A MICROPHONE ARRAY FOR SPEECH ENHANCEMENT USING MULTIRESOLUTION WAVELET TRANSFORM

Djamila Mahmoudi

Signal Processing Laboratory, Swiss Federal Institute of Technology at Lausanne
CH-1015 Lausanne, Switzerland. e-mail:mahmoudi@lts.de.epfl.ch

ABSTRACT

This paper addresses the problem of enhancing a speech signal corrupted by interfering signals. A new noise reduction algorithm based on a logarithmic microphone array and the multiresolution wavelet transform is described. The proposed processing is applied in the time-spectral domain with respect to the logarithmic subband decomposition of the spectrum of each microphone signal. The advantage of the proposed method is that both the sub-array based beamforming operation and the postfiltering are performed in the same transform domain without adding FFT processing.
Computer simulation results show that our approach is effective for noise reduction. The technique can be used in hands-free voice communication applications operating in an adverse environment. In particular, it can be applied to improve speech signal pick-up for voice communication terminal.

1. INTRODUCTION

Microphone arrays are increasingly replacing the head-mounted microphones as a speech acquisition system in many applications like man-machine voice communication, hands-free telephone and teleconference. This system often operates in an adverse environment where interfering sources, such as ambient noise, the reverberation effect and speakers other than the desired one, are present.

Three types of multi-microphone speech enhancement systems are commonly used: the delay and sum beamforming, also called conventional beamforming, the adaptive beamforming and noise reduction systems based on adaptive postfiltering. All these approaches assume that the direction of arrival (DOA) of the desired signal is known a priori. In real applications, the DOA of the desired signal is unknown and can only be estimated. The DOA estimate of the desired signal is obtained from the time delays between the microphone output signals. These time delays are due to the spatial distribution of microphones.

Conventional beamforming is the simplest method [1, 2]. It consists in summing the microphone output signals to improve signal reception in the presence of noise. Unfortunately, this method requires a large number of microphones, in case of a uniform array, to yield good spatial resolution at low frequencies while minimizing the gain in the direction of interference. However, this results in narrow beams in high frequencies, making the system very sensitive to the disadjustements.

The adaptive beamforming system is based on the least mean square (LMS) algorithm [3, 4]. Unfortunately, the DOA estimate is usually biased and the method is very sensitive to the estimation error. The convergence problem cannot be solved easily. In addition, the system needs a large number of filter coefficients to be applied successfully to broadband signals and reverberant environment. Furthermore, the number of degrees of freedom, which is defined as the number of interfering signals that the array can attenuate and that is given by the number of microphones minus one, is an important parameter. Consequently, the computational load becomes important for real-time applications.

The noise reduction system is based on beamforming and adaptive postfiltering operation. The postfiltering often uses a Wiener filter, applied in time or frequency domain [5, 6]. This system works satisfactorily in case of uncorrelated noise sources. However, a high spatial coherence between noise sources is measured, especially at low frequencies. This makes the noise reduction system unable to suppress noise at low frequencies. Of course, this coherence can be decreased by imposing a large microphone spacing, but this results in spatial aliasing. Furthermore, these systems suffer from the distortion of the desired signal and the introduction of a residual noise with a musical structure in the enhanced signal, which can be more annoying to a human listener than the original background noise.

Substantial progress was made concerning the realization of a microphone array system able to work correctly in adverse conditions. However, we are still far from the ultimate goal.
In this paper, we propose a new noise reduction system for an array which has a small number of microphones and is able to cover the speech bandwidth up to 4 KHz. The proposed approach is based on the wavelet transform, and the optimization of the postfilter with respect to this wavelet transform is considered.

In Sec. 2, we describe the microphone array used in this work as speech acquisition system. We also justify the use of the wavelet transform. Sec. 3 presents a derivation of the minimum mean square error (MMSE) filter design based on the wavelet transformation. Sec. 4 addresses the problem of the estimation of the Wiener filter. The description of the simulation method and the evaluation of the system is done in Sec. 5. Finally, conclusions are given in Sec. 6.

2. MICROPHONE ARRAY DESIGN

The microphone array is viewed as a discrete aperture. An inappropriate choice of the microphone spacing, d, can induce spatial aliasing in the form of grating lobes. These lobes can cause an erroneous estimate of DOA of the desired signal. In case of a uniform array, a large number of microphones, M, is required to ensure sufficient directivity while avoiding spatial aliasing in the defined speech bandwidth. Since this solution is unsuitable for practical reasons, a nonuniform (non symmetric) array with a logarithmic distribution of microphones, $M_1, M_2 \cdots M_n$ is proposed, where M_1 is the reference microphone. It should be noted that a better spatial selectivity is obtained at low frequencies compared with the uniform array for the same number of microphones, but the grating lobes are still present.

To reduce the level of the grating lobes, a better performance is obtained by splitting the logarithmic array into uniform sub-arrays [7, 2], and passing each microphone signal through a logarithmic filter bank. Then, the beamforming operation is performed in each sub-array with its appropriate frequency band [8]. The logarithmic array proposed in this paper is formed by 6 microphones, where the smallest distance between two adjacent microphones is 4.3 cm, and 4 uniform sub-arrays (see Table 1).

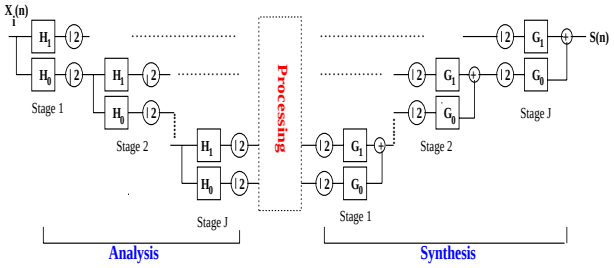

Figure 1. *Wavelet transform filter bank with octave-band decomposition.*

The multiresolution wavelet transform gives a logarithmic decomposition of the frequency axis with a good frequency resolution at low frequencies and good time resolution at high frequencies. If we observe the frequency decomposition given in Table 1, we remark that this is an octave-band structure which can be easily achieved using the wavelet transform. A perfect reconstruction of the signal can also be achieved and it is possible to reduce the computational requirements thanks to a subsampling factor as shown in Fig. 1. Thus, the wavelet transform is very appropriate to the proposed decomposition.

Subarray	Frequency band
M_1, M_5, M_6	250 Hz - 500 Hz
M_1, M_4, M_5	500Hz - 1000Hz
M_1, M_3, M_5	1000 Hz - 2000 Hz
M_1, M_2, M_3	2000 Hz - 4000 Hz

Table 1. *Space-frequency decomposition.*

Each sub-array consists of three microphones as shown in Table 1. The spatial response of each sub-array has a moderate beamwidth due to the small aperture. Limitations associated with this small aperture can be overcome by adding a postfiltering based on the Wiener filter. Furthermore, this moderate beamwidth is desirable in case of a moving desired source since it allows small fluctuations of the desired speech source location without steering the microphone array continuously.

To achieve the frequency decomposition of the microphone signals using the wavelet transform, the Daubechies's prototype filters of first order are used for the analysis and the synthesis task.

3. WAVELET TRANSFORM BASED WIENER FILTERING

The wavelet transform (WT) based Wiener filtering is a special case of the transform domain Wiener filtering, very often associated with the discrete Fourier transform (DFT). Fig. 2 shows a block diagram of the generalized Wiener filter where $\mathbf{A}$ can be any $N \times N$ orthogonal transformation matrix. The signal, $s(n)$, and the additive noise, $n(n)$, are assumed uncorrelated. To simplify the formulation, we note $s(n)$ and $n(n)$ as data vectors s and n respectively.

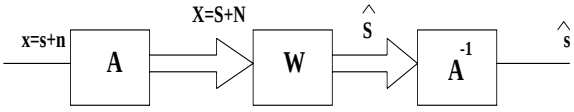

Figure 2. *Block diagram of the generalized Wiener filtering.*

In the system adopted here, an orthogonal transform operation, utilizing a wavelet transform matrix, $\mathbf{A}$, is performed on the corrupted input signal, x, yielding

$$X = A \cdot x = As + An = S + N, \tag{1}$$

where S, N and X are the WT coefficients of the desired speech, noise and corrupted signal respectively.
The resulting signal estimate from the WT Wiener filter is:

$$\hat{S} = \mathbf{W}^T X = \mathbf{W}^T (S + N). \tag{2}$$

where $\mathbf{W}^T$ is the transpose of the Wiener filter matrix. Then, the inverse WT is performed to obtain the signal

estimate

$$\hat{s} = \mathbf{A}^{-1}\hat{S} \tag{3}$$

In the design of the Wiener filter, the filter matrix, $\mathbf{W}$, is chosen so as to minimize the mean square value of the estimation error (MSE). In other words:

$$min \;\; E\{e^2\} = min \;\; E\{[s-\hat{s}]^2\} \tag{4}$$

Using Eq. 3, the expression of the MSE can be written in terms of WT domain quantities as

$$E\{e^2\} = E\{[\mathbf{A}^{-1}S - \mathbf{A}^{-1}\mathbf{W}(S+N)]^2\} \tag{5}$$

Developing Eq. 5 and taking into account the fact that s and n are uncorrelated, one obtains

$$E\{e^2\} = E\{[\mathbf{C}_s - 2\mathbf{W}\mathbf{C}_s + \mathbf{W}\mathbf{W}(\mathbf{C}_s + \mathbf{C}_n)]\}, \tag{6}$$

where $\mathbf{C}_s$ and $\mathbf{C}_n$ are the data covariance matrices of the WT of s and n respectively. The minimization of Eq. 6 yields the following optimum filter matrix:

$$\mathbf{W} = \frac{\mathbf{C}_s}{\mathbf{C}_s + \mathbf{C}_n} \tag{7}$$

In Wiener filtering, as shown in Eq. 7, the minimum MSE is independent of the type of the orthogonal transform employed [9]. For practical reasons, we reduce the computational requirements by performing a scalar Wiener filter. For this, we assume that $\mathbf{C}_s$ and $\mathbf{C}_n$ are diagonal matrices. s and n are zero-mean processes. The filter matrix $\mathbf{W}$ becomes:

$$\mathbf{W} = diag\left[\frac{E\{S^2\}}{E\{S^2\} + E\{N^2\}}\right] \tag{8}$$

The scalar filter is expressed as follows:

$$W = \frac{E\{S^2\}}{E\{X^2\}} \tag{9}$$

By analogy, the quantities $E\{S^2\}$ and $E\{X^2\}$ are considered as the wavelet power spectra of the desired and the noisy speech signals, respectively. We note that the wavelet power spectra summarize the information in the frequency spectrum by using just one value per octave frequency band.

Of course, this kind of assumption corresponds to a suboptimal design of the Wiener filter and may lead to some filtering error. This is the price to pay for obtaining a reasonable compromise between the optimality of the filter and the computational load [9, 10].

4. ADAPTIVE POSTFILTERING IN THE WAVELET PACKET DOMAIN

The postfiltering operation is performed using the orthogonal wavelet transform. The wavelet power spectra of X and S are estimated using the microphone output signals in the same manner as presented in [5]. We note that the

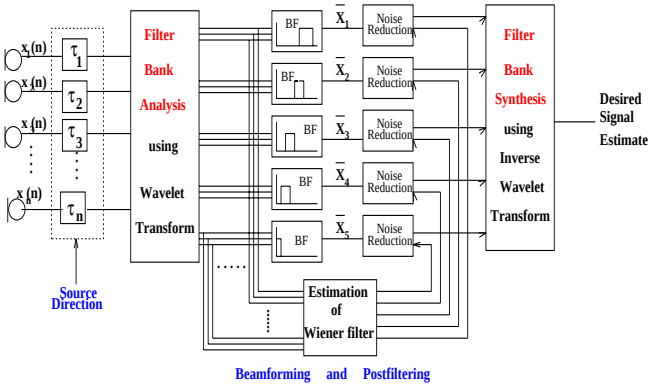

Figure 3. *Block diagram of the proposed system.*

output of the conventional beamformer, performed according to the sub-array decomposition, $\bar{X}$ is used as the input of the Wiener filter. Thus, the output of the noise reduction system is:

$$\hat{S}_\phi(n) = \bar{X}_\phi(n) W_\phi(n) \qquad \phi = 1, \cdots, L \tag{10}$$

where L is the number of subbands and ϕ is the index of the subband. Fig. 3 shows the block diagram of the proposed system. Obviously, the nonstationarity of the signal is not neglected and the Wiener filter is performed, frame-by-frame and separately, in each subband.

By analogy to the transfer function of the Wiener filter proposed in Zelinski's and Simmer's methods [5, 6], the Wiener filter in the wavelet domain has the following expression:

$$W_\phi(n) = \frac{\frac{2}{M \cdot M - 1}\sum_{i=1}^{M-1}\sum_{j=i+1}^{M} X_{i,\phi}(n) \cdot X_{j,\phi}(n)}{\frac{1}{M}\sum_{i=1}^{M}|X_{i,\phi}(n)|^2} \tag{11}$$

where $X_{i,\phi}(n)$ are the wavelet components of the i^{th} microphone signal in the spectral band of order ϕ.

5. EXPERIMENTS AND RESULTS

Since the voice communication systems are destined to the human listener, we use the Log Area Ratio (LAR) for the performance evaluation. This objective measure, based on LPC analysis, presents a good correlation with the human auditory system.
We note that the problem of the estimation of DOA of the desired signal is not addressed in this work. The DOA of the desired signal is assumed to be known. Let us also point out that for the octave-band decomposition, Daubechies compactly supported wavelets with variable lengths filter lengths are used. No remarkable improvement was obtained when using longer filters. Thus, we have limited the Daubechies filters to the lower-order filters (2 coefficients). We note that these filters are equivalent to Haar's wavelet. This provides a minimum delay in the filter banks making the system attractive for real-time processing.

The proposed new algorithm was evaluated by performing simulations with SNR decreasing down to 0dB and

with several types of noise sources. The simulations were realized assuming to have a single speech source and a single competitive noise source. With this approach, a segmental SNR improvement up to 16dB is achieved and Fig. 5 shows the achieved improvement in terms of LAR.

As far as the implementation is concerned, the method exploits an efficient and rapid algorithm for multiresolution filter bank [11]. The usefulness of the proposed approach is proved by the high attenuation of noise level as shown in Fig. 4).

Work is currently undertaken to obtain better approximations of the Wiener filter applied in the wavelet domain than the one used in this paper. In other words, the optimality of the filter is still matter of search. The present system is being tested in real situations and future work will be devoted to confirm the advantages of the proposed method for hands-free acquisition systems operating in a noisy environment.

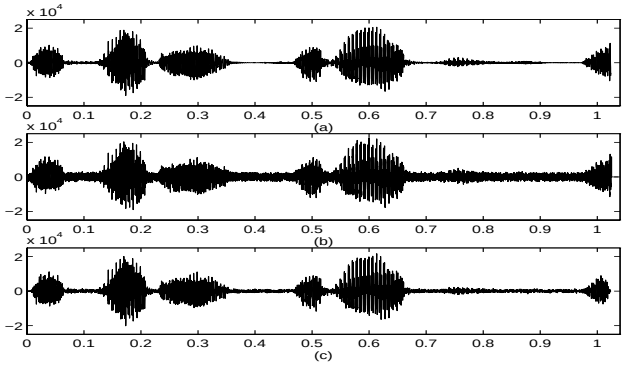

Figure 4. *The noise reduction system's results: (a) original signal, (b) noisy signal, (c) reconstructed output.*

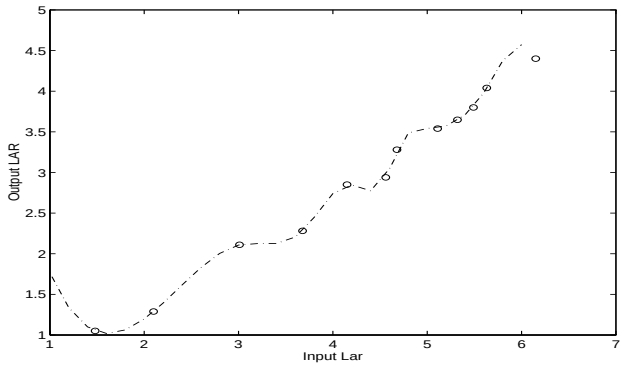

Figure 5. *Performance measure in terms of LAR of the proposed algorithm.*

6. CONCLUSION

We have presented a new noise reduction system based on subband decomposition using the wavelet transform. It compares favorably with the methods proposed before (e.g. [5, 6]). No noticeable distorsion was perceived by human listeners. Particularly, the resulting speech signal is free from any musical noise. Finally, it should be noted that one important aspect which has been addressed in this paper is the multiresolution property. Consequently, considerable computational savings can be obtained by applying the multirate technique.

7. ACKNOWLEDGEMENT

The author would like to thank Dr. A. Drygajlo for useful discussions on the subject of this paper.

REFERENCES

[1] J. L. Flanagan, D. A. Berkley, G. W. Elko, J. E. West, and M. M. Sondhi, "Autodirective Microphone Systems", *Acustica*, vol. 73, pp. 58–71, 1991.

[2] W. Kellermann, "Self-Steering Digital Microphone Array", *in Proc. of ICASSP'91*, vol. 4, pp. 3581–3584, July 1991.

[3] O. L. Frost, "An Algorithm for Linearly Constrained Adaptive Array Processing", *Proceedings of IEEE*, vol. 60, pp. 916–935, 1972.

[4] K. M. Buckley and L. J. Griffiths, "Broad-Band Signal-Subspace Spatial-Spectrum(BASS-ALE) Estimation", *IEEE Trans. on ASSP*, vol. 36, pp. 953–964, July 1988.

[5] R. Zelinski, "A Micropohone Array with Adaptive Postfiltering for Noise Reduction in Reverberant Rooms", *in Proc. of ICASSP'88*, pp. 2578–2581, 1988.

[6] K. U. Simmer and A. Wasiljeff, "Adaptive Microphone Arrays for Noise Suppression in the Frequency Domain", *in 2nd Cost 229 Workshop on Adap. Algo. in Com., France*, pp. 185–194, Sep. 1992.

[7] R. Zhan J. L. Flanagan, Don H. Johnson and G. W. Elko, "Computer-Steered Microphone Array for Sound Tranduction in Large Rooms", *J. Acoust. Soc. Am.*, vol. 78, pp. 1508–1518, 1985.

[8] D. Mahmoudi, "Multiresolution Array Processing for Speech Source Tracking in Voice Communication Systems", *in Proc. of ICSPAT*, pp. 346–350, Oct. 1996.

[9] W. K. Pratt, "Generalized Wiener Filtering Computation Techniques", *IEEE Trans. on Computers*, vol. c-21, pp. 636–641, July 1972.

[10] K. Rumatowski, "Walsh Transform Applied to Digital Filtering", *Signal Processing*, pp. 253–263, 1986.

[11] A. Drygajlo, "Butterfly Orthogonal Structure for Fast Transforms, Filter Banks and Wavelets", *in Proc. of ICASSP'92*, vol. 5, pp. 81–84, March 1992.