Aspects of source separation in beamforming

Ennes Sarradj[1], Christian Schulze[1] and Andreas Zeibig[2]

[1]Brandenburgische Technische Universität Cottbus, Institut für Verkehrstechnik,
Siemens-Halske-Ring 14, 03046 Cottbus, GERMANY
[2]Technische Universität Dresden, Institut für theoretische Grundlagen der Fahrzeugtechnik,
01062 Dresden, GERMANY

ABSTRACT

One of the most important elements to consider in the application of beamforming is the identification of sources from the beamforming result. This result is often given in form of a map of sound pressure contributions from a number of locations. While on first look the dominating sources may be easily identified, a number of problems arise during practical application: (a) sources closely spaced cannot be separated due to the width of the directional beam, (b) minor sources cannot be detected at all due to the limited signal-to-noise ratio, (c) it is not possible to separate sound pressure contributions from different source mechanisms. Some signal processing methods to deal with these problems are explained in short. The implications of each this methods are discussed. Practical examples using a 32-microphone array are given.

1 Introduction

The ultimate goal in the application of a beamforming phased microphone array is most often the identification and localization of sound sources, or in other words, the source separation. This task occurs usually when a complex acoustic situation is at hand and the number, strength and location of sources are not known a priori. The result of a phased microphone array measurement is conveniently given as a map of sound pressure contributions from a large number of locations on a virtual surface. The information that is to be drawn from the result includes the number of sources or source mechanisms contributing and the location and strength of the sources. It is only possible to gain that information if the spatial resolution is good enough to distinguish between multiple sources. Further, to detect minor sources the map must have a high dynamic range or so-called signal-to-noise ratio. Lastly, to decide on the number of sources and source mechanisms some sort of an indicator is necessary that marks individual source mechanisms as such. In a practical application, these three prerequisites are not always met, thus resulting in one or more of the following problems:

- sources closely spaced cannot be separated due to limited resolution,
- minor sources cannot be detected at all if the signal-to-noise ratio is too low,
- it is not possible to separate sound pressure contributions from different source mechanisms if their source regions are too closely spaced or overlapping.

These aspects of the separation of sound sources using a beamforming technique are discussed in what follows. After a brief introduction into the theory of beamforming the relevant quality parameters for the first two problems are introduced. Then, the information about the sources inherent in the sound field is analyzed. Some signal processing methods that can help improving the result are explained in short and example results are presented for illustration.

2 Theory

An array of sound receivers or microphones together with an appropriate signal processing constitutes a directional sound receiver. The working principle of such device may be explained using an analogy to concave mirrors equipped with a single microphone. Sound waves from a certain direction or origin will be reflected at such mirror and then focused in one focal point. If a microphone is put into that focal point, the sound from that origin is much more strongly received than that from other origins – this is the characteristic of a directional receiver. A concave mirror works this way because the time of travel of the sound waves from the source to the focal point is the same regardless of the location in the mirror where the reflection occurs. Thus, due to constructive interference the resulting sound pressure in the focal point will be much higher as it would be without the mirror.

Such constructive interference may be also achieved by using a number of microphones placed at different locations, often but not necessarily within a plane. Assume a certain individual delay Δt_i is applied to each microphone signal y_i and the signals are amplified by appropriate factors of A_i. Then, the summation of the suchlike modified signals may result in constructive interference of the signals if the sound wave captured by the microphones stems from a source at a certain location. This is an analogue to a concave mirror.

While a parabolic mirror focuses the sound from a certain direction and an elliptic mirror does the same for sound originating from the second focal point, in both cases it is necessary to modify position and maybe also the shape of the mirror to alter the point $\vec{x_0}$ or direction on which the array is focused. Using a microphone array, this change can be accomplished by using a different set of Δt_i and A_i. Thus, it is possible to conveniently construct maps of sound pressure contributions.

The procedure of applying certain delays and amplitude factors is also called beamforming; the device to do so is a beamformer. In short, the output of the beamformer with M microphone signals y_i is given by:

$$z(t) = \sum_{i=1}^{M} w_i A_i y_i (t - \Delta t_i). \tag{1}$$

w_i is an optional weighting factor that may be applied to a signal. A_i and Δt_i depend on the point $\overrightarrow{x_0}$ to which the array is focused. The output spectrum of the beamformer reads then as:

$$Z(\omega) = \sum_{i=1}^{M} w_i e_i(\overrightarrow{x_0}) Y_i(\omega), \tag{2}$$

where e_i incorporates the combined effect of A_i and Δt_i. If the beamformer output spectrum is multiplied by its conjugate $Z^*(\omega)$, the result is a power spectrum:

$$|Z(\omega)|^2 = Z(\omega)Z^*(\omega) = \mathbf{e}(\overrightarrow{x_0})^H \mathbf{W}\mathbf{Y}(\omega)\mathbf{Y}(\omega)^H \mathbf{W}^H \mathbf{e}(\overrightarrow{x_0}). \tag{3}$$

$\mathbf{e}(\overrightarrow{x_0})$ is the so called steering vector (that steers the directional characteristic to $\overrightarrow{x_0}$) and has all e_i as its column elements. $\mathbf{W}$ is a diagonal matrix where w_i weighting factors populate the main diagonal and $\mathbf{Y}$ is a column vector containing the spectra of all microphone signals. The most inner term in equation (3) yields the cross spectral matrix $\mathbf{S_{xy}}$ of the microphone signals. If the power spectrum from (3) is normalized by the squared trace of $\mathbf{W}$, the result is the normalized (or true) beamformer output power spectrum:

$$S(\omega) = \frac{\mathbf{e}(\overrightarrow{x_0})^H \mathbf{W}\mathbf{S_{xy}}\mathbf{W}^H \mathbf{e}(\overrightarrow{x_0})}{(\operatorname{trace}\mathbf{W})^2}. \tag{4}$$

This spectrum corresponds to the sound level contribution from $\overrightarrow{x_0}$ in the array centre location:

$$L_S(\omega, \overrightarrow{x_0}) = 10 \lg \left(\frac{S_(\omega)}{p_0^2} \right) \text{dB}. \tag{5}$$

If the phased array is seen as a directional sound receiver, the directional efficiency of that receiver is called the array pattern:

$$W(\omega, \overrightarrow{x}, \overrightarrow{x_0}) = \frac{S(\omega, \overrightarrow{x})}{S(\omega, \overrightarrow{x_0})}. \tag{6}$$

The array pattern comprises one main lobe at the location of maximum sensitivity $\overrightarrow{x_0}$ to which the array is steered and a number of side lobes. The width of the main lobe is an important quality parameter that controls the spatial resolution of the phased array. If the spacing between two sources is less than the main lobe width it would be not possible to separate them from the beamforming result. Another quality parameter is the signal-to-noise ratio (SNR), that is the amplitude ratio of the highest side lobe to the main lobe. It governs the dynamic range of the sound pressure contribution map. Minor sources with a level difference to the dominating source that is greater than the SNR cannot be detected or separated because the images of such sources cannot be distinguished from 'ghost images' of the dominating source due to the side lobes of the array pattern.

Both the main lobe width and the signal-to-noise ratio depend mainly on the layout of the microphones (the array geometry). Consequently, the layout is an important aspect that influences the ability to separate sound sources [1].

3 The sound field

A sound field is in general composed of the contributions from different sound sources mechanism. These mechanisms may be of aerodynamic, thermal or mechanical nature. Each of them results in one sound source or, more likely, a distribution of sound sources that are not necessarily restricted to one distinct location only. Assume that the mechanisms are independent from each other. Then, the corresponding sound fields are temporally uncorrelated. Further, it can be presumed that these sound fields are also spatially uncorrelated. Consequently it is not possible to represent one of the sound fields by a superposition of the other sound fields. That means that the spatial sound pressure distributions belonging to the individual sound fields are orthogonal.

It is apparent that the overall sound field contains the necessary information to separate the sound sources. An array of microphones effectively samples the sound spatially. Therefore, the input signals of the beamformer contain only a part of that information. In the presence of N different source mechanisms having individual source strength spectra Q_i, the spectrum of the signal from i-th microphone is:

$$Y_i = \sum_{j=1}^{N} f_{ij} Q_j. \tag{7}$$

The f_{ij} are factors that depend on the location of the microphone and on the source location of the source mechanism. The equation (7) may be also written in matrix form:

$$\mathbf{Y} = \sum_{j=1}^{N} \mathbf{f_j} Q_j = \mathbf{fQ}. \tag{8}$$

Due to the orthogonality of the individual sound fields the vectors $\mathbf{f_j} Q_j$ are also orthogonal. Equation (8) is also a mathematical statement of the inverse problem of estimating the unknown $\mathbf{Q}$ (dimension j) from the known $\mathbf{Y}$ (dimension i). If the source locations are known, $\mathbf{f}$ can be computed and for $j \leq i$, a solution for $\mathbf{Q}$ may be obtained using the pseudo-inverse of $\mathbf{f}$. If the source locations are unknown, an estimation of $\mathbf{Q}$ along with the source locations would be still imaginable as solution to a mathematical optimization program. However, if $j > i$ there is no sufficient information available and the inverse problem has a multitude of possible solutions. One reasonable way to deal with this situation is to neglect a number of sources to let j become smaller than i. However, because the inverse problem is ill-posed, this step produces difficulties with a practical solution.

Unfortunately, besides the contribution from the sound field in a practical application there is also unwanted noise from each of the microphone channels. Therefore, the number of 'real' sound sources must be increased be these i noise sources and consequently, $j > i$. This fact makes the use of inverse techniques problematic. On the other hand, the application of the beamforming technique is straightforward and does not rely on the solution of an inverse problem. The question is how to improve the beamforming result to be better able to separate sound sources.

4 Signal processing methods

While the result from beamforming is always limited in spatial resolution and dynamic range, a number of signal processing methods exist that help to extract a maximum of information from a given measurement. Two of these methods will be highlighted here.

Central to the beamforming equation (4) is the cross spectral matrix $\mathbf{S_{xy}}$ of the the microphone signals. If the microphone signals are the sum of the sound pressure $\mathbf{Y}$ from the sound field and unwanted noise $\mathbf{Y_N}$ due to turbulent pressure fluctuations, noise from the amplifiers etc., then this matrix is:

$$\mathbf{S_{xy}} = (\mathbf{Y}+\mathbf{Y_N})(\mathbf{Y}+\mathbf{Y_N})^H = \mathbf{YY}^H + \underbrace{\mathbf{Y_N}\mathbf{Y}^H + \mathbf{Y}\mathbf{Y_N}^H}_{=0} + \mathbf{Y_N}\mathbf{Y_N}^H. \tag{9}$$

Because the elements of $\mathbf{Y_N}$ are considered uncorrelated, the last term of this equation is a diagonal matrix. Thus, by removing the diagonal from $\mathbf{S_{xy}}$, the resulting matrix does not contain any contribution from the noise. It also does not contain the diagonal elements of $\mathbf{YY}^H$. These are the auto spectra of the sound pressure at each microphone. They contain no information about the location of the sound sources, but only about the amplitude. As the main part of this information is also available from the non-diagonal elements, it is feasible to perform the beamforming with the diagonal removed and take this as a good estimate of the true result[2]:

$$S(\omega)_j = \frac{\mathbf{e}(\overrightarrow{x_0})^H \mathbf{W}\,(\mathbf{S_{xy}} - \operatorname{diag}\mathbf{S_{xy}})\,\mathbf{W}^H \mathbf{e}(\overrightarrow{x_0})}{(\operatorname{trace}\mathbf{W})^2 - \operatorname{trace}\mathbf{W}}. \tag{10}$$

The removal of the main diagonal is one method to improve the beamforming result. Another method that is aimed especially on the improvement of source separation, is the orthogonal beamforming.

Take (8) as basis of the calculation of the cross spectral matrix:

$$\mathbf{S_{xy}} = \mathbf{YY}^H = \mathbf{fQQ}^H\mathbf{f}^H. \tag{11}$$

As the sources are assumed to be uncorrelated, the product $\mathbf{QQ}^H$ is a diagonal matrix of rank j. Thus, j (the number of sources) is also the rank of $\mathbf{S_{xy}}$ if there are less sources than microphones. From this follows that the cross spectral matrix may be expressed as a sum of partial cross spectral matrices.

$$\mathbf{S_{xy}} = \sum_{j=1}^{N} \mathbf{S_{xy}}_j. \tag{12}$$

These matrices are constructed from $\mathbf{QQ}^H$ with all but one diagonal element removed and correspond to the j independent source mechanisms. As the matrix $\mathbf{f}$ is unknown, the $\mathbf{S_{xy}}_j$ have to be estimated by an appropriate decomposition of $\mathbf{S_{xy}}$. Then, with

$$S(\omega)_j = \frac{\mathbf{e}(\overrightarrow{x_0})^H \mathbf{W}\left(\mathbf{S_{xy}}_j - \operatorname{diag}\mathbf{S_{xy}}_j\right)\mathbf{W}^H \mathbf{e}(\overrightarrow{x_0})}{(\operatorname{trace}\mathbf{W})^2 - \operatorname{trace}\mathbf{W}} \tag{13}$$

a map for each source mechanism j can be constructed separately. Thus, the limited dynamic range due to the SNR has no influence and 'ghost images' from a dominating mechanism do not mask the contribution from minor sources.

5 Example results

For the practical example explained below, 32 microphones were arranged as a planar array in two circles with different diameters around the same center. Each circle incorporated 16 microphones. The outer circle had a diameter of 1.30 meters.

It is well known that a flow around a cylinder generates noise with a tonal character. The tonal frequency coincides with the vortex shedding frequency of the von Kármán vortex street

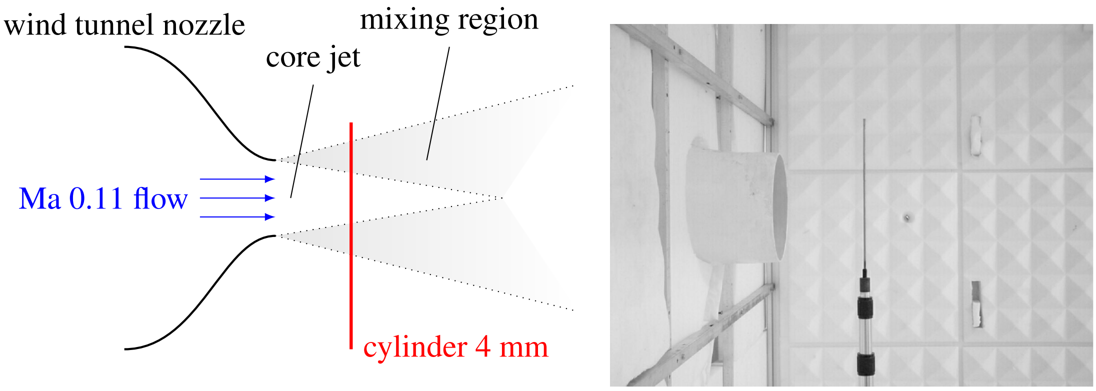

Figure 1: Experimental setup: a cylinder in an open jet

downstream of the cylinder. In an experiment a cylinder of 4 mm diameter was placed in the open jet of an aeroacoustic wind tunnel. In the core of the jet the flow is not as turbulent as in the mixing region of the jet boundaries. It can be expected that the degree of upstream turbulence affects the noise generation. In particular, the noise generated at the cylinder in the jet core will be more tonal and of more power. The noise generated in the mixing region will be more broadband. The noise from both regions can be thought of as two independent and thus uncorrelated sources.

A measurement using the classical beamforming technique shows that as expected the cylinder is a strong noise source. In Fig. 2 the result is shown for the 2 kHz and 8 kHz octave band. The result at 2 kHz already shows a good dynamic range (see also Fig. 3), so the removal of the main diagonal results in only minor improvements. At 8 kHz the removal of the main diagonal yields a considerable improvement.

With the classical beamforming, it is not possible to separate the two source mechanisms at 2 kHz, but only at 8 kHz. Using the orthogonal beamforming even at 2 kHz the sources can be clearly separated. Note that the source from the mixing region makes only a minor contribution to the overall result. Thus, the level is so low that this source can not be noticed in the classical beamforming result.

6 Summary

Due to the limited spatial resolution and dynamic range of the map of sound pressure contributions, the separation of sound sources may pose a problem in the practical application of beamforming. If the properties of the sound field and the unwanted noise at the beamformer input are considered, appropriate signal processing methods can be developed to improve the result.

References

[1] C. Schulze, E. Sarradj and A. Zeibig: Characteristics of Microphone Arrays. Proc. Inter-Noise 2004, Prague

[2] T. Brooks, W. Humphreys: Effect of Directional Array Size On The Measurement of Airframe Noise Components. AIAA Paper No. 99-1958, 1999

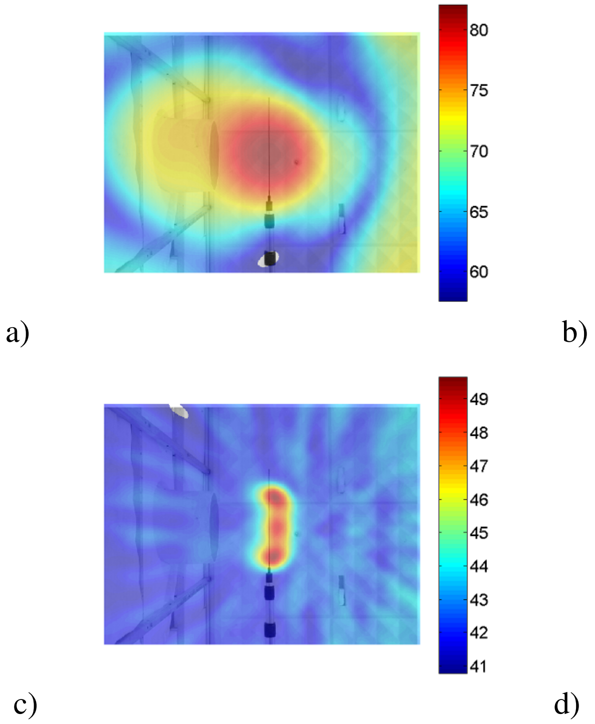

Figure 2: a) beamforming result map for the 2 kHz octave band plotted with 25 dB dynamic range, b) same as a) but with main diagonal of cross spectral matrix removed, c) same as a) but for 8 kHz and with 9 dB dynamic range d) same as c) but with main diagonal of cross spectral matrix removed

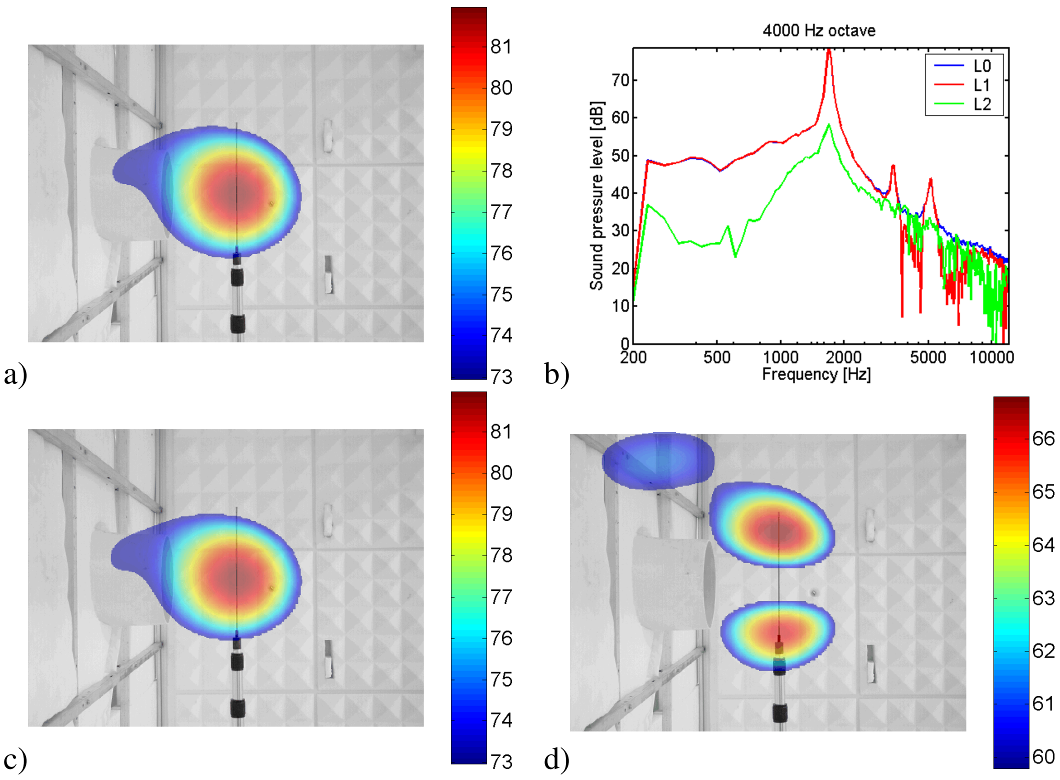

Figure 3: a) beamforming result map for the 2 kHz octave band (same as Fig. 2, but with limited dynamic range), b) spectra of the overall result (L0) and the two first components (L1,L2) of the orthogonal beamforming result, c) first and d) second component of orthogonal beamforming result map for the 2 kHz octave band