A Signal Subspace Tracking Algorithm for Speech Acquisition and Noise Reduction with a Microphone Array

Sofiène AFFES and *Yves GRENIER*

Ecole Nationale Supérieure des Télécommunications de Paris

Département Signal - 46 rue Barrault, 75634 Paris, Cedex 13, France

E-mail: affes@sig.enst.fr - Tel: ++331 45.81.77.82 - Fax: ++331 45.88.79.35

Abstract: The underlying idea of array processing is to synchronize and average the signals coherently propagating along the direct path covered from the desired source to the sensors. Simultaneously, any other source arriving from a different path (*i.e.* jammers or noise) is incoherently received and thus suppressed. Unfortunately, sources do not propagate along known and direct paths in real applications such as speech acquisition in an acoustic environment. Particularly in hands-free telephony, we have to deal with undesired propagation phenomena such as reflections or reverberations commonly encountered in other multipath environments. We also have to deal with the presence of a strong and close echo emitted from the loudspeaker. Prior methods developed adaptive microphone arrays for noise reduction after a time delay compensation of the direct path. This simple synchronization is insufficient to produce an acceptable speech quality, and makes adaptive beamforming unsuitable. In this contribution, we first propose to partially estimate impulse responses in the frequency domain, using a signal subspace tracking procedure we adapted from previous works. A significant acoustic feature revealed by the characterization of the environment finally makes a quite full identification of these responses possible. Simulations made with synthetic and real data show both objective and subjective results in speech acquisition and noise reduction, including echo cancellation, to be very satisfying.

I Introduction

In array processing, sources are usually modeled or approximated to propagate as planar or spherical waves. Steering can be seen in that case as a simple synchronization or compensation of the direct path. In real applications such as speech acquisition in acoustic environments, undesired multipath propagation phenomena such as reflections or reverberations can no longer be neglected by the processing stage. Particularly in adaptive beamforming, early reflections can be considered as coherent jammers, and may cause the cancellation of the desired source. Hence, steering should be rather seen as a first deconvolution step. In that case, the array processor outputs should correspond to a simple time delay response when summed.

Many algorithms were proposed for speech enhancement with adaptive microphone arrays [1-5]. Unfortunately, most of them turn down the first stage of steering (*i.e.* synchronization) and put the emphasis on noise reduction alone. In [6], we tested these methods for speech acquisition in cars and noticed their poor performance in noise reduction. Kaneda and Ohga measure the impulse responses (IRs) from a selected speaker position to the microphone array, then use it to train the array processor (*i.e.* beamformer) with recorded noise [1]. This requires real conditions difficult to reach with mobile speakers and nonstationary signals. The others assume a perfect synchronization of the signals performed by a time delay compensation of the direct path [2-5]. Even this simple task is hard to achieve properly in the presence of usually long reverberations and strong reflections, which besides remain uncompensated in that case. In addition, it prevents the use of adaptive beamformers such as the GSC (generalized sidelobe canceller) [11] often sensitive to steering errors and coherent interference [12], but efficient for jammers suppression. Consequently, this scheme imposes a delay-sum (DS) structure. With a restricted number of microphones, performances are hence limited to a poor reduction of only spatial diffuse noise.

We also proved in [6] the advantage of steering over time delay compensation achievable by beamforming in terms of clarity index gain of the total impulse response (IR). This leaves a significant potential for producing a very natural quality of speech and a higher intelligibility at the output.

We finally tested an improved version of [5] described in [7] for speech acquisition and noise reduction in a banker market operating room [8]. This method still proposes the use of a computationally expensive time delay compensation unit based on cross-correlation [9]. Despite the relative improvement in performance, we definitely mentionned in our contribution to [8] the necessity of performing steering instead of time delay compensation to really achieve satisfactory results.

We alternatively proposed in [10] an algorithm for wideband robust adaptive beamforming based on signal subspace tracking with simultaneous projection on an array manifold (*i.e.* IRs identification). We actually studied the algorithm with a simple manifold of far-field sources as a particular case of a more general array characterization. This flexible formulation makes room for a possible adaptation to acoustic environments. In addition, the high performance of the algorithm and its low complexity observed in that simple case offer a significant perspective for further implementation in real applications.

In this paper, we propose a signal subspace tracking algorithm similar to [10] adapted to speech acquisition in a banker market operating room. In section II, we first make an acoustic characterization of the array to possibly find the underlying features of the IRs. We will actually notice their total energy at any frequency component to be quite constant for emitting locations around a central speaker position. From this key observation, we introduce some significant constraints characterizing the array, then adapt the tracking procedure to the studied environment in section III. We also introduce a voice activity detector for the tracking inspired from [9], and a parallel GSC structure [11] for speech acquisition and noise reduction. Simulation results described in section IV show a very good quality of speech and an efficient noise reduction in both objective and subjective evaluations. The proposed algorithm outperforms a similar GSC structure combined with a simple time delay compensation as suggested in [13,14]. In addition, it is even able to cancell a strong echo emitted from a close loudspeaker without any knowledge of its reference signal. Conclusion is finally given in section V.

II Acoustic Characterization and Model

We consider for our application an array of 12 microphones located around the screen of a computer workstation (6 linearly placed along the top edge, and 2×3 placed on both the left and right edges). The spacing between each couple of adjacent sensors is 0.07 m. This array is the front-end receiver of a hands-free telephone installed on a banker market operator desk. The loudspeaker is fixed to the keyboard. We can now model the signals received from the microphone array at time t as follows:

$$\mathcal{X}(t) = \mathcal{G}(t) \otimes s(t) + \mathcal{N}(t) \ , \tag{1}$$

where $\mathcal{X}(t)$ denotes the m-dimensional observation vector when $m = 12$ is the number of microphones, and where $s(t)$ is the emitted speech signal uttered from the operator. $\mathcal{G}(t) \triangleq [g_1(t), \cdots, g_m(t)]^T$ is the m-dimensional vector of IRs, $\mathcal{N}(t)$ is the noise vector, and $\otimes$ denotes time convolution. All the quantities considered in (1) are real.

We have already measured the impulse responses at a sampling frequency of 8 kHz for 4 selected positions of the speaker's mouth (center, right, left, bottom). In Figure 1-a, the plot of the 6^{th} IR of the central position over 1024 coefficients clearly shows early reflections and reverberations. A time delay compensation unit is likely to synchronize the IRs over the direct path, but reflections and reverberations are not negligible and would be still present. Besides their undesirable effect on speech quality at the output, they can cause the cancellation of the desired signal if an adaptive noise reduction algorithm is used. For instance, Van Compernolle used a cross-correlation based unit similar to [9] with a GSC structure [13]. He also replaced this unit by adaptive filters in [14] to improve the accuracy of time delay estimates. Nevertheless, he reported with both schemes predictable signal cancellation phenomena at a positive SNR (signal to noise ratio) [12]. To overcome this problem, he proposed to block the GSC adaptation during speech activity. This suboptimal solution restricts signal cancellation to some extent, but speech quality and noise reduction are still affected. Actually, we will confirm later by simulation the drawbacks of approximating the IRs by simple time delays as usually made in [2-5,7,9,13,14].

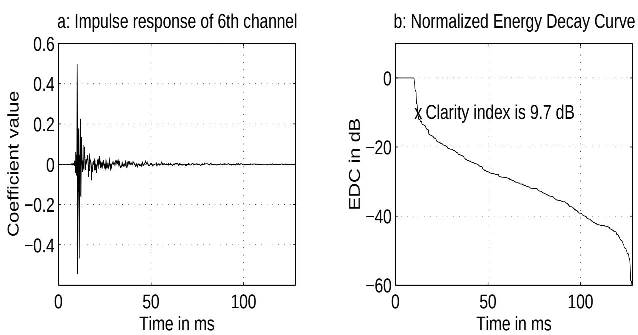

Figure 1: 6th IR of speaker position at center.

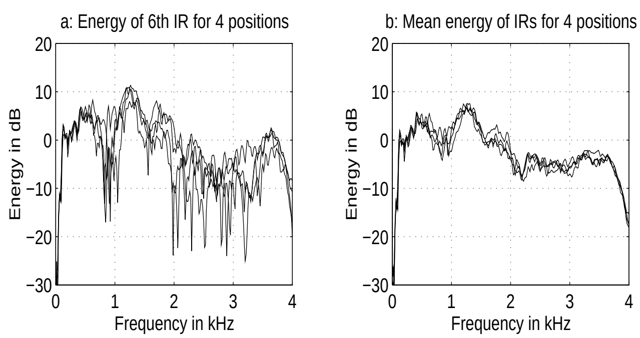

Figure 2: Energy characterization of IRs.

Let us now define the energy decay curve (EDC) of the i^{th} IR $g_i(t)$ $(i = 1, \cdots, m)$ as follows:

$$E_{g_i}(t) \triangleq \sum_{\tau=t}^{\infty} g_i^2(\tau) \ . \tag{2}$$

The clarity index is then defined by $C(g_i) \triangleq 10 \log_{10}(\frac{E_{g_i}(0)}{E_{g_i}(T_d)})$, where T_d is the total duration of the direct path and the early reflections. This index which specifies the quality of an acoustic channel for speech transmission, is the ratio of the total energy of the associated impulse response to the energy contained in its late reverberation part. The normalized curve of E_{g_6} plotted in Figure 1-b shows a relatively low clarity index of 9.7 dB at $T_d \approx 128$. It also shows that more than 99% of the energy is located in the first 256 samples, whereas the first 512 coefficients contain only less than 0.9% more. In this case, the identification of the IRs can be reasonably made over $L = 256$ coefficients.

We hence take the FFT (Fast Fourier Transform) of equation (1) over $2L$ snapshots each $K \leq L$ sampling periods. For $f = 0, \cdots, 2L-1$, we have:

$$X_{f,n} = G_f s_{f,n} + N_{f,n} \ , \tag{3}$$

where the subscripts f and n denote respectively in (3) the FFT of the indexed quantity at the frequency bin f and the $(m \times 2L)$-block number n. We actually assume the time variations of $\mathcal{G}(t)$ to be very slow and practically constant in comparison to the variations of $s(t)$ and $\mathcal{N}(t)$. We hence assume $G_{f,n} \approx G_f$ for simplicity.

In Figure 2-a, we plot $|g_{6,f}|^2$ for the 4 selected positions of the speaker. The curves show relatively high variations of the IRs from one position to another. On the other hand, the curves of $\frac{||G_f||^2}{m} = \frac{\Sigma_i |g_{i,f}|^2}{m}$ plotted for the same 4 positions in Figure 2-b show small variations. Their standard deviation is actually smaller than 10% of the mean value at any frequency component. In this case, we can assume the mean energy $\frac{||G_f||^2}{m}$ to be equal to a constant say β_f^2 for any location of the speaker around the central position. This constant can be measured as a weighted combination of the curves plotted in Figure 2-b. Intuitively, some kind of "local energy conservation principle" gives support to this feature which underlines the array characterization of our IRs.

In this case, we can rewrite equation (3) as follows:

$$X_{f,n} = \alpha_{f,n} \ U_f + N_{f,n} \ , \tag{4}$$

where the complex vector $U_f \triangleq \sqrt{\frac{m}{||G_f||^2}} \ G_f$ is the signal subspace basis vector with norm $\sqrt{m}$, and the complex scalar $\alpha_{f,n} \triangleq \sqrt{\frac{||G_f||^2}{m}} \ s_{f,n}$ is the modulated signal parameter. We can finally write $\alpha_{f,n} \approx \beta_f \ s_{f,n}$. If it is possible to track the signal subspace properly, the idea is to recover the modulated signal parameter $\alpha_{f,n}$ and consequently estimate $s_{f,n}$.

III The Proposed Algorithm

Let us first assume that an estimation of the signal subspace basis U_f at iteration n say $\hat{U}_{f,n}$ is available. We can immediately estimate $\alpha_{f,n}$ using a simple DS beamformer where the usual steering vector is to be replaced by $\hat{U}_{f,n}$. However, a proper identification of the signal subspace is associated with a reliable identification of the jointly orthogonal noise subspace. It is then possible to further reduce the residual noise still present in the DS output from available and quite clean noise references. To better estimate the signal parameter $\alpha_{f,n}$ $(f = 0, \cdots, L)$, we use a GSC structure [11] as follows:

$$
\begin{aligned}
y_{f,n} &\triangleq \frac{\hat{U}_{f,n}^H \; X_{f,n}}{m} \;, \\
X_{f,n}^N &\triangleq P_f^N \; X_{f,n} \;, \\
\hat{\alpha}_{f,n} &= y_{f,n} - W_{f,n}^H \; X_{f,n}^N \;, \\
W_{f,n+1} &= W_{f,n} + \eta_{f,n} \; X_{f,n}^N \; \hat{\alpha}_{f,n}^H \;,
\end{aligned}
\tag{5}
$$

where P_f^N is a $(m-1) \times m$ blocking matrix projecting $X_{f,n}$ on the noise subspace to obtain $X_{f,n}^N$. A^H denotes the conjugate transpose of A, and $\eta_{f,n}$ is the step-size of the GSC possibly including a normalization factor (*i.e.* $\eta_{f,n} \triangleq \frac{\eta_0}{\|X_{f,n}^N\|^2}$). $W_{f,n}$ is a $(m-1)$-dimensional vector initially set to 0. Actually, we start the algorithm with $\hat{U}_{f,0} = [e^{-j2\pi\hat{\tau}_1 \frac{f}{2L}}, \cdots, e^{-j2\pi\hat{\tau}_m \frac{f}{2L}}]^T$ where $\hat{\tau}_i$ $(i = 1, \cdots, m)$ are time delay estimates of the direct path.

Second, we use the observation vectors $X_{f,n}$ and the modulated signal parameters $\hat{\alpha}_{f,n}$ in the same way as in [10] to correct $\hat{U}_{f,n}$ and track the basis vectors U_f. In fact, we observed that the tracking procedure is less stable with the GSC output $\hat{\alpha}_{f,n}$ when applied to nonstationary signals. Instead, we used the DS-like (Delay-Sum) output $y_{f,n}$ in (5) as follows:

$$
\tilde{U}_{f,n+1} = \hat{U}_{f,n} + \mu_{f,n}(X_{f,n} - \hat{U}_{f,n} y_{f,n}) \; y_{f,n}^H \;, \tag{6}
$$

where $\mu_{f,n}$ is the step-size of the LMS-like tracking equation (6) possibly including a normalization factor, and where $\tilde{U}_{f,n+1}$ denotes an unconstrained estimate of $\hat{U}_{f,n+1}$. In this case, we obtain the simplified neuron model as a principal component analyzer proposed by Oja [15]. In a generalization of (6) to the case of a multi-dimensional signal subspace made in [16], this unconstrained tracking procedure is particularly proved to converge to the one dimensional signal subspace with the highest energy and the basis vector $\tilde{U}_{f,\infty}$ such that $\|\tilde{U}_{f,\infty}\|^2 = m$.

At a reasonable SNR, this equation converges to any solution of the form $\tilde{U}_{f,\infty} = e^{j\phi_f} U_f$ where ϕ_f is a phase shift. Our experience is that $\phi_f \approx 2\pi\tau_\infty \frac{f}{2L}$ is close to a linear phase where τ_∞ is a short time delay. In addition, the human auditory system is not very sensitive to phase distortion [2]. Hence, the effect of ϕ_f on speech quality is not significant.

When the SNR is very low particularly during the periods of silence, equation (6) is likely to track noisy sources. It would be better then to stop the adaptation of the algorithm so as to keep the estimates of U_f from being attracted in the noise subspace. To do so, we first define the steered input signals by $\hat{Y}_{f,n} \triangleq \text{diag}[\hat{U}^H_{f,n}]X_{f,n}$ where diag$[V]$ is a diagonal matrix with the elements of vector V on the main diagonal. We then introduce a modified version of the voice activity detector presented in [9] as follows:

$$a(n+1) = (1-\gamma)a(n) + \gamma \times \frac{\sum_{f\in\Phi}\{\frac{2}{m(m-1)} \times \sum_{j=1}^{m-1}\sum_{k=j+1}^{m} \text{Re}(\hat{Y}_{j,f,n}\hat{Y}^H_{k,f,n})\}}{\sum_{f\in\Phi}\left|\frac{1}{m} \times \sum_{l=1}^{m}\hat{Y}_{l,f,n}\right|^2}, \tag{7}$$

where speech activity $a(n)$ is given by a smoothed ratio of the sum of the cross-spectrum components at a selected set of frequencies Φ, over the sum of the auto-spectrum components at the same frequencies. γ is a smoothing factor, Re(.) denotes the real part of a complex, and $\hat{Y}_{j,f,n}$ is the j^{th} component of $\hat{Y}_{f,n}$. We found it also better to select 10 frequencies around 1.5 kHz and 2.8 kHz rather than defining $\Phi = \{0, 1, \cdots, \frac{L}{2}\}$ (*i.e.* the low frequency region going to 2 kHz) as proposed in [9]. To test the presence of speech or silence, speech activity $a(n)$ is compared to a given threshold a_{min} as follows:

$$\delta_n \triangleq \begin{cases} 1 \text{ if } a(n) \geq a_{min}, \\ 0 \text{ otherwise (silence)}. \end{cases} \tag{8}$$

We finally replace the step-size of the tracking equation in (6) by $\bar{\mu}_{f,n} \triangleq \delta_n\ \mu_{f,n}$. Notice that δ_n simultaneously rules the adaptation of (6) at any frequency f, though it can be splitted into frequency regions with sets other than Φ.

We now form the matrix $\tilde{G}_{n+1} \triangleq [\beta_0\tilde{U}_{0,n+1}, \cdots, \beta_L\tilde{U}_{L,n+1}, \cdots, \beta_{2L-1}\tilde{U}_{2L-1,n+1}]$ whose lines approximate the FFT of the IRs. We hence have to constraint the corresponding fast convolution to be linear [17]. This is likely to limitate any deviation of the tracking procedure from the true IRs, even at reasonably low SNRs. We hence compute the matrix $\tilde{\mathcal{G}}_{n+1}$ as the line by line IFFT of $\tilde{G}_{n+1}$, then set its $m \times L$ right half part to 0 to have $\hat{\mathcal{G}}_{n+1}$. We again take the line by line FFT of $\hat{\mathcal{G}}_{n+1}$ to estimate $\hat{G}_{n+1}$ or equivalently have $\hat{U}_{f,n+1}$ for $f = 0, 1, \cdots, L$. More details can be found in [17-19] about constrained adaptive filtering in the frequency domain and fast linear convolution.

We finally estimate the speech signal at the block $n+1$ by:

$$[\hat{s}(K(n+1)), \cdots, \hat{s}(K(n+1)+2L-1)] \triangleq \text{Re}\left\{\text{IFFT}\left(\left[\frac{\hat{\alpha}_{0,n}}{\beta_0}, \cdots, \frac{\hat{\alpha}_{2L-1,n}}{\beta_{2L-1}}\right]\right)\right\}. \tag{9}$$

With blocks shifted each $K < L$ samples, input data is oversampled at a rate higher than required to update (6) more frequently. This is shown to improve the tracking performance of the algorithm [18,19]. As blocks overalp over $2L - K$ samples, we only keep the following segment of length K:

$$[\hat{s}(K(n+1)+L), \cdots, \hat{s}(K(n+1)+L+K-1)].$$

An averaging over K output blocks can be alternatively made for subsampling [17-19].

IV Simulation Results

In this section, we propose to assess the performance of the studied algorithm for speech acquisition and noise reduction. We also want to compare it to prior methods based on simple time delay compensation. For this reason, we start the proposed scheme with time delay identified IRs as stated in section III. For objective evaluation, we first select an original signal of 2 speech sentences uttered from a female speaker and recorded in an anechoic room. We also convolve the original waveform plotted in figure 3-a with the recorded IRs from the nominal central position of the speaker to the array of microphones. The convolved signals are finally corrupted at a mean SNR of 7 dB by a background noise containing cocktail party speech, the noise of keyboards and the workstation fans. In figure 3-b, we plot one of the synthesized signals simulating the noisy speech received at the 6^{th} microphone.

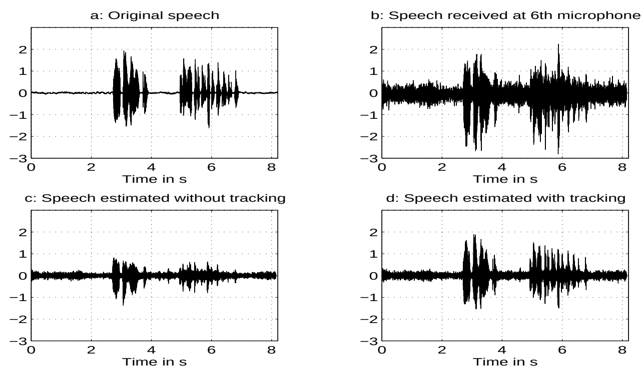

Figure 3: Speech signal at different stages.

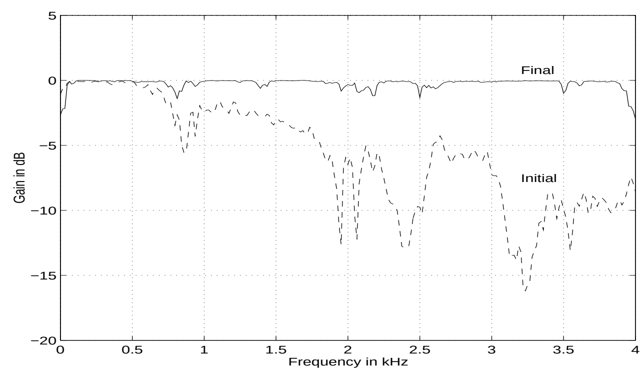

Figure 4: Gain of total IR in dB.

For comparison, we skip the tracking step illustrated by equation (6). This corresponds to a simple time delay compensation usually admitted by [2-5,7,9,13,14]. In this case, we clearly observe in Figure 3-c the cancellation of the speech signal as reported in [13,14]. On the other hand, the proposed algorithm avoids this phenomena as shown in figure 3-d, and proves the efficiency of the subspace tracking procedure of equation (6). In figure 4, we plot the gain in energy of the total response to the central position of the speaker. The initial curve corresponds to time delay compensation, and shows the approximation usually made in [2-5,7,9,13,14] to be inadequate beyond a small region in the low frequency. The final curve corresponds to the identified IRs after convergence of (6), and shows the leak from the signal to the noise subspace to be reduced to a quite negligible range. Despite the small distortions in phase and modulus, the quality of output speech is very natural, and point jammers are significantly reduced. However, the gain in SNR of approximately 7 dB is less than the optimal 10.8 dB reduction of spatial diffuse noise (*i.e.*

$10\ \log_{10}(m) \approx 10.8$). To improve the performance in SNR gain, we propose a post-processing stage of the residual noise as made in [2-5,7], using however a spectral subtraction method developed by Ephraim and Malah [20]. The additional gain is of 5 dB at an output SNR of 19 dB.

For subjective evaluation, we now test the algorithm with data recorded in real conditions. Since objective results are very satisfying, we only keep the 6 microphones located at the screen top edge of the workstation in the microphone array for the rest of our experiments. Four tests are run with sentences uttered from both male and female speakers at average SNRs ranging from 0 to 8 dB. The quality of both the output speech and residual noise still sounds good and natural. It shows a significant improvement when compared to the results of [8]. The total gain in SNR is ranging from 9 to 12 dB after post-processing. Other tests proved the algorithm to be even able to cancell a strong echo emitted from a close loudspeaker without any knowledge of its reference signal, and without any damage of the output speech.

V Conclusion

We proposed in this paper an algorithm based on a signal subspace tracking procedure we adapted from previous works, to partially identify IRs in the frequency domain. In an acoustic characterization of the environment, we noticed the total energy of IRs from any location of the speaker close to a nominal central position to be quite constant at any frequency component. From this key observation, we proved a full identification of IRs to be possible and more advantageous than simple time delay compensation in terms of speech acquisition and noise reduction. Consequently, adaptive structures for noise redcution such as the GSC can be efficiently used without considerable risks of speech cancellation.

At present, we are trying to explicitly incorporate echo cancellation in the proposed scheme using techniques developed in [18,19]. In parallel, we are adapting the algorithm to an application of hands-free telephony in a mini-teleconference mode where not only one but many speakers are free to move around in an office. Although some issues are still under investigation, the first experimental results we obtained are very encouraging.

Acknowledgment: The authors would like to acknowledge all the partners of ENST in the European ESPRIT project 6166 FREETEL, who authorized us to use their database. This data was recorded by ENST and PAGE Iberica in a banker market operating room of Banesto, Madrid, Spain.

References

[1] Y. Kaneda and J. Ohga, "Adaptive microphone-array system for noise reduction", *IEEE Trans. on ASSP*, vol. 34, no. 6, pp. 1391-1400, December 1986.

[2] M.M. Sondhi and G.W. Elko, "Adaptive optimization of microphone arrays under a nonlinear constraint", *Proc. of ICASSP'86*, pp. 981-984, Tokyo, Japan, April 7-11, 1986.

[3] R. Zelinski, "A microphone array with adaptive post-filtering for noise reduction in reverberant rooms", *Proc. of ICASSP'88*, pp. 2578-2581, New York City, USA, April 11-14, 1988.

[4] S. Gierl, "Noise reduction for speech input systems using an adaptive microphone-array", 22^{nd} *ISATA*, Florence, Italy, pp. 517-524, May 1990.

[5] K.U. Simmer and A. Wasiljeff, "Adaptive microphone arrays for noise suppression in the frequency domain", 2^{nd} *Cost 229 Workshop on Adaptive Algorithms in Communications*, Bordeaux-Technopolis, France, September 30 - October 2, 1992.

[6] S. Affes and Y. Grenier, "Test of adaptive beamformers for speech acquisition in cars", *Proc. of the* 5^{th} *International Conference on Signal Processing Applications and Technology ICSPAT'94*, Dallas, USA, vol. I, pp. 154-159, October 18-21, 1994.

[7] Z. Yang, K.U. Simmer, and A. Wasiljeff, "Improved performance of multi-microphone speech enhancement systems", *Proc. of* 14^{th} *GRETSI Symposium*, Juan-Les-Pins, France, vol. 1, pp. 479-482, September 13-16, 1993.

[8] R. Le Bouquin-Jeannès and G. Faucon, Ed., *Advanced solutions for noise reduction*, Deliverable 4.123.2, ESPRIT Project 6166 FREETEL, University of Rennes, Rennes, France, July 1994.

[9] K.U. Simmer, P. Kuczynski, and A. Wasiljeff, "Time delay compensation for adaptive multichannel speech enhancement systems", *ISSSE'92*, pp. 660-663, Paris, France, September 1-4, 1992.

[10] S. Affes, S. Gazor, and Y. Grenier, "Wideband robust adaptive beamforming via target tracking", *Proc. of the* 7^{th} *IEEE SP Workshop on SSAP*, Quebec City, Canada, pp. 141-145, June 26-29, 1994.

[11] L.J. Griffiths and C.W. Jim, "An alternative approach to linearly constrained adaptive beamforming", *IEEE Trans. on Antennas and Propagation*, vol. 30, no. 1, pp. 27-34, 1982.

[12] B. Widrow, K.M. Duvall, P.R. Gooch, and W.C. Newmann, "Signal cancellation phenomena in adaptive antennas: causes and cures", *IEEE Trans. on ASSP*, vol. 30, no. 3, pp. 469-478, May 1982.

[13] D. Van Compernolle, "Adaptive filter structures for enhancing cocktail party speech from multiple microphone recordings", *Proc. of* 12^{th} *GRETSI Symposium*, Juan-Les-Pins, France, vol. 1, pp. 513-516, June 12-16, 1989.

[14] D. Van Compernolle, "Switching adaptive filters for enhancing noisy and reverberant speech from microphone array recordings", *Proc. of ICASSP'90*, Albuquerque, USA, April 2-8, 1990.

[15] E. Oja, "A simplified neuron model as a principal component analyzer", *Journal of Math. Biology*, Springer-Verlag, vol. 15, pp. 267-273, 1982.

[16] B. Yang, "Subspace tracking based on the projection approach and the recursive least squares method", *Proc. of ICASSP'93*, Minneapolis, USA, vol. IV, pp. 145-148, April 27-30, 1993.

[17] J.J. Shynk, "Frequency-domain and multirate adaptive Filtering", *IEEE SP Magazine*, pp. 14-37, January 1992.

[18] E. Moulines O. Ait Amrane, and Y. Grenier, "The generalized multi delay adaptive filter: structure and convergence analysis", *IEEE Trans. on Signal Processing*, January 1995.

[19] J. Prado and E. Moulines, "Frequency-domain adaptive filtering with applications to acoustic echo cancellation", *Annales des Télécommunications*, vol. 49, no. 7-8, pp. 414-428, July 1994.

[20] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean square error log-spectral amplitude estimator", *IEEE Trans. on ASSP*, vol. 33, no. 2, pp. 443-445, April 1985.