A MULTI-MICROPHONE BEAMFORMER WITH AN EIGENDOMAIN POST-FILTER

Firas Jabloun and Benoît Champagne

Department of Electrical & Computer Engineering, McGill University
3480 University Street, Montreal, Canada, H3A 2A7
firas@tsp.ece.mcgill.ca, champagne@ece.mcgill.ca

ABSTRACT

In a previous paper, we proposed a new method to extend the single microphone signal subspace approach , to a multi-microphone design. This method is based on a delay-and-sum beamformer followed by a post-filter in the eigen domain. The post-filter is designed using a composite covariance matrix obtained from the outputs of all the channels. In this paper, we propose a second method for the post-filter design. The new method consists of estimating an average covariance matrix from the noisy covariance matrices of every microphone signal. We present a comparison between these methods in terms of their performance under adverse environments. The trivial approach of summing signals, of every channel, enhanced separately using a different signal subspace filter is also considered. Experimental results using different objective measures are reported.

1. INTRODUCTION

Single microphone speech enhancement techniques rely on noise statistics gathered during non-speech activity periods. Errors in noise parameters estimation result in the annoying residual noise known as musical noise. One of the proposed ways to overcome this artifact is to use more than one microphone for noise reduction.

Microphone arrays improve the speech quality by rejecting interfering signals coming from directions different from a desired look direction. Adaptive beamforming has been widely used to reduce the interference level by adaptively steering zeros in the direction of the interfering signals [1, 2]. These methods are efficient when the number of noise sources is smaller than the number of sensors but they fail to provide a good performance in reverberant rooms or when the interfering signals are correlated with the desired signal. In such conditions conventional adaptive beamforming methods exhibit cancellation of the desired signal [3].

Alternatively, a conventional delay-and-sum beamformer followed by a post-filter for noise reduction has been proposed. The noisy signals at the different channels are used to compute the parameters of the post-filter which can take different forms, for example a Wiener filter [4, 5] or a coherence filter [6].

In [7] we presented a similar post-filter designed using the signal subspace approach. The designed filter was based on a composite covariance matrix which allows to estimate the speech signal subspace by averaging in the eigendomain. In this paper we present a second method to estimate the eigendomain post-filter parameters. This method uses averaging in the time domain to approximate the signal covariance matrix. The eigenvalue decomposition of this estimated matrix is used to obtain the post filter parameters.

This paper is organized as follows. In Section 2 the signal subspace approach for speech enhancement is briefly presented. In Section 3 the multi-microphone approach is introduced and the methods used to calculate the eigendomain postfilter are described. Experimental results are reported in Section 4 and a conclusion is given in Section 5.

2. THE SIGNAL SUBSPACE APPROACH

In this section we briefly present the signal subspace approach for speech enhancement. More details can be found in [8].

Let $\mathbf{x} = \mathbf{s} + \mathbf{w}$ be a P-dimensional noisy observation vector where $\mathbf{s}$ is the desired vector and $\mathbf{w}$ is the noise vector with covariance matrix $\mathbf{R}_w$. The eigenvalue decomposition (ED) of the covariance matrix $\mathbf{R}_s$ of the clean vector is given by $\mathbf{R}_s = \mathbf{U}\mathbf{\Lambda}_s\mathbf{U}^T$ where $\mathbf{\Lambda}_s = \mathrm{diag}(\lambda_{s_1}, \ldots, \lambda_{s_P})$ with the eigenvalues λ_{s_i}'s in decreasing order. In this paper, we also assume that the noise is white with $\mathbf{R}_w = \sigma^2\mathbf{I}$ so that $\mathbf{R}_x$, the covariance matrix of $\mathbf{x}$, will have the same eigenvectors as $\mathbf{R}_s$. We also assume that $\mathrm{rank}(\mathbf{R}_s) = I < P$ so that $\lambda_{s_i} = 0$ for $i = I+1, \ldots, P$. Hence $\mathbf{U}$ can be written as $\mathbf{U} = [\mathbf{U}_1\mathbf{U}_2]$ where $\mathbf{U}_1$ spans the so-called signal subspace and $\mathbf{U}_2$ spans the noise subspace.

We want to find a linear estimate of $\mathbf{s}$ given by $\hat{\mathbf{s}} = \mathbf{H}\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{H}\mathbf{w}$. The residual error signal is given by

$$\mathbf{r} = \hat{\mathbf{s}} - \mathbf{s} = (\mathbf{H} - \mathbf{I})\mathbf{s} + \mathbf{H}\mathbf{w} = \mathbf{r}_s + \mathbf{r}_w \qquad (1)$$

In the spectral domain constrained approach (SDC), the enhancement filter $\mathbf{H}$ is obtained by minimizing the signal

distortion

$$\min_{\mathbf{H}} \quad E\{\|\mathbf{r}_s\|^2\} \tag{2}$$

subject to

$$E\{|\mathbf{u}_i^H \mathbf{r}_w|^2\} \leq \alpha_i \sigma^2 \quad \text{for} \quad i = 1, \ldots, I \tag{3}$$

which ensures that the i^{th} spectral component of the residual noise is below some threshold. Here $\mathbf{u}_i$ is the i^{th} eigenvector of $\mathbf{R}_s$ with eigenvalue λ_{s_i}. The solution to this problem is given by [8]

$$\mathbf{H} = \mathbf{U}_1 \mathbf{Q} \mathbf{U}_1^T \tag{4}$$

where $\mathbf{Q}$ is a $I \times I$ diagonal gain matrix with entries

$$q_i = \alpha_i^{1/2} = e^{-\nu\sigma^2/\lambda_{s_i}} \quad \text{for} \quad i = 1, \ldots, I. \tag{5}$$

3. THE MULTI-MICROPHONE APPROACH

We now show how this approach can be extended to a multi-microphone design. Suppose we have M microphones for signal acquisition followed by a time delay compensation module to ensure that all microphone signals are correctly synchronized. Under these conditions, we have

$$\mathbf{x}_m = \mathbf{s} + \mathbf{w}_m \quad \text{for} \quad m = 1, \ldots, M. \tag{6}$$

where $\mathbf{x}_m$ and $\mathbf{w}_m$ are the corrupted speech vector and the noise vector at the m^{th} microphone respectively, $\mathbf{s}$ is the desired speech vector.

Now as in [4] and [6], we assume that the noise and reverberation form a diffuse acoustic field. Therefore these perturbations, in addition to being uncorrelated with the direct path signal, are considered to be incoherent at different microphones. These assumptions coincide with real life applications when the microphones are close to the speaker relative to the interfering sources like a car engine or air conditioning noise inside a room. Hence the covariance matrix of the input signal at two particular microphones with index i and j is given by

$$\mathbf{R}_{ij} = E\{\mathbf{x}_i \mathbf{x}_j^T\} = \mathbf{R}_s + \sigma^2 \delta(i-j)\mathbf{I} \tag{7}$$

where $\mathbf{I}$ is the identity matrix.

To obtain the enhanced signal, a different eigenfilter $\mathbf{H}_m$ can be designed for every microphone as in the single microphone approach. The average of the M filter outputs is taken to yield the enhanced signal,

$$\hat{\mathbf{s}} = \frac{1}{M} \sum_{m=1}^{M} \mathbf{H}_m \mathbf{x}_m \tag{8}$$

We will refer to this method as SSM where the "M" denotes the number of microphones used. So SS1 refers to the single channel signal subspace method while SS4 means that 4 microphones were used.

Now in the post-filter approach, the signals in every microphone are used to design a single post-filter $\mathbf{H}$ and the enhanced signal is obtained as follows

$$\hat{\mathbf{s}} = \frac{1}{M} \mathbf{H} \sum_{m=1}^{M} \mathbf{x}_m \tag{9}$$

In addition to the Composite Covariance matrix (CCM) method proposed in [7], we provide, in this paper, a second design method which we refer to as the Averaged Covariance Matrix (ACM) method. The two methods differ in the way the ED of $\mathbf{R}_s$ is approximated. The so obtained eigenvalues $\hat{\lambda}_{s_i}$'s and eigenvectors $\hat{\mathbf{u}}_i$'s (for $i = 1 \ldots I$) are used to calculate the post-filter $\mathbf{H}$ according to Equations (4) and (5).

Experimental results involving objective measures and informal listening tests are provided to demonstrate the superiority of CCM and ACM over SS1 and SS4. These results also show that the ACM method does not perform as good as CCM but it has the advantage of having much less computational load.

3.1. The Composite Covariance Matrix (CCM) Method

In the CCM method [7], we define a combined vector $\mathbf{y}$ of length $D = MP$, by stacking the individual input vectors of every microphone in the following way

$$\mathbf{y} = [\mathbf{x}_1^T, \ldots, \mathbf{x}_M^T]^T \tag{10}$$

Then the overall composite covariance matrix $\mathbf{R}_y = E\{\mathbf{y}\mathbf{y}^T\}$ can be written as

$$\mathbf{R}_y = \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} & \cdots & \mathbf{R}_{1M} \\ \mathbf{R}_{21} & \mathbf{R}_{22} & \cdots & \mathbf{R}_{2M} \\ \vdots & & \ddots & \vdots \\ \mathbf{R}_{M1} & \cdots & \cdots & \mathbf{R}_{MM} \end{bmatrix} \tag{11}$$

Then given the eigen-pair $(\lambda_{y_i}, \mathbf{q}_i)$ for $i = 1, \ldots, P$ of $\mathbf{R}_y$, the eigenvalue λ_{s_i} and the corresponding unit-norm eigenvector $\mathbf{u}_i$ of $\mathbf{R}_s$ can be approximated as follows [7]:

$$\hat{\lambda}_{s_i} = \frac{\lambda_{y_i} - \sigma^2}{M} \tag{12}$$

$$\hat{\mathbf{u}}_i = \frac{1}{\sqrt{M}} \sum_{m=1}^{M} \mathbf{q}_{mi} \tag{13}$$

where $\mathbf{q}_i = [\mathbf{q}_{1i}^T, \ldots, \mathbf{q}_{Mi}^T]^T$, and $\mathbf{q}_{mi}$'s $(m = 1, \ldots, M)$ are P-dimensional vectors. The rank I is chosen to be the number eigenvalues such that $\hat{\lambda}_{s_i} > 0$.

3.2. The Averaged Covariance Matrix (ACM) Method

In the second proposed method, the eigen-pairs of $\mathbf{R}_s$ are approximated in a different way. The average of the individual covariance matrices in every channel, as defined in (7) for $i = j = m$, is calculated as follows

$$\bar{\mathbf{R}}_x = \frac{1}{M} \sum_{m=1}^{M} \mathbf{R}_{mm} \tag{14}$$

Now if λ_{x_i} and $\mathbf{u}_{x_i}$ are the i^{th} eigen-pair of $\bar{\mathbf{R}}_x$ then we have

$$\hat{\lambda}_{s_i} = \lambda_{x_i} - \sigma^2 \tag{15}$$

$$\hat{\mathbf{u}}_i = \mathbf{u}_{x_i} \tag{16}$$

The advantage of this method is its computational savings. In ACM, the ED of one single $P \times P$ covariance matrix needs to be computed. In CCM, however, the ED of an $MP \times MP$ matrix is required while SS4 needs the ED of a $P \times P$ matrix to be performed four times.

3.3. Implementation

For the ACM and the CCM methods the covariance matrices are approximated as follows

$$\hat{\mathbf{R}}_{ij} = \text{toeplitz}\{\hat{r}_{ij}(0), \ldots, \hat{r}_{ij}(P-1)\} \tag{17}$$

where the cross-correlation function $\hat{r}_{ij}(p)$ is calculated using N observations from microphones i and j as follows

$$\hat{r}_{ij}(p) = \frac{1}{N} \sum_{n=0}^{N-1-|p|} x_i(n) x_j(n-p) \quad \text{for} \tag{18}$$

In [5] the coherence of interfering signals at two different microphones is reduced by taking the real part of the cross-power spectrum. Following a similar approach, the coherence is reduced here by taking the even part of the cross-correlation $\hat{r}_{ij}(p)$ as follows

$$\hat{r}_{ij}(p) = \frac{1}{2}[\hat{r}_{ij}(p) + \hat{r}_{ij}(-p)] \tag{19}$$

To reduce the computational load the enhancement process is performed on a frame-by-frame basis as described in [9]. Briefly, this approach consists of dividing the speech signal into overlapping frames of length N. The samples of this frame are used to approximate one single eigenfilter used to enhance all P-dimensional vectors within that frame.

4. EXPERIMENTAL RESULTS

In the following results, $M = 4$ microphones were used. Computer generated white noise is added at every microphone. The enhanced signals are obtained using SS1, SS4, ACM and CCM with $P = 32$ and $N = 256$ at 8 KHz sampling rate.

Informal listening tests show that the postfilter methods outperform SS1 and SS4. While ACM and CCM have comparable performances relative to that of SS4 and SS1, signals enhanced with CCM are more natural and the residual noise is less prominent. The advantage of ACM, however, is its relatively lower computational complexity.

Two objective measures were used for performance evaluation. The Segmental SNR and the Perceptual Evaluation of Speech Quality (PESQ) score [10]. The PESQ score compares the enhanced signal with a reference signal (the clean speech signal) and provides a value consistent with the Mean Opinion Score (MOS) used to quantify the performance of speech signal algorithms.

The Segmental SNR used is defined as follows

$$\text{SegSNR}(j) = 10 \log_{10} \frac{E\{|s_j(n)|^2\}}{E\{|s_j(n) - \hat{s}_j(n)|^2\}} \tag{20}$$

where $s_j(n)$ and $\hat{s}_j(n)$ are the clean and enhanced signals in the j^{th} frame respectively. The average SegSNR for all, length 64, frames (with 50% overlap) was used for evaluation. The reported results are the average results of 5 runs of the experiment with different noise signals at every run.

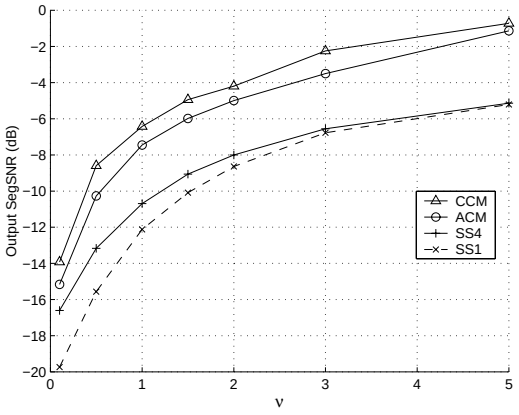

Fig. 1. Output SegSNR for different values of ν.

Figure 1 shows the output segmental SNR for a 0 dB global SNR noisy speech signal, for different values of ν. One important observation here is that for $\nu > 3$, SS1 and SS4 have almost the same performance. This can be explained by the fact that due to over-suppression of noise (with the expense of more signal distortion), no more gain can be achieved by taking the average output. However using PESQ score [10], shown in Figure 2, we can see that the use of M microphones improves the overall performance.

The improvement in performance achieved using the postfilter approach (ACM and CCM) is obvious with both objective measures. It can also be seen that CCM outperforms

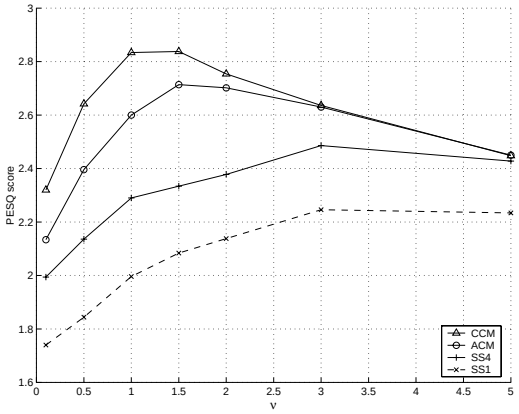

Fig. 2. PESQ scores for different values of ν.

ACM and that the gain achieved by CCM over ACM is comparable to the gain achieved by SS4 over SS1. We also conclude from these results (together with the informal listening experiments) that the optimum value for ACM and CCM is around 1 and around 2 for SS1 and SS4. These values are used in the next experiment.

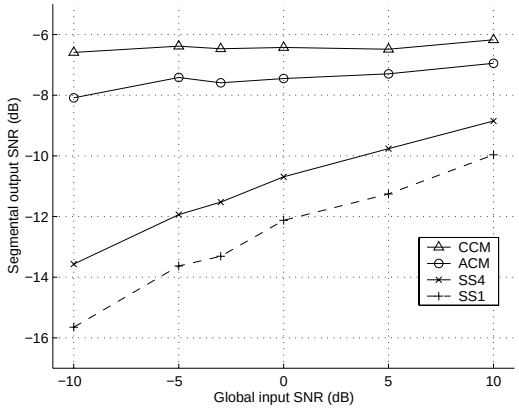

Fig. 3. Output SegSNR for different global input SNR's

In Figure 3, the output SegSNR's for different global input SNR's are shown. Again it can be seen that the CCM outperforms the other methods and that it has a relatively constant performance under different noise levels.

5. CONCLUSION

In this paper, we presented a microphone array delay-and-sum beamformer with an eigendomain post-filter. We proposed a new method for the post-filter design and compared its performance with a previously presented method. Experimental results show the superiority of these two methods over the straight forward solution of taking the average of separately enhanced signals. The postfilter approach have a relatively constant performance under different noise levels according to the segmental SNR objective measure. The advantage of the newly presented method is its low computational load while having a comparable performance to the previously proposed more complex method.

6. REFERENCES

[1] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. on Antennas and Propagation*, vol. 30, pp. 27–34, January 1982.

[2] Y. Kaneda, "Adaptive microphone array system for noise reduction (AMNOR) and its performance studies," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 34, pp. 1391–1400, December 1986.

[3] B. Widrow, K. M. Duvall, R. P. Gooch, and W. C. Newman, "Signal cancellation phenomena in adaptive antennas: Causes and cures," *IEEE Trans. on Antennas and Propagation*, vol. 30, pp. 469–478, 5 1982.

[4] C. Marro, Y. Mahieux, and K. U. Simmer, "Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering," *IEEE Trans. on Speech and Audio Processing*, vol. 6, pp. 240–259, May 1998.

[5] R. Zelinsky, "A microphone array with adaptive post-filtering for noise reduction in reverberant rooms," *in Proc ICASSP88*, pp. 2578–2580, 1988.

[6] R. Le Bouquin-Jeannes, A. A. Azirani, and G. Faucon, "Enhancement of speech degraded by coherent and incoherent noise using a cross-spectral estimator," *IEEE Trans. on Speech and Audio Processing*, vol. 5, pp. 484–487, November 1997.

[7] F. Jabloun and B. Champagne, "A multi-microphone signal subspace approach for speech enhancement," *in Proc ICASSP01*, vol. 1, pp. 205–208, 2001.

[8] Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech enhancement," *IEEE Trans. on Speech and Audio Processing*, vol. 3, pp. 251–266, July 1995.

[9] F. Jabloun and B. Champagne, "On the use of masking properties of the human ear in the signal subspace speech enhancement approach," *in Proc IWAENC, Darmstadt*, pp. 199–202, 2001.

[10] "Perceptual evaluation of speech quality (PESQ)." Recommendation ITU-T P.862, International Telecommunication Union, February 2001.