A ROBUST DATA-INDEPENDENT NEAR-FIELD BEAMFORMER FOR LINEAR MICROPHONE ARRAYS

F. Borra, L. Bianchi, F. Antonacci, S. Tubaro, A. Sarti

Dipartimento di Elettronica, Informazione e Bioingegneria – Politecnico di Milano via Ponzio 34/5 - 20133 Milano, Italia

ABSTRACT

This manuscript presents a robust data-independent near-field beamformer using a linear microphone array. We consider a scenario with a desired sound source and an interferer; our goal in this manuscript is to provide a technique to extract the desired source signal while attenuating the interferer. We formulate the beamformer design problem as a convex optimization problem with additional constraints that aim at controlling the spatial response of the beamformer, with emphasis on the robustness of the approach to errors in the localization of the interferer sound source. We validate our approach through simulations.

Index Terms — Microphone array, near-field beamforming, constrained optimization, source extraction

1. INTRODUCTION

Extraction of desired acoustic signals in presence of noise and interferences has been a challenging problem for the signal processing community for years [1]. A classical example is the separation of audio sources observed in a real room, known as a cocktail party environment, where a number of people is talking concurrently [2]. A prominent class of techniques that aim at separating the sources is based on microphone arrays and the design of a spatial filter, which combines microphone signals so that components related to the desired signal are constructively added, while others are attenuated.

Conventional techniques [3, 4] rely on the assumption that sources are in the far field, so that all wave fronts impinging on the array are assumed to be planar. However, this assumption is invalid when we consider acoustic sources in the proximity of the array. Here the wave-front curvature can not be neglected [5]. In particular, a source should be considered to be in the near field of the microphone array when its distance from the latter’s center is less than \(2\ell^2/\lambda\), where \(\ell\) is the array length and \(\lambda\) the wavelength [6]. One approach to deal with the near-field effect of microphone arrays is to effectively remove the wave-front curvature by applying a gain and phase compensation to each microphone signal. This compensation makes a curved wave-front appear as a plane wave allowing standard far-field design techniques to be used [6, 7]. Another option presented in [8, 9] is to divide the array into smaller subarrays and assume the wave-front to be planar at each sub-array.

In this contribution we propose a data independent near-field beamformer that considers the array as a whole, thus following the paradigm in [6, 7]. The only information required to design the spatial filter is the location of desired and interferer sources. This information can be estimated using several methods available from the recent literature on acoustic signal processing: from sound field image [8] to particle filtering [10, 11, 12, 13, 14] and to methods in the space range [15, 16]. However, the estimated locations may be affected by errors, thus degrading the separation capabilities of the beamformer. In order to attenuate the impact of localization errors, the design criteria of the spatial filter include three additional linear constraints: i) the spatial filter response must be unitary in the position of the desired signal; ii) the spatial filter response has to be minimal in the interference signal’s position; iii) the spatial derivative of the filter response at the interferer location must be null. We demonstrate through simulations that the robustness of the beamformer against localization errors of the interferer is increased. We do so by comparing the proposed method with the near-field beamformer in [7], which does not impose the constraints ii and iii.

The rest of the paper is structured as follows. Sec. 2 introduces the array signal model used throughout the work and formulates the near-field beamformer design problem. In Sec. 3 we derive the beamforming filter. Sec. 4 features a simulative validation of the proposed beamformer. Finally, Sec. 5 draws some conclusions.

2. SIGNAL MODEL AND PROBLEM STATEMENT

Consider a uniform linear microphone array of \(N\) microphones deployed on the \(y\) axis, as depicted in Fig. 1, where the origin of the reference frame coincides with the center of the array, so that the \(n\)th microphone position is \(\mathbf{m}_n = [0, m_{y,n}]^T = [0, (n - N/2)d]^T\), where \((\cdot)^T\) denotes the transposition operator and \(d\) is the microphone spacing. We consider a simple scenario where the array senses the acoustic field generated by two acoustic sources: one is the desired source and is located, in polar coordinates, at \(r_d = \rho_d[\cos(\theta_d), \sin(\theta_d)]\); the other is the interferer and it is located, in polar coordinates,
The signal (2) can be modeled as the superposition of three components: i) the contribution from the desired source; ii) the contribution from the interfering source; iii) the additive sensor noise term. In mathematical form we have

\[ x(\tau, \omega) = \mathbf{g}^{(d)}(\omega) S^{(d)}(\tau, \omega) + \mathbf{g}^{(i)}(\omega) S^{(i)}(\tau, \omega) + \mathbf{v}(\tau, \omega), \]

where \( S^{(d)}(\tau, \omega) \) denotes the short-time Fourier transform of the desired / interfering signal and \( \mathbf{v}(\tau, \omega) \) is additive noise. The statistics of which are assumed to be time-invariant. The vectors

\[ \mathbf{g}^{(d)}(\omega) = [g_{1}^{(d)}(\omega), \ldots, g_{N}^{(d)}(\omega)]^{T} \in \mathbb{C}^{N \times 1} \]

\[ \mathbf{g}^{(i)}(\omega) = [g_{1}^{(i)}(\omega), \ldots, g_{N}^{(i)}(\omega)]^{T} \in \mathbb{C}^{N \times 1} \]

denote the time-invariant acoustic transfer function between the desired / interfering source and the microphones. In this work we assume free space, near field propagation, hence the acoustic transfer function between a source in \( r_{(d)} \) and the \( n \)th microphone takes the form \([6, \text{Eq. (1)}]\)

\[ g_{n}^{(d)}(\omega) = \frac{\rho_{(d)}(\omega)}{||r_{(d)} - m_{n}||} e^{-j \frac{c}{||r_{(d)} - m_{n}||} \left( ||r_{(d)} - m_{n}|| - \rho_{(d)}(\omega) \right)}, \]

where \( c \) is the speed of sound.

Our goal is to estimate \( S^{(d)}(\tau, \omega) \) from (3) while minimizing the impact of interference and noise signals. We aim at doing so in a data-independent fashion, in such a way that the obtained spatial filter is independent on the statistics of array data. Moreover, we want our filter to be robust to errors in the estimation of the interferer location.

### 3. Near-Field Beamformer Design

The design criteria stated at the end of Sec. 2 can be formulated in a formal way as a convex optimization problem. Our goal is to design a spatial filter \( \mathbf{h}(\omega) \) such that the signal at the output of the filter approximates \( S^{(d)}(\tau, \omega) \), i.e.,

\[ y(\tau, \omega) = \mathbf{h}^{H}(\omega) \mathbf{x}(\tau, \omega) \approx S^{(d)}(\tau, \omega), \]

where \((\cdot)^{H}\) denotes Hermitian transpose. For this purpose, we impose that the filter \( \mathbf{h}(\omega) \), for each frequency \( \omega \), minimizes the power of the filter output subject to the following constraints: i) the filter must leave undistorted signals coming from the desired source location; ii) the filter must attenuate signals coming from the interferer location; iii) the first derivative of the filter response with respect to the source position must be zero in the interferer location. In mathematical terms, this means that

\[ \mathbf{h}_{0}(\omega) = \arg \min_{\mathbf{h}} \mathbf{h}^{H}(\omega) \mathbf{\Phi}_{xx}(\tau, \omega) \mathbf{h}(\omega) \]

subject to \( \mathbf{h}^{H}(\omega) \mathbf{B}(\omega) = \mathbf{f} \),

where \( \mathbf{\Phi}_{xx}(\tau, \omega) \) denotes the auto-covariance matrix of the array signals in (3), and all the constraints have been written in the matrix form \( \mathbf{h}^{H}(\omega) \mathbf{B}(\omega) = \mathbf{f} \). The matrix \( \mathbf{B}(\omega) \) takes the form

\[ \mathbf{B}(\omega) = \begin{bmatrix} \mathbf{g}^{(d)}(\omega), \mathbf{g}^{(i)}(\omega), \frac{\partial \mathbf{g}^{(d)}(\omega)}{\partial \rho} \bigg|_{r=r}, \frac{\partial \mathbf{g}^{(d)}(\omega)}{\partial \theta} \bigg|_{r=r} \end{bmatrix} \]

where we have expressed the derivative of the steering vector with respect to the source position in polar coordinates, i.e., \( r = \rho \cos(\theta), \sin(\theta) \). Following the definition in (8), the constraints vector is given by

\[ \mathbf{f} = [1, \varepsilon, 0, 0], \]

where \( \varepsilon \in \mathbb{R} \) is the amplitude of the spatial response in the interferer location. The explicit expression for the derivatives of the propagation function with respect to \( \rho \) and \( \theta \) in (8) is

\[ \frac{\partial g_{n}(\omega)}{\partial \rho} = \frac{g_{n}(\omega)}{r_{1}} \]

\[ \frac{\partial g_{n}(\omega)}{\partial \theta} = \frac{g_{n}(\omega) r_{1} \cos(\theta_{1})}{||r_{1} - m_{n}||} \left( j \frac{\omega}{c} ||r_{1} - m_{n}|| + 1 \right) \]

The solution to (7) is known from the literature and it is given by \([17, \text{p. 354}]\)

\[ \mathbf{h}_{0}(\omega) = \mathbf{\Phi}_{xx}^{-1}(\tau, \omega) \mathbf{B}(\omega) \left( \mathbf{B}^{H}(\omega) \mathbf{\Phi}_{xx}^{-1}(\tau, \omega) \mathbf{B}(\omega) \right)^{-1} \mathbf{f}^{H}. \]
In order to obtain a solution that is independent on the statistics of array data, we assume a diagonal structure for the auto-covariance matrix, i.e. \( \Phi_{xx}(\tau, \omega) = I_N \), \( I_N \) being the identity matrix of dimensions \( N \times N \).

In the following section, we will investigate the role played by the three constraints individually. For this purpose, besides \( h_0 \), we also define the filters \( h_1 \) and \( h_2 \), which account only for the first; and the first and the second criteria, respectively. These filters can be obtained from (12) by using the first and the first two columns of \( B \) and \( f \), respectively, i.e.

\[
\begin{align*}
    h_1 &= \Phi_{xx}^{-1}B(:,1) \left( B_H(:,1) \Phi_{xx}^{-1}B(:,1) \right)^{-1} f_H(:,1) \\
    h_2 &= \Phi_{xx}^{-1}B(:,1:2) \left( B_H(:,1:2) \Phi_{xx}^{-1}B(:,1:2) \right)^{-1} f_H(:,1:2),
\end{align*}
\]

where we omitted the dependency on \( (\tau, \omega) \) for the sake of compactness. In (13) and (14), \( B(:,a:z) \) denotes the columns from \( a \) to \( z \) of the matrix \( B \) and the same notation holds for \( f \). Notice that \( h_1(\omega) \) is the same filter derived in [7, p. 24], which is a near-field extension of the classical beamforming filter [17, Chap. 6].

4. VALIDATIONS

For the simulations we use the setting introduced in Sec. 2, with \( N = 16 \) microphones spaced by \( d = 10 \text{ cm} \). We place the sources at \( r_1 = [56 \text{ cm}, 30^\circ] \) and \( r_2 = [56 \text{ cm}, -30^\circ] \). We consider a sensor noise \( v(\tau, \omega) \) with a power level such that the SNR between the signal emitted by the desired source and the noise at the 8th microphone is 30 dB. In the following, we consider a set of frequencies \( \omega_p \), which uniformly sample the interval between \( 0 \text{ rad} \text{s}^{-1} \) and \( \pi F_0 \) with \( P \) samples.

4.1. Performance Metrics

In this section we present the metrics used to assess the proposed filter methodology.

- **Signal–to–Interference–plus–Noise Ratio** (SINR), which evaluates the energy ratio between the extracted signal and the sum of interference plus noise signals. The SINR is defined in [18, p. 731] as a function of frequency. In order to provide consistent results, we evaluate the SINR with a Monte-Carlo simulation using \( R \) realizations. The SINR for a single realization is

\[
    \text{SINR}_r(\omega) = \frac{|S_r^{(d)}(\omega_p)|^2}{|S_r^{(i)}(\omega_p)|^2 + |S_r^{(n)}(\omega_p)|^2},
\]

where \( S_r^{(d)}(\omega_p) \), \( S_r^{(i)}(\omega_p) \), \( S_r^{(n)}(\omega_p) \) are the STFT of desired signal, interference and noise, respectively. We then compute the average across multiple realizations, i.e. \( \text{SINR} = \frac{1}{R} \sum_{r=1}^{R} \text{SINR}_r \).

- **White noise gain** (WNG), which evaluates the amplification of the self-noise [18, p. 65], i.e.

\[
    \text{WNG}(\omega_p) = \frac{|h^H(\omega_p)g_n^{(d)}(\omega_p)|^2}{|h^H(\omega_p)h(\omega_p)|^2}. \quad (16)
\]

- **Mean square error** (MSE), which evaluates the error between the output signal and the desired one, i.e.

\[
    \text{MSE} = \frac{1}{MP} \sum_{m=1}^{M} \sum_{p=1}^{P} |S_d(\tau_m, \omega_p) - y(\tau_m, \omega_p)|^2. \quad (17)
\]

4.2. Simulations

In this section we show the validation through four different simulations. In all simulations the frequency goes beyond the aliasing frequency of the array that, for our setting, is 1.7 kHz and the number of realizations is set to \( R = 100 \).

4.2.1. SINR evaluation

We start examining the behavior of the SINR for our filter project varying the constraints as explained in Sec. 3. In this simulation we set \( F_0 = 8 \text{ kHz} \) and we evaluate the metric up to the temporal Nyquist frequency, i.e. \( F_0/2 \). The desired and interference signals are white noise signals with equal constant PSD. We remark that for this simulation the position of the two sources is assumed known a priori. As can be observed from Fig. 2, the addition of the constraint on the level of the interferer (cfr. \( h_0, h_1 \)) leads to an improvement on the SINR level with respect to \( h_1 \). Moreover, a more stable behavior along the frequency axis can be observed.

4.2.2. Robustness to localization errors

In Sec. 4.2.1 we assumed that the source positions used in the project of the spatial filters perfectly match the actual ones. The next simulation removes this assumption and shows the impact of interferer localization errors on the SINR metric.
for two different filters design: $h_2(\omega_p)$, shown in Fig. 3a and $h_0(\omega_p)$ shown in Fig. 3b. The two sources and the frequency axes are the same described in Sec. 4.2.1. We use a uniformly distributed random variable with increasing standard deviation to simulate localization errors. As we can observe from the comparison between Fig. 3a and Fig. 3b, adding derivative constraints with respect to the interferer location increases the robustness against the localization error. In fact, curves in Fig. 3b are closer to the ideal case of perfect localization for a wider range of frequencies than those in Fig. 3a.

4.2.3. WNG evaluation

The simulative setup is the same to that of Sec. 4.2.1. We compare the results with $h_1(\omega_p)$, which has the maximum white noise gain achievable [19]. We observe from Fig. 4 that $h_2(\omega_p)$ and $h_0(\omega_p)$ exhibit close resemblance to $h_1(\omega_p)$ for a large range of frequencies although at low frequencies the WNG worsens.

4.2.4. MSE evaluation

This simulation aims at providing a quality measure for the filtered signal $y(\tau_m, \omega_p)$. In this simulation we set $F_S = 44.1 \, \text{kHz}$, $P = 1024$ samples. For the STFT computation we use a frame of 1024 samples and a Hanning window with a 75% overlap. We run two different types of simulations to evaluate (17). In the first case we use two speech signals [20, track 49 and 50] for both the desired source and for the interferer; in the second case, we set the interferer as a white noise with the same power at the 8th microphone equal to the one generated by the desired signal. As we can see from Table 1, and in particular in the first case, adding constraints on the level of the interferer leads to appreciable improvements in terms of the MSE. Contrary we can note that the addition of the constraint on the derivatives does not influence the metric significantly. In the second case, we can see that the performances of $h_1$ show an important improvement while for $h_2$ and $h_0$ the results remain almost equal to the first case. This is due to the properties of the filter in (12) that optimally rejects noise-like interferences.

5. CONCLUSION

In this manuscript we have proposed a technique for the extraction of a desired sound source from an acoustic scene captured by a linear microphone array. The filter project was based on a quadratic data-independent optimization problem subject to three linear constraints on the response of the spatial filter at desired and interferer source locations. We have provided numerical simulations to prove the effectiveness of the proposed solution. Results demonstrate that the constraint on the attenuation of the interferer improves the MSE, whereas the constraint on the derivative of the response of the spatial filter at the interferer location increases the robustness against uncertainties in the location of the interferer.
6. REFERENCES


