Applied Acoustics 139 (2018) 267–281

Contents lists available at ScienceDirect

Applied Acoustics

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

3D sound source localization based on coherence-adjusted monopole dictionary and modified convex clustering

Tomoya Tachikawa*, Kohei Yatabe, Yasuhiro Oikawa

Department of Intermedia Art and Science, Waseda University, 3-4-1 Ohkubo, Shinjuku-ku, Tokyo 169-8555, Japan

ARTICLE INFO

Keywords:
Sparsity
Direction-of-arrival (DOA)
Distance estimation
Convex optimization
Primal-dual splitting

ABSTRACT

In this paper, a sound source localization method for simultaneously estimating both direction-of-arrival (DOA) and distance from the microphone array is proposed. For estimating distance, the off-grid problem must be overcome because the range of distance to be considered is quite broad and even not bounded. The proposed method estimates the positions based on a modified version of the convex clustering method combined with the sparse coefficients estimation. A method for constructing a suitable monopole dictionary based on the coherence is also proposed so that the convex clustering based method can appropriately estimate distance of the sound sources. Numerical and measurement experiments were performed to investigate the performance of the proposed method.

1. Introduction

Sound source localization techniques using a microphone array have been interested in many applications including speaker localization, robot communications and electronic surveillance. A lot of methodologies have been studied for estimating the direction-of-arrival (DOA) of the sound sources only from observed sound by a microphone array [1–8]. To formulate the estimation problems, models of the sound wave, or a dictionary, must be chosen beforehand. The plane waves are used as a dictionary in these methods because the far-field assumption allows us to approximate the wavefront of far-field sources as the plane waves.

For estimation, DOA is usually discretized into a grid so that only a finite number of directions in the dictionary must be considered. Although it is possible to formulate the DOA estimation problem whose directions are treated as continuous variables, such strategy often suffers from the initialization issue. Therefore, to make the problem simpler, the variables are discretized into a grid. This on-grid model is preferable in many cases as it is tractable for many algorithms. However, the grid discretization can be a source of estimation error because only a direction which coincides with the grid can be perfectly represented by the dictionary. It can be regarded as a trade-off between accuracy and computational cost: A fine grid provides better accuracy with more computational complexity. Such discretization issue is called the off-grid problem, and several methods have been proposed to overcome it [8–11].

While the DOA estimation problems (only considering directions) are the common settings, a problem of estimating both direction and distance is also important. When a source is closely located to the microphone array, the sound source must be considered as a function of both direction and distance because the far-field assumption is violated and the plane-wave modeling does not valid. Then, estimation of the distance of the sound source is necessary in addition to the DOA estimation.

For estimating distance, sound sources must be parameterized by both direction and distance. In the previous research, one popular methodology for that is to extend the subspace method, in particular the MUSIC method, to include distance as the parameter [12–14]. Although the concept of the subspace methods is well-accepted, its accuracy may decrease if some model mismatch exists. Such mismatch between the model and observed signal is caused by excessive simplification of the model. This modeling issue can be circumvented by considering a more sophisticated model within the formulation.

Some recent research has shown that sophisticated models combined with sparse estimation techniques are effective for localizing sound sources [15]. For example, if the room impulse responses (RIR) at many locations are known, the source localization performance can be improved as the reverberation time increases [16]. Another example showed that complete modeling of a sound field by the finite element method (FEM) can localize an impulsive source from an observation obtained by a single microphone in principle [17]. Although these results indicate the effectiveness of a richer model for localization, constructing a model consisting of RIR or FEM in advance is difficult because complete information of the sound field or room geometry is

* Corresponding author.
E-mail address: tachi0611s.w@gmail.com (T. Tachikawa).

https://doi.org/10.1016/j.apacoust.2018.04.033
Received 29 September 2017; Received in revised form 23 March 2018; Accepted 25 April 2018

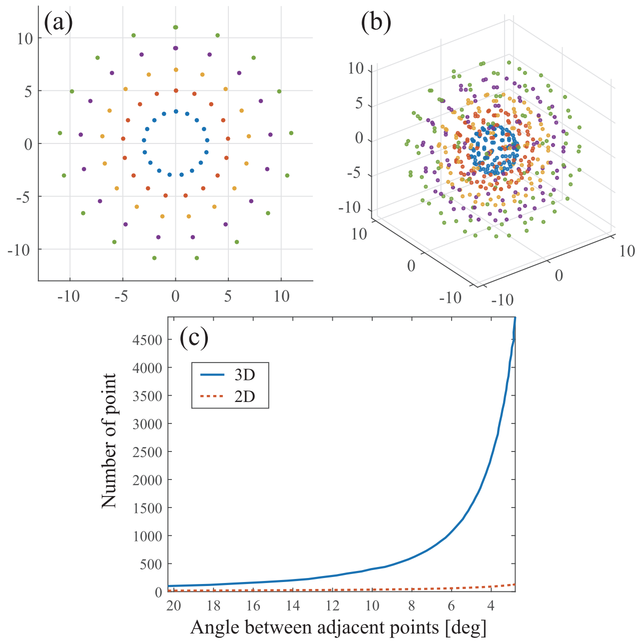

Fig. 1. Illustration of the growth of the number of grid points in 2D and 3D spaces. 2D and 3D grid points are shown in (a) and (b), respectively, where the spacing between the points are set to the same value (angle: 20.3 degree, distance: 3, 5, 7, 9 and 11 m), which ends up with 90 points for (a) and 500 points for (b). The relation between the angle and number of the grid points for a single circle and sphere is shown in (c). The number of grid points rapidly grows in 3D space as the grid becomes dense.

required. Their computational costs can also be problematic when a large sound field is considered. That is, too much sophistication of the sound source and sound field models causes another issues related to model complication and/or computational cost.

A simpler but relatively sophisticated model parameterized by both direction and distance is a set of monopoles [18]. As later shown in Section 2.1, multiple monopoles, or the fundamental solution of the Helmholtz equation, can approximate the sound source distribution generating the sound field [19,20]. That is, sufficiently large number of monopoles, which are positioned appropriately, can approximate the observed sound field which satisfies the Helmholtz equation. Modeling by many monopoles does not require complicated information such as room geometry, and thus it is easier to apply in real situation. For these reasons, a set of monopoles is considered for modeling sound fields in this paper.

However, the off-grid problem becomes a quite serious problem for such monopole model in the three-dimensional (3D) space because the number of grid points grows extremely fast. Fig. 1 illustrates how the grid points increase in the 2D and 3D spaces. Although spacing of the grid points of Fig. 1(a) and (b) is the same, the number of points is notably different: 90 points for 2D space and 500 points for 3D space. The growth of the points is summarized in Fig. 1(c), where the grid becomes dense as the horizontal axis goes right. These figures indicate that increasing the grid points for reducing the effect of the off-grid problem results in enormous points which can be impossible to solve owing to the high computational requirement. Note that the aforementioned methods [12–18] did not suffer form this severe issue because they considered only 2D space (source locations are parametrized by two variables) [21].

Existing 3D sound source localization methods are based on the MUSIC methods [22–25], and there is no method which utilizes a set of monopoles as the model of a 3D sound field, to the best of our knowledge. This should be because of the trade-off between the off-grid problem and the problematic growth of the number of grid points. In order to apply monopoles as the model in 3D space, resolving the off-grid problem even for a very coarse grid is necessary. This situation prevents from using the ordinary methods for tackling the off-grid problem [26,27] because continuous modeling as in [26] is difficult in the 3D space while an interpolative method as in [27] cannot perform well for a coarse grid.

In this paper, methods for preventing unreasonable increase of the size of the monopole dictionary and for solving the off-grid problem are proposed as parts of a 3D sound source localization method.[1] The main part of 3D localization is a sparsity based method which is not new. Instead, our contributions are on the dictionary construction part and the post processing part, which can be summarized as follows: (i) A method of constructing a monopole-only dictionary is proposed so that the number of monopoles in the dictionary can be controlled based on the similarity measure, or coherence; (ii) A clustering method based on the convex optimization is proposed as a post processing method to overcome the off-grid problem; (iii) An algorithm for solving the convex clustering method based on the primal-dual splitting method is derived. These proposals can be combined with any sparsity based localization method, and thus our contributions may be in a broader context rather than a single 3D sound source localization method.

The ideas behind the proposed methods (dictionary construction and post processing) are the following. In the dictionary construction part, the proposed method tried to adjust the positions of the monopoles based on a similarity measure called coherence. The coherence indicates how two monopoles are similar in the domain of observed signals. Since excessively similar monopoles end up with an overly redundant dictionary which requires high computational resources, similarity among each monopoles should be arranged in the same order. However, optimizing the position of a lot of monopoles is quite a difficult problem because of the non-convexity [30]. The proposed method avoids the difficulty by fixing the direction of monopoles and simplifying the problem so that an easy and tractable algorithm can be obtained.

In the post processing part, a modified version of the convex clustering [31–33] is proposed to gather the candidates of the location so that the off-grid problem is resolved. After the sparsity based localization, several coefficients corresponding to a single sound source contain values. Then, one has to find a peak value from the coefficients which is regarded as the estimated position. This process of selecting the large coefficient occurs as post processing in the most of localization methods. The proposed method formulates this post processing as a convex optimization problem so that the coefficient selection problem and the off-grid problem are simultaneously solved. The positions are treated as continuous variables in the method, and distances among the positions are sparsified. This process gathers the nearby grid points with large coefficients into a single point. Therefore, the gathered point is regarded as the selected location which is not limited on the grid.

The rest of this paper is organized as follows. The sparsity based sound source localization used in this paper is briefly explained in Section 2. The importance of the selection of large coefficient is also demonstrated in that section. The proposed dictionary construction method is explained in Section 3, and the proposed post processing method is developed in Section 4. Numerical simulations in Section 5 and measurement experiments in Section 6 show the validity of the proposed methods. Finally, Section 7 concludes the paper.

2. Sparse localization using monopoles

In this section, the sparsity based localization method used in this paper is briefly introduced. An algorithm for solving the optimization problem arising in the method is also shown. In addition, the importance of post processing to determine the estimated position from

[1] The preliminary versions of this work have appeared in conference proceedings [28,29].