpful comments.

References

[1] R. O. Schmidt, "A new approach to geometry of range difference location," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 8, no. 6, pp. 821–835, 1972.

[2] H. C. Schau and A. Z. Robinson, "Passive source localization employing intersecting spherical surfaces from time-of-arrival differences," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 8, pp. 1223–1225, 1987.

[3] J. O. Smith and J. S. Abel, "Closed-form least-squares source location estimation from range-difference measurements," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 12, pp. 1661–1669, 1987.

[4] Y. T. Chan and K. C. Ho, "A simple and efficient estimator for hyperbolic location," *IEEE Transactions on Signal Processing*, vol. 42, no. 8, pp. 1905–1915, 1994.

[5] M. S. Brandstein, J. E. Adcock, and H. F. Silverman, "A closed-form location estimator for use with room environment microphone arrays," *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 1, pp. 45–50, 1997.

[6] Y. Huang, J. Benesty, G. W. Elko, and R. M. Mersereau, "Real-time passive source localization: a practical linear-correction least-squares approach," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 8, pp. 943–956, 2001.

[7] M. D. Gillette and H. F. Silverman, "A linear closed-form algorithm for source localization from time-differences of arrival," *IEEE Signal Processing Letters*, vol. 15, pp. 1–4, 2008.

[8] P. Stoica and J. Li, "Source localization from range-difference measurements," *IEEE Signal Processing Magazine*, vol. 23, no. 3, pp. 63–66, 2006.

[9] W. R. Hahn and S. A. Tretter, "Optimum processing for delay-vector estimation in passive signal arrays," *IEEE Transactions on Information Theory*, vol. 19, no. 5, pp. 608–614, 1973.

[10] M. Wax and T. Kailath, "Optimum localization of multiple sources by passive arrays," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 31, no. 5, pp. 1210–1217, 1983.

[11] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and cramer-rao bound: further results and comparisons," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 12, pp. 2140–2150, 1990.

[12] M. Segal, E. Weinstein, and B. R. Musicus, "Estimate-maximize algorithms for multichannel time delay and signal estimation,"

IEEE Transactions on Signal Processing, vol. 39, no. 1, pp. 1–16, 1991.

[13] J. C. Chen, R. E. Hudson, and K. Yao, "Maximum-likelihood source localization and unknown sensor location estimation for wideband signals in the near-field," *IEEE Transactions on Signal Processing*, vol. 50, no. 8, pp. 1843–1854, 2002.

[14] P. G. Georgiou and C. Kyriakakis, "Robust maximum likelihood source localization: the case for sub-gaussian versus gaussian," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 4, pp. 1470–1480, 2006.

[15] G. Destino and G. Abreu, "On the maximum likelihood approach for source and network localization," *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 4954–4970, 2011.

[16] P. Aarabi, "The fusion of distributed microphone arrays for sound localization," *EURASIP Journal on Applied Signal Processing*, vol. 2003, no. 4, pp. 338–347, 2003.

[17] M. Omologo and P. S. R. DeMori, "Acoustic transduction," in *Spoken Dialogue with Computers*, Academic Press, London, UK, 1998.

[18] J. H. DiBiase, H. F. Silverman, and M. S. Brandstein, "Robust localization in reverberant rooms," in *Microphone Arrays: Signal Processing Techniques and Applications*, Digital Signal Processing, pp. 157–180, Springer, Berlin, Germany, 2001.

[19] D. B. Ward, E. A. Lehmann, and R. C. Williamson, "Particle filtering algorithms for tracking an acoustic source in a reverberant environment," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 6, pp. 826–836, 2003.

[20] P. Pertilä, T. Korhonen, and A. Visa, "Measurement combination for acoustic source localization in a room environment," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2008, Article ID 278185, pp. 1–14, 2008.

[21] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 4, pp. 320–327, 1976.

[22] J. Benesty, "Adaptive eigenvalue decomposition algorithm for passive acoustic source localization," *The Journal of the Acoustical Society of America*, vol. 107, no. 1, pp. 384–391, 2000.

[23] Y. Huang and J. Benesty, "Adaptive multichannel time delay estimation based on blind system identification for acoustic source localization," in *Adaptive Signal Processing: Applications to Real-World Problems*, Signals and Communication Technology, pp. 227–247, Springer, Berlin, Germany, 2003.

[24] Y. Huang and J. Benesty, "A class of frequency-domain adaptive approaches to blind multichannel identification," *IEEE Transactions on Signal Processing*, vol. 51, no. 1, pp. 11–24, 2003.

[25] D. Salvati and S. Canazza, "Adaptive time delay estimation using filter length constraints for source localization in reverberant acoustic environments," *IEEE Signal Processing Letters*, vol. 20, no. 5, pp. 507–510, 2013.

[26] M. S. Bartlett, "Smoothing periodograms from time-series with continuous spectra," *Nature*, vol. 161, no. 4096, pp. 686–687, 1948.

[27] J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.

[28] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," in *Proceedings of the RADC Spectrum Estimation Workshop*, pp. 243–258, Rome, NY, USA, October 1979.

[29] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[30] A. Paulraj, R. Roy, and T. Kailath, "A subspace rotation approach to signal parameter estimation," *Proceedings of the IEEE*, vol. 74, no. 7, pp. 1044–1046, 1986.

[31] R. Roy, A. Paulraj, and T. Kailath, "ESPRIT—a subspace rotation approach to estimation of parameters of cisoids in noise," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 5, pp. 1340–1342, 1986.

[32] R. Roy and T. Kailath, "ESPRIT—estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984–995, 1989.

[33] J. Chen, J. Benesty, and Y. Huang, "Robust time delay estimation exploiting redundancy among multiple microphones," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 6, pp. 549–557, 2003.

[34] J. Benesty, J. Chen, and Y. Huang, "Time-delay estimation via linear interpolation and cross correlation," *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 5, pp. 509–519, 2004.

[35] J. Benesty, Y. Huang, and J. Chen, "Time delay estimation via minimum entropy," *IEEE Signal Processing Letters*, vol. 14, no. 3, pp. 157–160, 2007.

[36] F. Wen and Q. Wan, "Robust time delay estimation for speech signals using information theory: a comparison study," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2011, article 3, pp. 1–10, 2011.

[37] H. Sawada, R. Mukai, and S. Makino, "Direction of arrival estimation for multiple source signals using independent component analysis," in *Proceedings of the 7th International Symposium on Signal Processing and Its Applications*, vol. 2, pp. 411–414, Paris, France, July 2003.

[38] B. Loesch, S. Uhlich, and B. Yang, "Multidimensional localization of multiple sound sources using frequency domain ICA and an extended state coherence transform," in *Proceedings of the IEEE/SP 15th Workshop on Statistical Signal Processing (SSP '09)*, pp. 677–680, Cardiff, UK, September 2009.

[39] A. Lombard, Y. Zheng, H. Buchner, and W. Kellermann, "TDOA estimation for multiple sound sources in noisy and reverberant environments using broadband independent component analysis," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 19, no. 6, pp. 1490–1503, 2011.

[40] N. Strobel, S. Spors, and R. Rabenstein, "Joint audio-video object localization and tracking: a presentation general methodology," *IEEE Signal Processing Magazine*, vol. 18, no. 1, pp. 22–31, 2001.

[41] N. Strobel, S. Spors, and R. Rabenstein, "Joint audio-video signal processing for object localization and tracking," in *Microphone Arrays: Signal Processing Techniques and Applications*, Digital Signal Processing, pp. 203–225, Springer, Berlin, Germany, 2001.

[42] D. Bechler, M. Grimm, and K. Kroschel, "Speaker tracking with a microphone array using kalman filtering," *Advances in Radio Science*, vol. 1, pp. 113–117, 2003.

[43] I. Potamitis, H. Chen, and G. Tremoulis, "Tracking of multiple moving speakers with multiple microphone arrays," *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 5, pp. 520–529, 2004.

[44] U. Klee, T. Gehrig, and J. McDonough, "Kalman filters for time delay of arrival-based source localization," *EURASIP Journal on Applied Signal Processing*, vol. 2006, Article ID 012378, pp. 1–15, 2006.

[45] S. Gannot and T. G. Dvorkind, “Microphone array speaker localizers using spatial-temporal information,” *EURASIP Journal on Applied Signal Processing*, vol. 2006, Article ID 059625, pp. 1–17, 2006.

[46] Z. Liang, X. Ma, and X. Dai, “Robust tracking of moving sound source using multiple model Kalman filter,” *Applied Acoustics*, vol. 69, no. 12, pp. 1350–1355, 2008.

[47] C. Seguraa, A. Abad, J. Hernando, and C. Nadeu, “Multispeaker localization and tracking in intelligent environments,” in *Multimodal Technologies for Perception of Humans*, vol. 4625 of *Lecture Notes in Computer Science*, pp. 82–90, Springer, Berlin, Germany, 2008.

[48] D. N. Zotkin, R. Duraiswami, and L. S. Davis, “Joint audio-visual tracking using particle filters,” *EURASIP Journal on Applied Signal Processing*, vol. 2002, Article ID 162620, pp. 1154–1164, 2002.

[49] F. Antonacci, D. Riva, D. Saiu, A. Sarti, M. Tagliasacchi, and S. Tubaro, “Tracking multiple acoustic sources using particle filtering,” in *Proceedings of the 14th European Signal Processing Conference*, pp. 1–4, Florence, Italy, September 2006.

[50] J.-M. Valin, V. F. Michaud, and J. Rouat, “Robust localization and tracking of simultaneous moving sound sources using beamforming and particle filtering,” *Robotics and Autonomous Systems*, vol. 55, no. 3, pp. 216–228, 2007.

[51] F. Talantzis, A. Pnevmatikakis, and A. G. Constantinides, “Audio-visual active speaker tracking in cluttered indoors environments,” *IEEE Transactions on Systems, Man, and Cybernetics B*, vol. 38, no. 3, pp. 799–807, 2008.

[52] A. Quinlan, M. Kawamoto, Y. Matsusaka, H. Asoh, and F. Asano, “Tracking intermittently speaking multiple speakers using a particle filter,” *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2009, Article ID 673202, pp. 1–11, 2009.

[53] A. Levy, S. Gannot, and E. A. P. Habets, “Multiple-hypothesis extended particle filter for acoustic source localization in reverberant environments,” *IEEE Transactions on Audio, Speech and Language Processing*, vol. 19, no. 6, pp. 1540–1555, 2011.

[54] T. Nishiura, T. Yamada, S. Nakamura, and K. Shikano, “Localization of multiple sound sources based on a CSP analysis with a microphone array,” in *Proceedings of the IEEE Interntional Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 1053–1056, Istanbul, Turkey, June 2000.

[55] J. Scheuing and B. Yang, “Disambiguation of TDOA estimation for multiple sources in reverberant environments,” *IEEE Transactions on Audio, Speech and Language Processing*, vol. 16, no. 8, pp. 1479–1489, 2008.

[56] J.-S. Hu and C.-H. Yang, “Estimation of sound source number and directions under a multisource reverberant environment,” *EURASIP Journal on Advances in Signal Processing*, vol. 2010, Article ID 870756, pp. 1–14, 2010.

[57] A. Brutti, M. Omologo, and P. Svaizer, “Multiple source localization based on acoustic map de-emphasis,” *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2010, Article ID 147495, pp. 1–17, 2010.

[58] D. Salvati, A. Rodà, S. Canazza, and G. L. Foresti, “A real-time system for multiple acoustic sources localization based on ISP comparison,” in *Proceedings of the 13th International Conference on Digital Audio Effects*, pp. 201–208, Graz, Austria, September 2010.

[59] D. Salvati, A. Rodà, S. Canazza, and G. L. Foresti, “Multiple acoustic sources localization using incident signal power comparison,” in *Proceedings of the 8th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS ’11)*, pp. 77–82, Klagenfurt, Austria, September 2011.

[60] H. Cox, R. Zeskind, and M. Owen, “Robust adaptive beamforming,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 35, no. 10, pp. 1365–1376, 1987.

[61] B. D. Carlson, “Covariance matrix estimation errors and diagonal loading in adaptive arrays,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 24, no. 4, pp. 397–401, 1988.

[62] J. Dmochowski, J. Benesty, and S. Affès, “On spatial aliasing in microphone arrays,” *IEEE Transactions on Signal Processing*, vol. 57, no. 4, pp. 1383–1395, 2009.

[63] B. V. Veen and K. M. Buckley, “Beamforming: a versatile approach to spatial filtering,” *IEEE ASSP Magazine*, vol. 5, no. 2, pp. 4–24, 1988.

[64] Q.-H. Huang, Q. Zhong, and Q.-L. Zhuang, “Source localization with minimum variance distortionless response for spherical microphone arrays,” *Journal of Shanghai University*, vol. 15, no. 1, pp. 21–25, 2011.

[65] J. Makhoul, “Linear prediction: a tutorial review,” *Proceedings of the IEEE*, vol. 63, no. 4, pp. 561–580, 1975.

[66] R. J. McAulay, “Maximum likelihood spectral estimation and its application to narrow-band speech coding,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 2, pp. 243–251, 1984.

[67] L. L. Pfeifer, “Inverse filter for speaker identification ,” RADC TR-74-214, Speech Communications Research Lab Inc, Santa Barbara, Calif, USA, 1974.

[68] J. A. H. Gray Jr. and J. D. Markel, “Distance measures for speech processing,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 24, no. 5, pp. 380–391, 1976.