MICROPHONE ARRAYS FOR NOISE REDUCTION WITH LOW SIGNAL DISTORTION IN ROOM ACOUSTICS

Mehrez Souden, Jacob Benesty, and Sofiène Affes
INRS-ÉMT, 800, de la Gauchetière Ouest, Suite 6900, Montréal, H5A 1K6, Qc, Canada.
{souden,benesty,affes}@emt.inrs.ca

ABSTRACT

We propose a new method for noise reduction using a microphone array. The method takes advantage of the spatial diversity inherent to microphone arrays and optimizes certain criteria, namely, the output signal to noise ratio (SNR) or the mean squared error (MSE), subject to the constraint of spatial prediction that relates the noise free signals captured by the microphones. Simulation results demonstrate that resorting to this new method leads to high rate of noise reduction and low signal distortion.

Index Terms— Noise reduction, microphone array, spatial prediction, speech distortion, Wiener filter.

1. INTRODUCTION

Noise reduction has become an active area of research after the pioneering work of Schroeder [1]. This fact is due to its various applications including hand-free communications, hearing aids, teleconferencing, etc. [2].

So far, several noise reduction techniques have been proposed. The first and most popular ones were developed in the case of a single microphone and in the presence of an additive noise only. These techniques have been classified into three main classes [3]: spectral substraction, statistical-model-based, and subspace-decomposition-based. Unfortunately, noise reduction comes at the price of significant speech distortion in these techniques [4] because of the utilization of a single microphone. Microphone arrays have, however, theoretically the potential to reduce the noise while keeping the speech signal undistorted. Several works have been also carried out to enhance speech signals (dereverberation and denoising) using microphone arrays as in [5]. However, the resulting complexity therein is prohibitive. Beamforming techniques [6, 7] such as the generalized side lobe canceller (GSC) [8] have also the potential to perform this task by steering the array beam toward the direction of arrival of the source. But, these techniques are sensitive to reverberation and calibration errors [2]. Actually, reverberation itself remains a complicated task and one would rather focus on denoising only. In [9], for example, Doclo and Moonen generalized the single microphone noise reduction subspace-based techniques to the multichannel case by utilizing the so-called generalized singular value decomposition (GSVD). In [10], this multichannel GSVD-based technique has been incorporated in a GSC-type structure to reduce its complexity. In all of these techniques, a very important feature which is the spatial predictability of the speech signal captured by the microphone array has not been considered.

The very basic idea of the proposed approach is to take into account the spatial predictability of the speech components perceived by the microphones for the design of a denoising filter. Indeed, the utilization of multiple microphones renders the speech signal spatially predictable. In other words, in theory any of the noise free speech components received by one microphone can be obtained by interpolating any of the other noise free signals captured by another microphone. In this paper, we mathematically formulate this aspect and deduce two optimal filters that can achieve noise reduction with low speech distortion. Namely, these two filters consist in output SNR maximization and MSE minimization under the constraint of speech spatial predictability.

2. PROBLEM STATEMENT AND ASSUMPTIONS

Let $s(t)$ denote a speech signal impinging on an array of N microphones with an arbitrary geometry. The resulting observations are given by:

$$\begin{aligned} y_n(t) &= s(t) * g_n + v_n(t) \\ &= x_n(t) + v_n(t);\ n = 1, 2, \cdots, N, \end{aligned} \tag{1}$$

where $*$ is the convolution operator, g_n is the channel impulse response encountered by the source when impinging on the nth microphone, $x_n(t) = s(t) * g_n$ is the noise free speech component, and $v_n(t)$ is the noise at microphone n [the noise can be colored and is uncorrelated with $s(t)$]. We assume that all the noise components and $s(t)$ are zero-mean random processes and that all the involved entities are real valued.

The objective of this work is to recover one of the speech signal components, say $x_1(t)$ without loss of generality, the best way we can by either maximizing the output SNR or minimizing the MSE under some constraint of low speech distortion. Notice here that we are only interested in noise reduction and not in speech dereverberation which goes out of the scope of this paper [5]. Since we use the first microphone signal as a reference, we define the input SNR as:

$$\text{SNR} = \frac{E\{x_1^2(t)\}}{E\{v_1^2(t)\}} = \frac{\sigma_{x_1}^2}{\sigma_{v_1}^2}. \tag{2}$$

We aim at finding a linear filter, $\mathbf{h}$, of length L that will be applied to the observed signals to obtain:

$$z(t) = \mathbf{h}^T\mathbf{y}(t) = \mathbf{h}^T\mathbf{x}(t) + \mathbf{h}^T\mathbf{v}(t), \tag{3}$$

where $\mathbf{x}(t) = [\mathbf{x}_1^T(t)\ \mathbf{x}_2^T(t)\ \cdots\ \mathbf{x}_N^T(t)]^T$, $\mathbf{x}_n(t) = [x_n(t)\ x_n(t-1)\ \cdots\ x_n(t-L+1)]^T$ $(n = 1, 2, \cdots, N)$, and so are defined $\mathbf{y}(t)$ and $\mathbf{v}(t)$. At the output of this filter, the SNR is given by:

$$\text{SNR}(\mathbf{h}) = \frac{E\left\{\left[\mathbf{h}^T\mathbf{x}(t)\right]^2\right\}}{E\left\{\left[\mathbf{h}^T\mathbf{v}(t)\right]^2\right\}} = \frac{\mathbf{h}^T\mathbf{R}_{xx}\mathbf{h}}{\mathbf{h}^T\mathbf{R}_{vv}\mathbf{h}}. \tag{4}$$

In this paper, $\mathbf{R}_{da}$ denotes the correlation matrix of two random vectors $\mathbf{d}$ and $\mathbf{a}$. Our aim is to find optimal filters that provide us with the highest noise reduction and the lowest signal distortion simultaneously.

3. SPATIAL PREDICTION CONSTRAINT

It is well known that the speech signal is temporally partially predictable. Similarly, the utilization of an array of microphones in room acoustics makes the received signal spatially predictable. Indeed, the speech signal captured by any of the microphones $2, \cdots, N$ can be predicted from the one captured by the first microphone. In other words, for any $n \in \{1, 2, \cdots, N\}$, there exists an $L \times L$ matrix $\mathbf{W}_n$ such that:

$$\mathbf{x}_n(t) \approx \mathbf{W}_n \mathbf{x}_1(t) \tag{5}$$

with, of course, $\mathbf{W}_1 = \mathbf{I}$ ($\mathbf{I}$ is the identity matrix). Defining $\mathbf{W} = [\mathbf{W}_1^T \ \mathbf{W}_2^T \ \cdots \mathbf{W}_N^T]$, we can write

$$\mathbf{x}(t) \approx \mathbf{W}^T \mathbf{x}_1(t). \tag{6}$$

This relation is very important and will be used as a constraint to minimize the speech distortion while reducing the noise. Now, how to calculate $\mathbf{W}$? To this end, we minimize the following MSE criterion:

$$J(\mathbf{W}) = E\left\{ \left[\mathbf{W}^T \mathbf{x}_1(t) - \mathbf{x}(t)\right]^T \left[\mathbf{W}^T \mathbf{x}_1(t) - \mathbf{x}(t)\right] \right\}. \tag{7}$$

Straightforward calculations lead to the optimal filter:

$$\mathbf{W}_{\text{o}} = \mathbf{R}_{x_1 x_1}^{-1} \mathbf{R}_{x_1 x}. \tag{8}$$

In practice, $\mathbf{R}_{xx}$ is not available (so are $\mathbf{R}_{x_1 x_1}$ and $\mathbf{R}_{x_1 x}$, the first $L \times L$ and $L \times NL$ block matrices extracted from $\mathbf{R}_{xx}$, respectively), but can be estimated if the noise is stationary enough (its second order statistics do not change much with time). Indeed, using a voice activity detector, one can estimate $\mathbf{R}_{vv}$ during the periods of silence and use it during periods of speech jointly with the fact that $\mathbf{R}_{xx} = \mathbf{R}_{yy} - \mathbf{R}_{vv}$. Having $x_1(t)$ as a reference signal, a natural choice of the MSE is the following [9]:

$$\begin{aligned} J(\mathbf{h}) &= E\{[z(t) - x_1(t)]^2\} \\ &= \sigma_{e_x}^2 + \sigma_{e_v}^2, \end{aligned} \tag{9}$$

where $z(t)$ is defined in (3) and

$$\sigma_{e_x}^2 = E\{e_x^2(t)\} = E\left\{ \left[\mathbf{h}^T \mathbf{x}(t) - x_1(t)\right]^2 \right\}, \tag{10}$$

$$\sigma_{e_v}^2 = E\{e_v^2(t)\} = E\left\{ \left[\mathbf{h}^T \mathbf{v}(t)\right]^2 \right\}. \tag{11}$$

Ideally, we would like to have $e_x(t) = 0$ and $e_v(t) = 0$. Unfortunately, this is not the case, and any noise reduction leads to speech distortion in practice. For a given filter $\mathbf{h}$, the signal distortion is given by:

$$e_x(t) = \mathbf{h}^T \mathbf{x}(t) - x_1(t) \approx \mathbf{h}^T \mathbf{W}_{\text{o}}^T \mathbf{x}_1(t) - \mathbf{u}_1^T \mathbf{x}_1(t), \tag{12}$$

where $\mathbf{u}_1 = [1 \ 0 \ \cdots \ 0]^T$ is an $L-$dimensional vector. The approximation above is obtained by taking into account (6) and (8). Now, (12) can be easily rewritten as:

$$e_x(t) \approx (\mathbf{W}_{\text{o}} \mathbf{h} - \mathbf{u}_1)^T \mathbf{x}_1(t). \tag{13}$$

Hence, by imposing the constraint:

$$\mathbf{W}_{\text{o}} \mathbf{h} = \mathbf{u}_1 \tag{14}$$

while minimizing the MSE or maximizing the output SNR, we expect to obtain minimum signal distortion. This approach will lead to two new filters as explained below.

4. WIENER FILTER WITH SPATIAL PREDICTION CONSTRAINT

The classical Wiener filter is obtained by minimizing the MSE in (9):

$$\mathbf{h}_{\text{W}} = \arg\min_{\mathbf{h}} J(\mathbf{h}) \tag{15}$$

which leads to:

$$\mathbf{h}_{\text{W}} = \mathbf{R}_{yy}^{-1} \mathbf{R}_{xx_1} \mathbf{u}_1. \tag{16}$$

As stated previously, $\mathbf{R}_{xx_1}$ can be estimated if the noise is stationary enough. However, we would like to take into account the constraint (14) while minimizing the MSE. Namely, we are interested in solving this optimization problem:

$$\mathbf{h}_{\text{CW}} = \arg\min_{\mathbf{h}} J(\mathbf{h}) \tag{17}$$

$$\text{s.t.} \quad \mathbf{W}_{\text{o}} \mathbf{h} = \mathbf{u}_1. \tag{18}$$

The Lagrangian is then given by:

$$\mathcal{L}(\mathbf{h}, \lambda) = J(\mathbf{h}) + \lambda^T \mathbf{W}_{\text{o}} \mathbf{h} - \lambda^T \mathbf{u}_1. \tag{19}$$

Setting the derivative of this function with respect to $\mathbf{h}$ to zero leads to:

$$\mathbf{h}_{\text{CW}} = \mathbf{R}_{yy}^{-1} \left(\mathbf{R}_{xx_1} \mathbf{u}_1 + \frac{1}{2} \mathbf{W}_{\text{o}}^T \lambda \right). \tag{20}$$

Using the constraint (18), we find:

$$\lambda = 2 \left(\mathbf{W}_{\text{o}} \mathbf{R}_{yy}^{-1} \mathbf{W}_{\text{o}}^T \right)^{-1} \left(\mathbf{W}_{\text{o}} \mathbf{R}_{yy}^{-1} \mathbf{R}_{xx_1} - \mathbf{I} \right) \mathbf{u}_1. \tag{21}$$

Using (6), (20), and (21), we obtain:

$$\mathbf{h}_{\text{CW}} = \mathbf{R}_{yy}^{-1} \mathbf{W}_{\text{o}}^T \left(\mathbf{W}_{\text{o}} \mathbf{R}_{yy}^{-1} \mathbf{W}_{\text{o}}^T \right)^{-1} \mathbf{u}_1. \tag{22}$$

5. MAXIMUM OUTPUT SNR WITH SPATIAL PREDICTION CONSTRAINT

Here, our aim is to maximize $\text{SNR}(\mathbf{h})$ defined in (4) under the constraint of spatial prediction (6). The problem can be formulated as:

$$\mathbf{h}_{\text{CS}} = \arg\max_{\mathbf{h}} \text{SNR}(\mathbf{h}) \tag{23}$$

$$\text{s.t.} \quad \mathbf{W}_{\text{o}} \mathbf{h} = \mathbf{u}_1. \tag{24}$$

The Lagrangian can be written as:

$$\mathcal{L}_{\text{SNR}}(\lambda, \mathbf{h}) = \frac{\mathbf{h}^T \mathbf{R}_{xx} \mathbf{h}}{\mathbf{h}^T \mathbf{R}_{vv} \mathbf{h}} + \lambda^T \mathbf{W}_{\text{o}} \mathbf{h} - \lambda^T \mathbf{u}_1. \tag{25}$$

Setting the derivative of the above function to zero, we find:

$$\frac{2}{(\mathbf{h}_{\text{CS}}^T \mathbf{R}_{vv} \mathbf{h}_{\text{CS}})} \left[\mathbf{M}(\mathbf{h}_{\text{CS}})\right] \mathbf{h}_{\text{CS}} + \mathbf{W}_{\text{o}}^T \lambda = \mathbf{0}, \tag{26}$$

where

$$\mathbf{M}(\mathbf{h}_{\text{CS}}) = \mathbf{R}_{xx} - \text{SNR}(\mathbf{h}_{\text{CS}}) \mathbf{R}_{vv}. \tag{27}$$

When considering the maximization of the output SNR only, the optimal filter is the eigenvector associated to the largest eigenvalue of $\mathbf{R}_{vv}^{-1} \mathbf{R}_{xx}$. The corresponding output SNR is equal to this eigenvalue. However, the resulting signal distortion is very high. If $\text{SNR}(\mathbf{h}_{\text{CS}})$ is equal to any of the eigenvalues of $\mathbf{R}_{vv}^{-1} \mathbf{R}_{xx}$, $\mathbf{h}_{\text{CS}}$ is

an associated eigenvector and the effect of the constraint (24) disappears from (26). Hence, $\mathbf{M}(\mathbf{h}_{\mathrm{CS}})$ is invertible and

$$\mathbf{h}_{\mathrm{CS}} = -\frac{(\mathbf{h}_{\mathrm{CS}}^T\mathbf{R}_{vv}\mathbf{h}_{\mathrm{CS}})}{2}\left[\mathbf{M}(\mathbf{h}_{\mathrm{CS}})\right]^{-1}\mathbf{W}_{\mathrm{o}}^T\lambda. \quad (28)$$

Using the constraint (24), we obtain:

$$-\frac{\mathbf{h}_{\mathrm{CS}}^T\mathbf{R}_{vv}\mathbf{h}_{\mathrm{CS}}}{2}\mathbf{W}_{\mathrm{o}}\left[\mathbf{M}(\mathbf{h}_{\mathrm{CS}})\right]^{-1}\mathbf{W}_{\mathrm{o}}^T\lambda = \mathbf{u}_1 \quad (29)$$

meaning that

$$\lambda = -\frac{2}{\mathbf{h}_{\mathrm{CS}}^T\mathbf{R}_{vv}\mathbf{h}_{\mathrm{CS}}}\left\{\mathbf{W}_{\mathrm{o}}\left[\mathbf{M}(\mathbf{h}_{\mathrm{CS}})\right]^{-1}\mathbf{W}_{\mathrm{o}}^T\right\}^{-1}\mathbf{u}_1. \quad (30)$$

We obtain:

$$\mathbf{h}_{\mathrm{CS}} = \left[\mathbf{M}(\mathbf{h}_{\mathrm{CS}})\right]^{-1}\mathbf{W}_{\mathrm{o}}^T\left\{\mathbf{W}_{\mathrm{o}}\left[\mathbf{M}(\mathbf{h}_{\mathrm{CS}})\right]^{-1}\mathbf{W}_{\mathrm{o}}^T\right\}^{-1}\mathbf{u}_1. \quad (31)$$

Using (31), we propose an iterative approach to calculate $\mathbf{h}_{\mathrm{CS}}$ and the corresponding output SNR. Indeed, we first use an initial guess of the output SNR, say $\hat{\mathrm{SNR}}(\mathbf{h}_{\mathrm{CS}})$, and we deduce the corresponding filter $\hat{\mathbf{h}}_{\mathrm{CS}}$ using (31). After that, we use $\hat{\mathbf{h}}_{\mathrm{CS}}$ to calculate the output SNR and iterate few times to obtain the optimal filter and the corresponding output SNR. Moreover, notice that (31) involves the inversion of the matrix $\mathbf{M}(\mathbf{h}_{\mathrm{CS}})$ defined in (27). Remarkably, this matrix depends on the energy of the speech signal which might drop to very low values in noise-dominated frames [3]. This makes the inversion of $\mathbf{M}(\mathbf{h}_{\mathrm{CS}})$ a delicate task because it becomes badly conditioned and might lead to high signal distortions as we empirically found. Therefore, we will use the filter (31) with moderate to high energy frames only as we explain in Section 6.

6. SIMULATIONS

In this section, we provide some numerical examples to corroborate the potential of the proposed approaches. We are interested in providing the usefulness of the relation (6) which translates into the constraint (24) or (18). Hence, we put aside the problem of noise-second-order-statistics estimation and suppose that they are known for any processed data frame. For further details, we refer the readers to [3], Chapter 9, where noise estimation algorithms are investigated.

The starting point in the derivation of the spatial prediction matrix $\mathbf{W}$ is the data model (1) which assumes that the noise and speech are both present. However, the speech signal is not stationary and its energy may go to zero in some data frames. Consequently, the classical Wiener filter is used instead of the two proposed filters if the SNR< -15 dB. As we stated in Section 5, the filter (31) is used with moderate to high energy frames only (with $\mathrm{SNR} \geq 10$ dB here). For $-15 \leq \mathrm{SNR} < 10$ dB, we use the filter (22) instead. To sum up, three methods are compared, namely, *Method 1*: Wiener filter (16), *Method 2*: constrained Wiener filter (22) for $\mathrm{SNR} \geq -15$ dB and Wiener filter otherwise, and *Method 3*: constrained SNR maximization filter (31) for $\mathrm{SNR} \geq 10$ dB, constrained Wiener filter (22) for $-15 \leq \mathrm{SNR} < 10$ dB and Wiener filter otherwise. In practice, the exact value of the SNR for a given data frame is not available but can be estimated; see [9, 10, 11] for example.

We use some impulse responses that are measured at the Bell-Labs varechoic room. The simulated room dimensions[1] are: length $= 6.7$, width $= 6.1$, and height $= 2.9$ ($x \times y \times z$). We consider a uniform linear array of N microphones which are placed on the axis ($y_{\mathrm{m}} = 5.6$, $z_{\mathrm{m}} = 1.4$) with the first microphone at the coordinate $x_{\mathrm{m},1} = 2.437$ along the x-axis and the microphones spacing is $\Delta = 0.1$. The source is a 4 seconds-long male speech taken from the noisy speech corpus NOIZEUS [3] sampled at 8 kHz and located at ($x_{\mathrm{s}} = 1.337$, $y_{\mathrm{s}} = 3.162$, $z_{\mathrm{s}} = 1.6$). The reverberation condition is set to $T_{60} \approx 240$ ms. The N simulated impulse responses (4096 filter taps each) are convolved with the speech signal before adding a computer generated white Gaussian noise with a long-term input SNR $= 10$ dB in both simulated scenarios. The perceived signals are cut into small rectangular frames with 64 ms duration each and 50% overlapping. After processing them, each frame is multiplied by a Hamming window and overlap added to the other processed frames [3]. As performance indices we use the output SNR defined in (4) and the log-likelihood ratio [3] between the original speech and the filtered one, $\mathbf{h}^T\mathbf{x}(k)$, as a measure of speech distortion.

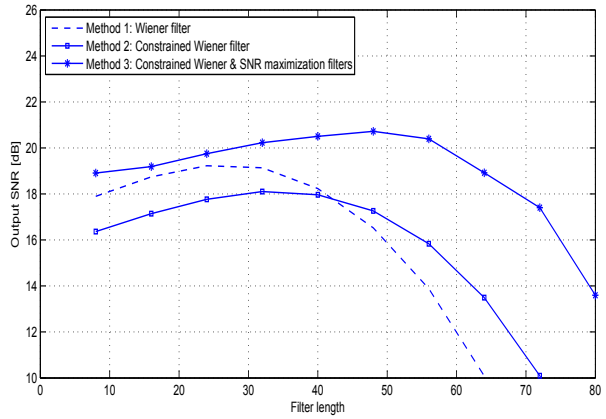
In the first simulation setup, we chose $N = 6$ and vary the number of filter taps L. The results are depicted in Fig. 1. We notice that the spatial prediction constraint remarkably reduces the signal distortion with both Wiener and output SNR maximization filters (unconstrained output SNR maximization is discarded from our comparisons as it introduces extremely high signal distortions). This comes at the price of lower output signal to noise ratio for the constrained Wiener filter (for moderate filter lengths). *Method 3*, however, provides the best output signal to noise ratio and comparable signal distortion especially when $L \approx 50$. As L becomes very high, the performance of all the filters collapse because the matrices involved in the calculations of $\mathbf{h}_{\mathrm{W}}$, $\mathbf{h}_{\mathrm{CW}}$, and $\mathbf{h}_{\mathrm{CS}}$ become large and ill conditioned.

In the second simulation setup, we choose $L = 16$ and increase the number of microphones from 2 to 10. Here, it is worth mentioning that the spatial prediction model is valid for $N \geq 2$ while the Wiener filter is applicable even when $N = 1$. In this case, we empirically found that the Wiener filter gives an output SNR ≈ 15.3 dB and the log-likelihood ratio is around 0.3. The results of this simulation are provided in Fig 2. We notice that, as expected, the output SNR increases with the number of microphones for the three methods. However, the signal becomes more and more distorted with the new proposed filters (especially when N increases from 2 to 6). This fact is due to the spatial predictability constraint which is imperfect. Indeed, recall that in (5), we are interpolating $\mathbf{x}_1(k)$ to obtain $\mathbf{x}_n(k)$ (both vectors have the same length L). Therefore, $\mathbf{x}_n(k)$ has some unpredictable entries (at its end) because of the time delay propagation between sensors 1 and n. When n increases, the unpredictable part of $\mathbf{x}_n(k)$ becomes larger due to the increase of the propagation time delay. We conclude that longer filters need to be used to take advantage of the proposed spatial prediction constraint when arrays of larger sizes are deployed. However, one has to pay attention to the potential deterioration of the conditioning of the covariance matrices which are involved in the calculations of the proposed filters.

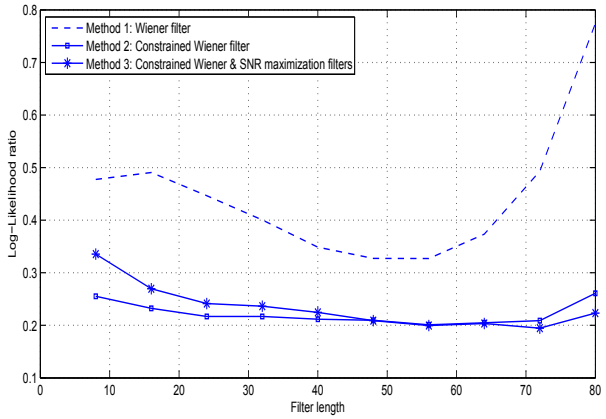
7. CONCLUSIONS

In this paper, we explored a new concept of spatial predictability of the speech signal received by a microphone array. This predictability was then exploited as constraint to optimize two criteria, namely, SNR maximization and MSE minimization. We empirically found that the constrained SNR maximization provides more SNR gains while the constrained MSE minimization provides less speech distortion especially for filters of moderate sizes. As the filter length increases, both filters provide comparable signal distortions with different output SNR values.

[1] All dimensions and coordinates are in meters.

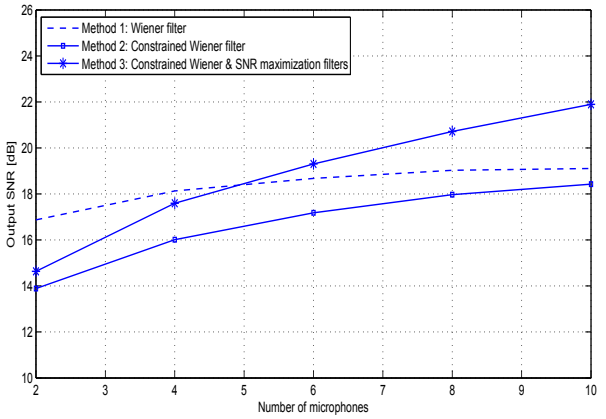


(a) Output SNR vs. L

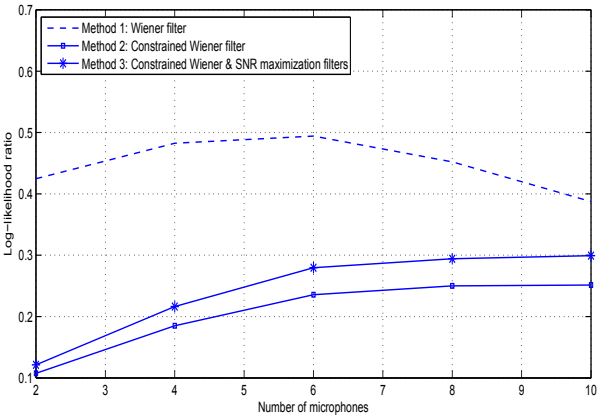


(b) Log-Likelihood ratio vs. L

Fig. 1. Filter length (L) effect; $T_{60} \approx 240$ ms and $N = 6$.



(a) Output SNR vs. N



(b) Log-Likelihood ratio vs. N

Fig. 2. Number of microphones (N) effect; $T_{60} \approx 240$ ms and $L = 16$.

8. REFERENCES

[1] M. R. Schroeder, "Apparatus for supressing noise and distortion in communication signals," U.S. Patent No 3,180,936,filed Dec. 1, 1960, issued Apr. 27, 1965.

[2] J. Benesty, S. Makino, and J. Chen, editors, *Speech Enhancement*. Springer-Verlag, Berlin, Germany, 2005.

[3] P. C. Loizou, *Speech enhancement: Theory and Practice*. CRC Press, New York, USA, 2007.

[4] J. Chen, J. Benesty, Y. Huang, and S. Doclo, "New insights into the noise reduction Wiener filter," *IEEE Trans. Audio, Speech, Language Process.,* vol. 14, pp. 1218-1234, July 2006.

[5] M. Delcroix,T. Hikichi, and M. Miyoshi,"Dereverberation and denoising using multichannel linear prediction," *IEEE Trans. Audio, Speech, Language Process.,* vol. 15, pp. 1791-1801, Aug. 2007.

[6] B. D. Van Veen and K. M. Buckley ,"Beamforming: a versatile approach to spatial filtering," *IEEE Audio, Speech, Signal Process. Mag.,* vol. 5, pp. 4-24, Apr. 1988.

[7] J. Benesty, J. Chen, Y. Huang, and J. Dmochowski, "On microphone-array beamforming from a MIMO acoustic signal processing perspective," *IEEE Trans. Audio, Speech, Language Process.,* vol. 15, pp. 1053-1065, Mar. 2007.

[8] L. J. Griffiths and C. W. Jim,"An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas Propagat.* vol. AP-30, pp. 27-34, Jan. 1982.

[9] S. Doclo and M. Moonen,"GSVD-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Trans. Signal Process.,* vol. 50, pp. 2230-2244, Sept. 2002.

[10] S. Doclo and M. Moonen,"Multimicrophone noise reduction using recursive GSVD-based optimal filtering with ANC post-processing," *IEEE Trans. Audio, Speech, Language Process.,* vol. 13, pp. 53-69, Jan. 2005.

[11] Y. Hu and P. C. Loizou,"A generalized subspace approach for enhancing speech corrupted by colored noise," *IEEE Trans. Speech Audio Process.*, vol. 11, pp. 334-341, Jul. 2003.