Journal of Sound and Vibration 401 (2017) 152–168

Contents lists available at ScienceDirect

Journal of Sound and Vibration

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

Performance analysis of microphone array methods

Gert Herold*, Ennes Sarradj

Institute of Fluid Mechanics and Engineering Acoustics, Technische Universität Berlin, Einsteinufer 25, D-10587 Berlin, Germany

ARTICLE INFO

Article history:
Received 15 September 2016
Received in revised form
22 March 2017
Accepted 23 April 2017
Handling Editor: Prof Z Su

Keywords:
Microphone array
Beamforming
Reliability

ABSTRACT

Microphone array methods aim at the characterization of multiple simultaneously operating sound sources. However, existing data processing algorithms have been shown to yield different results when applied to the same input data. The present paper introduces a method for estimating the reliability of such algorithms. Using Monte Carlo simulations, data sets with random variation of selected parameters are generated. Four different microphone array methods are applied to analyze the simulated data sets. The calculated results are compared with the expected outcome, and the dependency of the reliability on several parameters is quantified. It is shown not only that the performance of a method depends on the given source distribution, but also that the methods differ in terms of their sensitivity to imperfect input data.

1. Introduction

Microphone array methods are not merely a way to visualize acoustic sources, but an effective tool for quantitative measuring [1]. The catalog of methods for calculating source level distributions from synchronously recorded sound pressures comprises several different approaches. These include enhancements of classic beamforming techniques [2], deconvolution methods based on the beamforming result [3–5], and beamforming-independent inverse methods [6,7].

Microphone array methods map the spatial distribution of acoustic sources in terms of sound pressure level radiated towards a reference position. A desirable property of any array method is the ability to reconstruct sources with high spatial resolution, minimum artifacts and a high dynamic range. However, existing methods have been shown to perform differently depending on the given task. A previous evaluation of measurement data by Herold and Sarradj [8] revealed that different methods may yield different results even when applied to the same input from a basic experimental setup.

In the literature, many specific and exemplary comparisons of array methods can be found. Examples for experimental studies are the works conducted by Yardibi et al. [9] and by Chu and Yang [10]. While experiments are apt for yielding realistic results for the performance of the examined methods, in general, their focus is on a specific setup, whose parameters can only be varied to a very limited degree. Furthermore, with experiments, it is often not possible to define a reference with which the results obtained through a microphone array method are to be compared, which makes evaluating their performance difficult.

Using simulated data as input for the array methods has the advantage that the expected output is known and can be used as reference. Studies based on simulated input data have been done, amongst others, by Leclere et al. [11], who evaluated the qualitative differences between algorithms, and by Ehrenfried and Koop [12], who tested the algorithm

* Corresponding author.
E-mail addresses: gert.herold@tu-berlin.de (G. Herold), ennes.sarradj@tu-berlin.de (E. Sarradj).

http://dx.doi.org/10.1016/j.jsv.2017.04.030

performance of methods similar to the DAMAS algorithm. However, even there, the variation of input data is generally not very extensive if existing, and the scope of possible performance evaluations therefore limited.

In addition to that, the methods compared in any evaluation, including this study, only ever represent a subset of available algorithms. This often makes a comparison with other methods difficult, in particular in the case of experimental setups.

For testing a new algorithm, it would be of advantage to be able to systematically compare its performance with other algorithms, depending on several parameters that are suspected to have an influence on the reconstruction ability. Therefore, the first objective of the present study is to provide a data generation framework that simulates typical use cases for a microphone array method but is easily extendible and adaptable. As it should allow the examination of arbitrary dependencies, a statistic approach is appropriate in order to avoid a huge number of necessary systematic parameter variations.

The second objective is to provide rating criteria to quantify the deviation of the reconstruction from the correct solution. Approaches for evaluating the calculated results differ between method-comparing studies. The most basic way consists of visually assessing the map of source levels and examine whether the represented features correspond to the expected source distribution. While this constitutes a simple way to verify whether a method works at all and to illustrate its potential differences to other methods, it depends on subjective criteria and is not suited for evaluating a high number of maps.

The spatial resolution capability of a method can be specified by evaluating the extent of the representation of a point source on a map, i.e., the sharpness of the peak, as this limits the minimum distance at which two sources appear separate. For comparing methods, a common way is to use the Rayleigh resolution limit of classic beamforming [13] as reference and calculate the minimum resolvable distance as fraction of that limit. Dougherty [14] proposed applying the Sparrow limit instead, since it is explicitly defined as the closest possible distance at which a minimum between two peaks still appears [15]. These criteria are suitable for assessing the capability of an algorithm to separate point sources of similar levels. However, they do not allow evaluating the correct reconstruction of the position and level of the sources.

As a quantitative measure of the reconstruction result, Ehrenfried and Koop [12] proposed the standard deviation of the per-grid-point difference between the reconstructed map and the correct solution. With this, both the absolute level of the sources and their positions are evaluated within a single value. However, a small error in the positioning of a source is penalized as much as a completely wrong position or even a disappearing source. Furthermore, it is not possible to separately treat the reconstruction of multiple sources in one data set.

The quantitative comparison between microphone array methods can also be done by integrating the reconstructed sound pressures over an area of interest. As the choice of the integration sectors is flexible, this approach is not limited to point sources but can be used for arbitrary source shapes [3,5,16]. The challenge here is to define integration areas in such a way that the integrated levels allow deducing meaningful information regarding the qualitative and quantitative source reconstruction ability of the methods. The approach used in this paper bears similarities with that pursued in a previous study by Herold and Sarradj [17], where the evaluation of the maps is based on integrating circular areas around point source positions.

The remainder of this paper is organized as follows. After a short summary of the microphone array methods to be compared in this study, the evaluation methodology applied here is described in detail. It consists of defining the basic setup, the rules for generating simulated data, and the approach to evaluate the maps calculated with the array methods. Subsequently, results from these evaluations are discussed, and important findings are summarized in the concluding section.

2. Microphone array methods

Microphone array methods rely on the evaluation of phase and amplitude differences of signals recorded at a number of distributed sensors. The data processing can either be done in time domain or in frequency domain [18]. The methods considered in this study work in the frequency domain and are based on the cross-spectral matrix (CSM) of the microphone signals.

The CSM is estimated using Welch's method [19]: Each time signal is divided into K blocks, onto which an FFT is applied. For each discrete frequency, the resulting complex sound pressures are stored in a vector $\mathbf{p}_k \in \mathbb{C}^M$, which has as many entries as microphone channels used. The cross-spectra between the channels are calculated and stored in matrix form for every block. Finally, the CSM is calculated by averaging all cross-spectra:

$$\mathbf{C} = \frac{1}{K}\sum_{k=1}^{K} \mathbf{p}_k\, \mathbf{p}_k^{\mathrm{H}}. \tag{1}$$

The main diagonal of the CSM contains the autospectra of the microphone channels, which hold no information about the phase differences between microphones. Since, however, its entries may contain uncorrelated self-noise of the channels from measurements, it is common to omit them in the further calculations.

The classic delay-and-sum beamformer formulation in the frequency domain is

$$b(\mathbf{x}_t) = \mathbf{h}^{\mathrm{H}}(\mathbf{x}_t)\, \mathbf{C}\, \mathbf{h}(\mathbf{x}_t), \quad t = 1 \ldots N, \tag{2}$$

where $b(\mathbf{x}_t)$ contains the squared sound pressure characterizing a source at a focus point $\mathbf{x}_t$. For acoustic source mapping, a spatial domain assumed to contain sources is discretized with a grid consisting of N focus points. The steering vector $\mathbf{h}$