AN APPROACH TO PREVENT ADAPTIVE BEAMFORMERS FROM CANCELLING THE DESIRED SIGNAL

Tofigh Naghibi and Beat Pfister

Speech Processing Group, Computer Engineering and Networks Lab., ETH Zurich, Switzerland
{naghibi,pfister}@tik.ee.ethz.ch

ABSTRACT

Under real conditions, severe signal cancellation often occurs in adaptive beamformers because of reverberation, microphone transfer function mismatch or steering vector error. Therefore, usually voice activity information is necessary to pause the beamformer update during the speech activity. However, this information is not available or not sufficiently accurate in most applications. Here we propose a new algorithm in order to mitigate the signal cancellation effect. The algorithm extracts the desired source-to-microphones transfer functions from the data covariance matrix by using rough estimates of some source locations. We show that it is robust against reverberation, microphone mismatch and imprecisely estimated target direction.

Index Terms— Microphone array, signal cancellation effect, weighted Procrustes problem, dereverberation

1. INTRODUCTION

Interference that is coherent with the signal of interest often occurs in adaptive microphone arrays because of reverberations, microphone mismatches or error in steering vector. The commonly utilized minimum variance distortionless response (MVDR) beamformer minimizes the output power while maintaining a specified response to the desired signal. However, in the presence of coherent interferers or steering vector error, MVDR is not successful because coherent interferences are used to cancel the desired signal from the output. Many signal processing methods have been proposed to address this problem. These techniques are usually called robust adaptive beamforming. However, this robustness is achieved at the expense of less interference reduction or an increased number of microphones. For example in [5], the signal is averaged over the space to decorrelate the signal and interference. This technique can only be applied to uniform microphone arrays and needs many elements to achieve satisfactory performance. Using norm-constrained adaptive filters was proposed to constrain the power of the signal leakage and therefore improving adaptive beamformers robustness. This approach has been employed in [4] to address the target signal cancellation effect. However, it requires knowledge of the interference covariance matrix which may not be available in speech recognition applications. In [3], both quadratic and non-linear (truncation) constraints in three-block structure have been used to improve the interference reduction. Another approach is to estimate the transfer functions (TFs) with blind source separation techniques. TFs may also be estimated based on non-stationarity of signal and stationarity of noise assumption [2]. In this paper, we extract the TFs from the covariance matrix of the array data given the source locations. The mathematical formulation of this approach leads to a weighted Procrustes problem [6]. Typically, a Procrustes problem is about rotating and scaling a known set of data to fit another set. Here this method is used to estimate those TFs that are close enough to the ideal TFs (steering vectors) and still can reconstruct the data covariance matrix.

2. SIGNAL CANCELLATION

Let $s_m(n)$ be the sound waves emitted by M wide-band sources. They are received by an array of N microphones. The room impulse response $h^{k,i}_{room}(t)$ characterizes both direct and echo paths from the kth source to the ith microphone. Since the microphones may not be calibrated, they may introduce different transfer functions $h^{i}_{mic}(t)$. One can merge the room impulse response and microphone transfer function into a total transfer function $h^{k,i}(t)$ containing both room acoustic and microphone characteristics. Therefore the ith microphone output in the frequency domain can be written as:

$$x_i(f) = \sum_{m=1}^{M} h^{k,i}(f) s_m(f) + n_i(f), \tag{1}$$

where $n_i(f)$ is the additive white noise at the ith microphone. The microphone outputs can be aggregated into a column vector $\mathbf{x}$:

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{n}, \tag{2}$$

where $\mathbf{n}$ is the additive white noise vector, $\mathbf{H} = [\mathbf{h}_1, ..., \mathbf{h}_M]$ the channel matrix, $\mathbf{h}_i = [h^{k,i}, ..., h^{M,i}]^T$ the ith column of $\mathbf{H}$ and $\mathbf{s} = [s_1(f), ..., s_M(f)]$ the source vector. The number of sources M is assumed to be less than the number of microphones N. Without loss of generality, s_1 can be considered

as the desired signal. The goal of the beamformer is to obtain an estimate of the desired signal by filtering and summing the microphone outputs:

$$y(f) = \mathbf{w}(f)^H \mathbf{x}(f), \tag{3}$$

where $(.)^H$ is the Hermitian transpose operator, y the beamformer output and $\mathbf{w}$ the beamformer weight vector. The conventional MVDR beamformer chooses its weight vector $\mathbf{w}$ to minimize the output power while maintaining the signal from a specified direction of arrival:

$$\underset{\mathbf{w}}{\operatorname{argmin}} \quad \mathbf{w}^H \mathbf{R} \mathbf{w} \qquad \text{subject to} \quad \mathbf{w}^H \mathbf{d} = 1 \tag{4}$$

$\mathbf{R} = E\{\mathbf{x}\mathbf{x}^H\}$ is the covariance matrix of the received data $\mathbf{x}$. The target steering vector is defined by $\mathbf{d} = [e^{-j\omega\tau_1}, ..., e^{-j\omega\tau_N}]$, where $\tau_1, ..., \tau_N$ are delays matched to the desired direction. However, the conventional MVDR in (4) works well in very limited situations. In reality, severe signal cancellation occurs because of microphone mismatches, location estimate errors, signal-correlated noise and reverberant environments. To clarify this problem, we have to look more closely at the covariance matrix $\mathbf{R}$:

$$\mathbf{R} = E\{\mathbf{x}\mathbf{x}^H\} = \mathbf{H}\mathbf{R}_s\mathbf{H}^H + \sigma^2\mathbf{I}, \tag{5}$$

where $\mathbf{R}_s = E\{\mathbf{s}\mathbf{s}^H\}$ is the source covariance matrix and σ^2 the additive white noise power. As long as the sound sources are uncorrelated, $\mathbf{R}_s$ is diagonal and (5) can be decomposed into three additive terms:

$$\mathbf{R} = \mathbf{h}_1\mathbf{h}_1^H S_1(f) + \mathbf{H}_{2:M}\mathbf{R}_{s_{2:M}}\mathbf{H}_{2:M}^H + \sigma^2\mathbf{I}, \tag{6}$$

where $S_1(f)$ is the power spectrum of the desired signal and $\mathbf{H}_{2:M} = [\mathbf{h}_2, ..., \mathbf{h}_M]$. $\mathbf{R}_{s_{2:M}}$ can be obtained by removing the first row and the first column of $\mathbf{R}_s$. Substituting $\mathbf{w}^H\mathbf{h}_1 = 1$ in the MVDR objective function (4) yields:

$$\mathbf{w}^H\mathbf{R}\mathbf{w} = S_1(f) + \mathbf{w}^H\mathbf{H}_{2:M}\mathbf{R}_{s_{2:M}}\mathbf{H}_{2:M}^H\mathbf{w} + \sigma^2\mathbf{w}^H\mathbf{w} \tag{7}$$

Ignoring the first term, the MVDR optimization problem can be simplified to:

$$\begin{aligned} \underset{\mathbf{w}}{\operatorname{argmin}} \;& \mathbf{w}^H\mathbf{H}_{2:M}\mathbf{R}_{s_{2:M}}\mathbf{H}_{2:M}^H\mathbf{w} + \sigma^2\mathbf{w}^H\mathbf{w} \\ \text{subject to} \quad & \mathbf{w}^H\mathbf{h}_1 = 1 \end{aligned} \tag{8}$$

Note that (8) is independent of S_1. However, usually in reverberant rooms, no knowledge about $\mathbf{h}_1$ is available in advance. It has to be estimated from the array data or approximated with steering vector $\mathbf{d}$ calculated from the desired source location estimate. Since $\mathbf{h}_1$ includes both direct and indirect paths, it can be decomposed into a sum of steering vector and echos transfer function:

$$\mathbf{h}_1 = \mathbf{d} + \mathbf{h}_{echos} \tag{9}$$

Note that the attenuation factor has been intentionally dropped to avoid notational distraction. $\mathbf{R}$ can be rewritten as follows:

$$\begin{aligned} \mathbf{R} = (\mathbf{h}_{echos} + \mathbf{d})(\mathbf{h}_{echos} + \mathbf{d})^H S_1(f) + \mathbf{H}_{2:M}\mathbf{R}_{s_{2:M}}\mathbf{H}_{2:M}^H \\ + \sigma^2\mathbf{I} \end{aligned} \tag{10}$$

By using $\mathbf{w}^H\mathbf{d} = 1$ as an approximation of the target transfer function $\mathbf{h}_1$, the objective function becomes:

$$|\mathbf{w}^H\mathbf{h}_{echos} + 1|^2 S_1(f) + \mathbf{w}^H\mathbf{H}_{2:M}\mathbf{R}_{s_{2:M}}\mathbf{H}_{2:M}^H\mathbf{w} + \sigma^2\mathbf{w}^H\mathbf{w} \tag{11}$$

Looking closely at this function reveals that the first term can be vanished if $\mathbf{w}^H\mathbf{h}_{echos} = -1$ holds. Therefore the MVDR optimization problem tends to satisfy the $\mathbf{w}^H\mathbf{h}_{echos} = -1$ constraint. However, satisfying this constraint results in removing the signal from the beamformer output. To clarify it, note that the signal component in the beamformer output can be written as $y_s = \mathbf{w}^H\mathbf{h}_1 s_1(f)$. Using (9) and the fact that the weight vector satisfies $\mathbf{w}^H\mathbf{d} = 1$ and $\mathbf{w}^H\mathbf{h}_{echos} = -1$ constraints, the beamformer response to the signal transfer function $\mathbf{h}_1$ is $\mathbf{w}^H\mathbf{h}_1 = 0$ and thus $y_s = 0$. That is, by approximating the real TF with the steering vector $\mathbf{d}$, the MVDR beamformer tends to remove the desired signal. Actually, if we ignore the white noise term for sake of simplicity, by choosing $\mathbf{w}$ from $\mathbf{H}$ null space, the objective function reaches its minimum value, namely zero. It means $\mathbf{w}^H\mathbf{H} = 0$ and therefore $\mathbf{w}^H\mathbf{h}_1 = 0$. One natural solution is to update the sample covariance matrix at speech pauses. Then there will be no $S_1(f)$ in (10) to result in signal cancellation. However, this solution needs accurate voice activity information and also cannot track the non-stationary background noise. The second approach which will be described in the following section, is to estimate the acoustic transfer function and the microphone mismatches. This estimate is then used to achieve the signal-independent optimization problem in (8).

3. TRANSFER FUNCTION ESTIMATION

In the presence of several interferers, we have to estimate $\mathbf{h}_1$ from the corrupted data $\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{n}$. This information can be extracted from the covariance matrix $\mathbf{R}$. Using spectral decomposition, $\mathbf{R}$ can be written as $\mathbf{R} = \mathbf{U}(\mathbf{\Lambda} + \sigma^2\mathbf{I})\mathbf{U}^H$, where the first M columns of $\mathbf{U}$ are the orthonormal basis vectors of the signal and interferences subspace and σ^2 is the white noise power which can be estimated as the average of the $N-M$ smallest eigenvalues of $\mathbf{R}$. Given the source position, the ideal steering vector can be easily calculated. However, because of acoustic characteristics of a room, the steering vector $\mathbf{d}$ may not lie in the subspace spanned by the $\mathbf{U}_{1:M}$ columns. One approach to estimate $\mathbf{h}_1$ is to find in subspace $\mathbf{U}_{1:M}$, the closest vector to $\mathbf{d}$ in the Euclidean sense, i.e.,

$$\underset{\mathbf{x}}{\operatorname{argmin}} \quad ||\mathbf{U}_{1:M}\mathbf{x} - \mathbf{d}||_2^F \tag{12}$$

Consequently, $\mathbf{h}_1 = \mathbf{U}_{1:M}\mathbf{x}$ and it belongs to the subspace $\mathbf{U}_{1:M}$. This is a projection problem and the solution can be written as $\mathbf{h}_1 = \mathbf{U}_{1:M}\mathbf{U}_{1:M}^H\mathbf{d}$, where $P = \mathbf{U}_{1:M}\mathbf{U}_{1:M}^H$ is the projection operator onto the $\mathbf{U}_{1:M}$ subspace. However, although it can be a good guess if the steering vector and the real TF mismatch is fairly small, it may not work in more severe reverberant environments. To go one step further, we assume here the availability of some additional information about the interferer locations. That is, direction of arrival of k interferers ($k \leq M$) can be derived from the array data (which is reasonably simple for at least an imprecise estimate)[1]. Unlike some other methods, they do not necessarily need to be the k strongest sources. Defining $\mathbf{H}' = \mathbf{U}_{1:M}\mathbf{\Lambda}_{M\times M}^{1/2}\mathbf{V}$ and thus $\mathbf{R} = \mathbf{H}'\mathbf{H}'^H$ one can infer that $\mathbf{H}'$ can be estimated up to an unknown multiplicative unitary matrix $\mathbf{V}$ from $\mathbf{R}$ decomposition. Our aim is to estimate this unitary matrix $\mathbf{V}$ and reconstruct $\mathbf{h}_1$ by $\hat{\mathbf{h}}_1 = \mathbf{U}_{1:M}\mathbf{\Lambda}_{M\times M}^{1/2}\mathbf{v}_1$. Taking locations information into account, (12) can be extended as follows:

$$\underset{\mathbf{V},\mathbf{\Gamma}}{\operatorname{argmin}} \quad ||\mathbf{U}_{1:M}\mathbf{\Lambda}^{1/2}\mathbf{V}_{1:k}\mathbf{\Gamma} - \mathbf{D}||_2^F \quad \text{s.t} \quad \mathbf{V}_{1:k}^H\mathbf{V}_{1:k} = \mathbf{I} \tag{13}$$

where $\mathbf{\Gamma}$ is a $k\times k$ diagonal weighting matrix which is necessary to model the unknown attenuation factor. $\mathbf{D} = [\mathbf{d}_1, ..., \mathbf{d}_k]$ is the steering matrix. It is worth to note that in case of $k = 1$, (13) will reduce to the simple least squares problem in (12). However, unlike (12), there is no straightforward solution for (13). The optimization problem in (13) is a linear least squares problem defined on a Stiefel manifold, known as weighted orthogonal Procrustes problem (WOPP). A Stiefel manifold is the set of all $M\times N$ matrices $\mathbf{V}$ having orthonormal columns. Usually, solutions suggested for this problem have two steps: Given $\mathbf{\Gamma}$, they try to find the optimum $\mathbf{V}$ and use it in the second step to find the optimal $\mathbf{\Gamma}$. This forms an iterative solution that converges to a local minimum. Here, we employ the algorithm suggested in [1] with small modifications to work with complex matrices. The complete iterative channel matrix estimation algorithm is listed in Table 1. Simulations have shown that it converges in less than 20 iterations. This algorithm may be computationally expensive. However, it only needs to be run infrequently since the TF does not change rapidly. The output of this algorithm is an estimate of the channel matrix $\mathbf{H}$ which is used in (8) to achieve a signal-independent MVDR beamformer.

4. SIMULATION RESULTS

The adaptive beamformer is implemented in the frequency domain with overlap-add 2048 point FFT filterbank and sampling frequency of $f_s = 44100$ Hz. The non-uniform linear array consists of 8 microphones as depicted in Figure 1. For

[1] Since many systems nowadays use both audio and video channels to ease the human-machine interaction, like Microsoft's Kinect-Xbox, the location information could also come from the vision channel.

$\mathbf{A} = [\mathbf{a}_1, ..., \mathbf{a}_k] = \mathbf{\Lambda}^{1/2}\mathbf{U}_{1:M}^H\mathbf{D}$

$\mathbf{V}_0 = \mathbf{I}$

for each t = 0, ...

$\mathbf{Z} = [\mathbf{z}_1, ..., \mathbf{z}_k] = \mathbf{V}_t$

$\alpha_i = \dfrac{\mathbf{a}_i^H\mathbf{z}_i}{\mathbf{z}_i^H\mathbf{\Lambda}\mathbf{z}_i} \quad i = 1, .., k$

$\mathbf{C} = [\mathbf{c}_1, ...\mathbf{c}_k], \quad \mathbf{c}_i = \alpha_i\mathbf{a}_i + |\alpha_i|^2(\rho\mathbf{I} - \mathbf{\Lambda})\mathbf{z}_i \quad i = 1, ..., k$

$\mathbf{C} = \mathbf{U}_c\mathbf{\Lambda}_c\mathbf{V}_c^H$

$\mathbf{V}_{t+1} = \mathbf{U}_c\mathbf{V}_c^H \quad \mathbf{V}_{t+1} = [\mathbf{v}_1, ..., \mathbf{v}_k]$

$\Phi(t+1) = 2\sum_{i=1}^{k} real(\alpha_i\mathbf{v}_i^H\mathbf{a}_i) - \sum_{i=1}^{k} |\alpha_i|^2\mathbf{v}_i^H\mathbf{\Lambda}\mathbf{v}_i$

terminate if $\Phi(t+1) - \Phi(t) \approx 0$

Table 1: Iterative channel matrix estimate algorithm

evaluation of the proposed algorithm, two different simulation cases have been chosen.

Case I: In the first scenario, two speech signals in a reverberant room with $T_{60} = 100$ ms (T_{60} is the reverb time) have been assumed. The origin of the coordinate system is at the center of the microphone array. The desired speaker and the interferer are placed at $[0, 2.5, 0]$ and $[3, 2.5, 0]$ in meter, respectively. A mismatch of less than 2 dB is assumed between the microphone transfer functions. Beam pattern can be defined as $\text{BP}(\mathbf{h}(f)) = |\mathbf{w}^H\mathbf{h}(f)|^2$ and beam pattern values at $\mathbf{h} = \mathbf{h}_1$ (desired signal transfer function) and $\mathbf{h} = \mathbf{h}_2$ (interferer) can be interpreted as the beamformer responses to the signal and interference, respectively. These responses are shown in Figure 2 for the proposed, the conventional MVDR and the super-directive beamformers. Super-directive beamformers assume diffuse noise which is a very common model for reverberant environments. Note that $\text{BP}(\mathbf{h}_1)$ can be interpreted as $\frac{SNR_{out}}{SNR_{in}}$ and $\text{BP}(\mathbf{h}_2) = \frac{INR_{out}}{INR_{in}}$. Therefore, beam pattern can be seen as a performance evaluation measure. The average of $\frac{\text{BP}(\mathbf{h}_2(f))}{\text{BP}(\mathbf{h}_1(f))}$ over all frequencies is called relative mean attenuation and is shown with a horizontal dotted line for all beamformers in Figure 2. It reveals that the fixed super-directive beamformer can attenuate the interference up to 15 dB in average while the MVDR with updated constraint with our algorithm, can achieve 7 dB more interference reduction. Also, the beamformer response to $\mathbf{h}_1$ shows less fluctuations

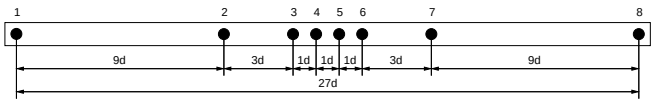

Fig. 1: Linear microphone array with eight non-uniformly spaced microphones at $d = 3$ cm.

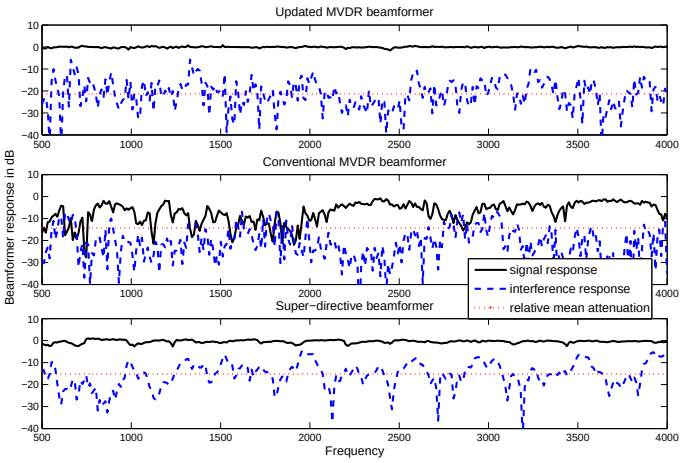

Fig. 2: Proposed MVDR (top), conventional MVDR (middle) and super-directive beamformer responses to the signal and the interferer.

and thus less signal distortion for the updated MVDR than the super-directive beamformer. However, as expected, the conventional MVDR response to the target TF shows that it severely distorts the desired signal.

Case II: In the second scenario, the traditional and the proposed MVDR beamformers' performances in the presence of large target steering vector error have been studied. Figure 3 shows the beam pattern of the MVDR beamformer with updated constraint and the traditional MVDR beamformer at 1 kHz. A steering vector error of 15° has been assumed for the signal and the interferer. Therefore both columns of the steering matrix in (13) are imprecise. Nevertheless, as it can be seen in Figure 3, the updated MVDR achieves both robustness against 15° steering vector error and high interference reduction which is around 30 dB at interference direction. The solid line also reveals a slight shift in the null position (from 40° to 33° which leads to about 10 dB degradation in interference reduction (from 40 dB to 30 dB). More experiments have shown that the proposed algorithm is robust against even larger target direction errors at the expense of this shift in null position which can be seen as a trade-off between the noise reduction and robustness. The proposed algorithm prevents signal attenuation by shifting the main lobe to the correct azimuth while the traditional MVDR performs dramatically worse. The main

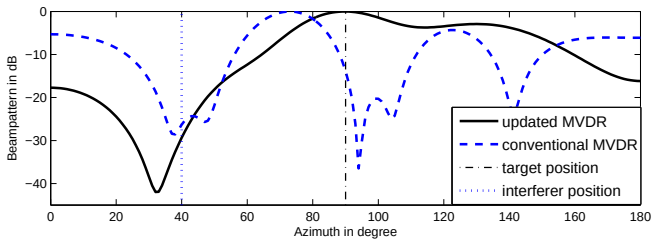

Fig. 3: Beam pattern of the updated constraint MVDR and the traditional MVDR at f=1000 Hz. Vertical lines mark the source and the interferer positions.

advantage of this method over other similar methods of robust constraint set design like [7] is that it doesn't widen the main lobe but shifts it to the correct angle by extracting covariance matrix information.

5. CONCLUSION

We proposed a new solution to overcome the signal cancellation effect caused by reverberation or target direction error. The proposed algorithm uses additional information about source locations to achieve a precise estimate of the transfer functions relating the source of interest to the microphones. The proposed algorithm shows a high robustness in the presence of reverberation or target direction error, where the performance of the conventional MVDR beamformer is unacceptable. However, the proposed algorithm may degrade in heavy reverberation environments. A possible solution for the heavy reverberation case is to replace the steering vectors with more informative vectors that reflect some of the room acoustic characteristics.

6. ACKNOWLEDGMENT

This work has been supported by Swiss National Science Foundation (SNSF).

7. REFERENCES

[1] M. B. Dosse. Anisotropic orthogonal Procrustes analysis. *Journal of Classification*, 27, 2010.

[2] S. Gannot, D. Burshtein, and E. Weinstein. Signal enhancement using beamforming and nonstationarity with applications to speech. *IEEE Trans. on Signal Processing*, 49:1614 –1626, 2001.

[3] O. Hoshuyama, A. Sugiyama, and A. Hirano. A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters. *IEEE Trans. on Signal Processing*, 47:2677–2684, 1999.

[4] F. Qian and B. Van Veen. Quadratically constrained adaptive beamforming for coherent signals and interference. *IEEE Trans. on Signal Processing*, 43:1890–1900, 1995.

[5] T. J. Shan and T. Kailath. Adaptive beamforming for coherent signals and interference. *IEEE Trans. on ASSP*, vol. 33:527–536, 1985.

[6] T. Viklands. *Algorithms for the Weighted Orthogonal Procrustes Problem and other Least Squares Problems*. PhD thesis, Umeå University, Umeå, Sweden, 2006.

[7] Y. Zheng, P. Xie, and S. Grant. Robustness and distance discrimination of adaptive near field beamformers. *ICASSP Proceedings*, 12:478–488, 2006.