Signal Processing 153 (2018) 1–10

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Sensitivity-based region selection in the steered response power algorithm

Daniele Salvati*, Carlo Drioli, Gian Luca Foresti

Department of Mathematics, Computer Science and Physics, University of Udine, Italy

ARTICLE INFO

Article history:
Received 28 August 2017
Revised 18 May 2018
Accepted 2 July 2018
Available online 3 July 2018

Keywords:
Acoustic source localization
Microphone array
SRP-PHAT
Sensitivity map
Region selection
Geometrically sampled grid

ABSTRACT

The steered response power (SRP) algorithm is a well-studied method for acoustic source localization using a microphone array. Recently, different improvements based on the accumulation of all time difference of arrival (TDOA) information have been proposed in order to achieve spatial resolution scalability of the grid search map and reduce the computational cost. However, the TDOA information distribution is not uniform with respect to the search grid, as it depends on the geometry of the array, the sampling frequency, and the spatial resolution. In this paper, we propose a sensitivity-based region selection SRP (R-SRP) algorithm that exploits the nonuniform TDOA information accumulation on the search grid. First, high and low sensitivity regions of the search space are identified using an array sensitivity estimation procedure; then, through the formulation of a peak-to-peak ratio (PPR) measuring the peak energy distribution in the two regions, the source is classified to belong to a high or to a low sensitivity region, and this information is used to design an ad hoc weighting function of the acoustic power map on which the grid search is performed. Simulated and real experiments show that the proposed method improves the localization performance in comparison to the state-of-the-art.

1. Introduction

Acoustic source localization using microphone arrays has received significant attention by the scientific community due to its importance in sound scene analysis, signal enhancement, and speaker recognition and tracking [1–10].

In general, the localization can be computed with indirect and direct methods. The former are based on the computation of a set of time differences of arrival (TDOAs), obtained by measurements across various combinations of microphones [11,12], and on the estimation of the source position using geometric reasoning [13–15]. Direct methods are based on the steered response power (SRP) beamformers [16–18], on subspace algorithms [19–21], or on maximum-likelihood estimators [22–24]. They are very attractive for acoustic applications due to their robustness in noisy and reverberant conditions.

The conventional SRP algorithm is based on the delay-and-sum beamforming technique [25]. Broadband SRP is typically implemented with the phase transform (PHAT) pre-whitening [11,26], which provides a normalization of narrowband SRPs and increases the spatial resolution [27]. This allows a better identification of direct path and early reflections in a reverberant environment. SRP-PHAT has the advantage that it can be computed by considering the generalized cross-correlation (GCC) [11] between each microphone pair, and by summing TDOA values related to the search space [17]. This implementation is computationally more efficient if compared to methods that require a computation of narrowband SRP maps and their fusion [27]. However, the search procedure can be very expensive. Thus, iterative volume-search-based procedures have been recently proposed [28–30], which aim at reducing the computational complexity of this step. These methods take into account the accumulation of TDOA information [29–31] to achieve the reduction of the spatial grid resolution without loss of information, and uses sequentially volumetric refinement steps for increasing the localization accuracy.

It has been demonstrated, using the geometrically sampled grid (GSG) algorithm [32], that the accumulation of all TDOA values from GCC functions is not uniform within the search space, and as a consequence the acoustic map is characterized by high and low sensitivity regions. The advantage of using all TDOA information is to obtain a robust localization in the high sensitivity region with adverse noisy and reverberant conditions. If the sound source is located in a low sensitivity region, however, its localization is more prone to be unstable and affected by errors. This is due to the fact that the acoustic map energy peak corresponding to the actual source position might be lower than the peaks correspond-

* Corresponding author.
E-mail addresses: daniele.salvati@uniud.it (D. Salvati), carlo.drioli@uniud.it (C. Drioli), gianluca.foresti@uniud.it (G.L. Foresti).

https://doi.org/10.1016/j.sigpro.2018.07.002

ing to noise and reverberation in the high sensitivity region, emphasized by the prominent TDOA accumulation. SRP-based methods that use all TDOA information were proposed in [29,31,32]. In [28], it was also proposed a SRP method that uses all TDOA information, providing however a power normalization in each volume with respect to the number of TDOA values. This approach mitigates the problem due to the nonuniform TDOA accumulation, but also reduces the robustness in the high sensitivity region. In [30], a SRP method based on the use of two grids (a coarser one, and a finer one) was proposed. This method uses a uniform TDOA accumulation in each volume, mitigating the problem of nonuniform distribution but meanwhile discarding part of the information available, thus reducing the TDOA accumulation that can be positively used in the high sensitivity region. In fact, the final resolution is given by the finer grid, implemented with a conventional SRP approach: for each microphone pair and for each point on the grid, a unique integer TDOA value is selected to be the acoustic delay information linked to that point. This uniform regular grid procedure does not guarantee that all TDOA samples are associated to points on the grid and does not exploit the accumulation of TDOA values that can be positively used in the high sensitivity region also with a finer grid. Note that it was demonstrated in [32] that using all TDOA samples can improve the localization performance in the high sensitivity region with a coarser grid and with a finer grid up to a resolution of 0.01 m.

In this paper, we consider the localization of a single source in noisy and reverberant conditions. This scenario can be of interest in different practical applications such as videoconferencing systems or in human-computer interaction systems. We propose a sensitivity-based region selection SRP algorithm, named R-SRP, which has the following characteristics: 1. it uses all the TDOA information provided by the GCC functions; 2. it exploits the localization robustness in the high sensitivity region; 3. with respect to other methods, it allows to use coarser search grids in a more effective way, thus reducing the computational cost. The algorithm is organized in two steps. First, it establishes if the source is positioned in a high or low sensitivity region, through the formulation of a peak-to-peak ratio (PPR) measuring the peak energy distribution in the high and low sensitivity regions of the array, determined through the GSG algorithm. Then, it proceeds with the search of the acoustic source in the selected region using, when opportune, the sensitivity map to weight the power acoustic map and reduce the impact of noise. It will be shown that this array sensitivity-informed method effectively reduces the localization errors due to the nonuniform distribution of the TDOA accumulation in the power acoustic map.

2. Steered response power

Let us consider a reverberant room G, M microphones positioned at coordinates $\mathbf{r}_m = [x_m, y_m, z_m]^T (m = 1, 2, \ldots, M)$, where $(\cdot)^T$ denotes the transpose operator, and a single source $\mathbf{r}_s(k) = [x_s(k), y_s(k), z_s(k)]^T$ active at frame index k. The SRP-PHAT based on all the TDOA information can then be expressed in terms of GCC functions as [17,28,29,31,32]

$$\phi(\mathbf{r}, k) = \sum_{m_1=1}^{M-1} \sum_{m_2=m_1+1}^{M} \sum_{\tau=\tau_{m_1m_2}^{\min}(\mathbf{r})}^{\tau_{m_1m_2}^{\max}(\mathbf{r})} R_{m_1m_2}(\tau, k), \tag{1}$$

where $\mathbf{r} = [x, y, z]^T \in G$ is a generic grid position with spatial resolution Δ, $\tau_{m_1m_2}^{\min}(\mathbf{r})$ and $\tau_{m_1m_2}^{\max}(\mathbf{r})$ denote the bounds of the accumulated TDOAs between microphones m_1 and m_2 for the position $\mathbf{r}$, and the GCC-PHAT [11] function is

$$R_{m_1m_2}(\tau, k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{X_{m_1}(w, k) X_{m_2}^*(w, k)}{|X_{m_1}(w, k) X_{m_2}^*(w, k)|} e^{jw\tau} dw, \tag{2}$$

where τ is the time lag, w is the angular frequency, $X_m(w, k)$ is the transform of the signal observed at microphone m, $(\cdot)^*$ denotes the complex conjugate, j denotes the imaginary unit, and $|\cdot|$ denotes absolute value. The GCC-PHAT is computed in the frequency domain using the discrete Fourier transform (DFT), and hence the SRP is computed on a block-by-block basis. If $\tau_{m_1m_2}^{\min}(\mathbf{r}) = \tau_{m_1m_2}^{\max}(\mathbf{r})$, Eq. (1) represents the conventional SRP-PHAT algorithm [17]. The accumulation limits can be determined with different strategies which can rely on the gradient of the inter-microphone time delay function corresponding to each microphone pair in the M-SRP [31], on the gradient of the inter-microphone time delay function exploiting the mean of the accumulated GCC-PHAT values for each volume in the I-SRP [28], on the surrounding cube taking into account vertices of the volume in the H-SRP [29] or selecting only some points related to a finer grid in the RV-SRP [30], or on discrete representations of the hyperboloids related to all possible TDOA values in the GSG-based method (G-SRP) [32].

Once the array steered response power function $\phi(\mathbf{r}, k)$ is available, the source position can be estimated by searching its maximum in the search region

$$\widehat{\mathbf{r}}_s(k) = \underset{\mathbf{r}}{\operatorname{argmax}}[\phi(\mathbf{r}, k)]. \tag{3}$$

3. Geometrically sampled grid

The proposed R-SRP algorithm extends the G-SRP [32] algorithm by including a region selection procedure. The G-SRP is based on the GSG method, in which the search space is obtained by discretizing, with a given spatial resolution, the hyperboloids representing the surface on which the TDOAs are constant, and by finally computing a grid related to the intersections between these discrete curves. It thus allows the accumulation of the whole TDOA information provided by the GCC functions into the search space, the design of an acoustically-coherent space grid, and the design of a sensitivity map for the array in use. The use of all the TDOA information available from the GCC-PHAT functions solves the problem of arbitrarily selecting the spatial grid resolution without loss of information. The acoustically-coherent space grid guarantees that every point of the grid is consistent with the condition of being the locus where at least three half-hyperboloids intersect. Note that the coherent grid may discard the points of the uniform regular grid (used in the conventional approach) which are not covered by sufficient acoustic information, especially when a finer grid is used. The sensitivity map refers to a quantified measure of the change of the response power with respect to the change of the spatial position, predicting where the search space will be characterized by higher and lower localization accuracy. Note that another approach to identify the spatial localization accuracy was proposed in [33], in which a discriminability measure is defined to distinguish a given point in space from its neighbors. This method does not consider the spatial resolution, and it does not give useful information when a larger resolution is used [32].

Let us now consider the dicretization of the search space G with a spatial resolution Δ. A discrete hyperboloid related to a microphone pair (m_1, m_2) and a TDOA $\tau_{m_1m_2}$ can be represented as a finite set $\Lambda_{\tau_{m_1m_2}}^{\Delta}$ of points in $\mathbb{R}^3$, describing the hyperboloid when the x, y, and z-axis are discretized with spatial resolution Δ (for a detailed discussion on the hyperboloid discretization procedure, see [32]).

In the implementation of the G-SRP, the discrete hyperboloids and the TDOA information are stored in four look-up tables. The tables are computed off-line, and then used on-line to estimate the acoustic energy and computing the accumulation of the GCC-PHAT function information due to all the sensor pairs involved. To each discrete hyperboloid point we assign an index q, so that we have a table $\gamma_r(q)$ that stores the position to which each hyperboloid