ACOUSTIC FILTER-AND-SUM BEAMFRMING BY ADAPTIVE PRINCIPAL COMPONENT ANALYSIS

Ernst Warsitz, Reinhold Haeb-Umbach

University of Paderborn
Dept. of Communications Engineering
33098 Paderborn, Germany
{warsitz,haeb}@nt.uni-paderborn.de

ABSTRACT

For human-machine interfaces in distant-talking environments multichannel signal processing is often employed to obtain an enhanced signal for subsequent processing. In this paper we propose a novel adaptation algorithm for a Filter-and-Sum beamformer to adjust the coefficients of FIR filters to changing acoustic room impulses e.g. due to speaker movement. A deterministic and a stochastic gradient ascent algorithm are derived from a constrained optimization problem, which iteratively estimate the eigenvector corresponding to the largest eigenvalue of the cross power spectral density of the microphone signals. The method does not require an explicit estimation of the speaker location. The experimental results show fast adaptation and excellent robustness of the proposed algorithm.

1. INTRODUCTION

Microphone array speech signal processing is an essential element for hands-free speech communication or recognition. For removing unwanted interference and noise from desired signals, multichannel techniques exploit the spatial diversity to discriminate between desired and undesired signal components. In this paper we are concerned with adaptive beamforming to direct a beam of increased sensitivity towards a possibly moving desired source, whose position is not known *a priori*.

There are essentially three classes of source localization algorithms [1]: a) Estimating the time differences of arrival, e.g. by the Generalized Cross Correlation method, b) Using high-resolution spectral estimation, and c) Using a beamformer which is steered in various directions and searching for peaks in the output signal. All these methods perform poorly in highly reverberant environments and many encompass considerable computational effort [2].

The primary objective is, however, often not the determination of the source location but the computation of an enhanced speech signal for subsequent processing (e.g. recognition or transmission). In the approach described below we avoided to explicitly localize the source. Instead we developed an adaptation algorithm for a Filter-and-Sum beamformer (FSB) to adjust the coefficients of the FIR filters to changing acoustic room impulse responses, e.g. due to speaker movement. If necessary, information about the source location can be derived from the filter coefficients [3].

In presteered arrays the microphone signals are appropriately delayed to steer the beam towards the desired source location. For broadband source signals, such as speech signals, subsequent FIR filters are used to compensate for the frequency-selectivity of the room impulse response. Frost developed a constrained LMS algorithm to adapt the FIR filter coefficients [4]. His concept has found many applications, with the generalized sidelobe canceller by Griffiths and Jim as one of the most important extensions [5].

Frost did not consider the estimation of the steering angle. In this paper we are concerned with the steering of the beamformer towards the desired location. Rather than using pure time delays (c.f. *Delay-and-Sum Beamforming*) we propose to use (typically short) FIR filters instead, which are capable of not only compensating for the time delays of the direct (*line-of-sight*) signal component between the individual sensors but which can also align early reflections. It is shown that this improves the signal-to-interference ratio of the output signal. The FIR filter coefficients are determined by solving a constrained optimization problem which turns out to be an eigenvalue problem: For each frequency bin, the eigenvector associated with the largest eigenvalue of the cross-power spectral density matrix of the input sensor signals is the optimum filter coefficient vector. To compute this eigenvector, we derive an iterative method, which is suitable for tracking this eigenvector in time-variant (e.g. due to speaker movement) scenarios. The well-known Oja-rule for adaptive principal component analysis [6] can be obtained as a special case of the proposed method.

The paper is organized as follows. In Sections 2, 3 and 4 we formulate the optimization problem, derive a deterministic gradient ascent solution, and present a stochastic gradient realisation, respectively. Section 5 presents experimental results which demonstrate that the proposed method is both fast and robust and thus well suited for complex time-varying acoustic scenarios.

2. FORMULATION OF CONSTRAINED OPTIMIZATION PROBLEM

We are given an array of M microphones, where each microphone signal $x_i(n)$ consists of two components: The desired signal component $s_i(n)$, which results from the convolution of the source signal $u(n)$ with the room impulse response $h_i(n)$, and the noise term $n_i(n)$:

$$\begin{aligned} x_i(n) &= s_i(n) + n_i(n) \\ &= h_i(n) * u(n) + n_i(n); \quad i = 1, \ldots, M. \end{aligned} \tag{1}$$

The goal of the beamforming is to obtain an estimate of $u(n)$ by filtering and then summing the microphone signals

$$y(n) = \sum_{i=1}^{M} \tilde{f}_i(n) * x_i(n). \tag{2}$$

Here, $\tilde{f}_i(n) = f_i(-n)$ is the impulse response of the filter of the i-th microphone signal. The filtering operation is preferably done in the frequency domain:

$$\begin{aligned} Y(k) &= \sum_{i=1}^{M} F_i^*(k) \cdot X_i(k) \\ &= \mathbf{F}^H(k) \cdot \mathbf{X}(k); \quad k = 0, \ldots, L-1. \end{aligned} \tag{3}$$

Here, k denotes the frequency bin and L is the DFT-length. The frame index has been omitted for ease of notation. In the following we prefer the vector notation, e.g. $\mathbf{X} = (X_1, \ldots, X_M)^T$. In (3) $(\cdot)^H$ denotes Hermitian transpose.

If the desired signal and the noise are uncorrelated the power spectral density of the FSB output is obtained as

$$\begin{aligned} \Phi_{YY}(k) &= \mathbf{F}^H(k)\mathbf{\Phi_{XX}}(k)\mathbf{F}(k) \\ &= \mathbf{F}^H(k)\left(\mathbf{\Phi_{SS}}(k) + \mathbf{\Phi_{NN}}(k)\right)\mathbf{F}(k) \end{aligned} \tag{4}$$

where $\mathbf{\Phi_{XX}}(k)$, $\mathbf{\Phi_{SS}}(k)$ and $\mathbf{\Phi_{NN}}(k)$ are the PSD matrices of the microphone signals, the speech and noise terms, respectively. Let us now assume spherically white noise

$$\mathbf{\Phi_{NN}}(k) = \sigma_N^2 \cdot \mathbf{I}_M \tag{5}$$

($\mathbf{I}_M$: identity matrix of dimension $M \times M$). Using this in eq. (4) one can readily see that the signal-to-noise ratio of the output signal $y(n)$ is maximized, if the power of $y(n)$ is maximized under the constraint

$$\sum_{i=1}^{M} |F_i(k)|^2 = 1, \qquad \forall k. \tag{6}$$

The constrained maximization problem

$$\max_{\mathbf{F}^H(k)} \mathbf{F}^H(k)\mathbf{\Phi_{XX}}(k)\mathbf{F}(k) \text{ subject to } \mathbf{F}^H(k)\mathbf{F}(k) = 1 \tag{7}$$

is solved using the method of Lagrange [7]. We define the real cost function

$$J = \mathbf{F}^H(k)\mathbf{\Phi_{XX}}(k)\mathbf{F}(k) + 2\lambda(\mathbf{F}^H(k)\mathbf{F}(k) - 1) \tag{8}$$

and compute the gradient

$$\nabla_{\mathbf{F}} J = 2\mathbf{\Phi_{XX}}(k)\mathbf{F}(k) + 2\lambda\mathbf{F}(k), \tag{9}$$

which is set to zero:

$$\mathbf{\Phi_{XX}}(k)\mathbf{F}(k) + \lambda\mathbf{F}(k) = 0. \tag{10}$$

Obviously we have to conduct an eigenvalue decomposition of the cross-power spectral density matrix of the sensor signals. Note that this has to be done for each frequency bin separately.

3. DETERMINISTIC GRADIENT ASCENT

To iteratively solve eq. (10) we develop a deterministic gradient ascent scheme:

$$\mathbf{F}(\kappa+1) = \mathbf{F}(\kappa) + \frac{\mu}{2}\nabla_{\mathbf{F}} J \Big|_{\mathbf{F}=\mathbf{F}(\kappa)}, \tag{11}$$

where κ counts the iterations, and μ is the step size parameter. Here and in the following we omit the frequency bin index k for ease of notation. Using eq. (8) and determining the Lagrange multiplier λ from the constraint $\mathbf{F}^H(\kappa+1)\mathbf{F}(\kappa+1) = 1$ we eventually arrive at the following iteration, after neglecting terms of order $\mathcal{O}(\mu^2)$,

$$\begin{aligned} \mathbf{F}(\kappa+1) = \mathbf{F}(\kappa) &+ \frac{1}{2}\left[\frac{1}{\mathbf{F}^H(\kappa)\mathbf{F}(\kappa)} - 1\right]\mathbf{F}(\kappa) \\ &+ \mu\left[\mathbf{\phi_{XX}}\mathbf{F}(\kappa) - \frac{\mathbf{F}^H(\kappa)\mathbf{\phi_{XX}}\mathbf{F}(\kappa)}{\mathbf{F}^H(\kappa)\mathbf{F}(\kappa)}\mathbf{F}(\kappa)\right]. \end{aligned} \tag{12}$$

If the constraint is satisfied, i.e. $\mathbf{F}^H(\kappa)\mathbf{F}(\kappa) = 1$, the term in the first pair of brackets vanishes, and if $\mathbf{F}(\kappa)$ is an eigenvector of $\mathbf{\phi_{XX}}$, also the term in the second pair of brackets is zero (recognize the Rayleigh coefficient!). We therefore have $\mathbf{F}(\kappa+1) = \mathbf{F}(\kappa)$, which shows that the iteration indeed computes the requested eigenvector. A detailed proof, that the weight vector $\mathbf{F}(\kappa)$ converges to the first principal component can be done similar to [8].

It is however important to note that it is not assumed beforehand that the constraint is met! The term in the first pair of brackets has a "stabilizing" effect: if $\mathbf{F}^H(\kappa)\mathbf{F}(\kappa) > 1$ the term is negative and the norm of $\mathbf{F}(\kappa)$ tends to be reduced. Similarly, a positive correction term results if $\mathbf{F}^H(\kappa)\mathbf{F}(\kappa) < 1$.

4. STOCHASTIC GRADIENT ASCENT

The cross power spectral density matrix of the sensor signals is not known in practice. We therefore replace it by the instantaneous estimate

$$\mathbf{\phi_{XX}} \approx \mathbf{X}(m)\mathbf{X}^H(m). \tag{13}$$

Further, in every iteration a new block of data is processed, i.e. the iteration index κ is replaced by the frame index m. Using this in eq. (12) we obtain a stochastic gradient ascent algorithm

$$\begin{aligned} \mathbf{F}(m+1) = \mathbf{F}(m) &+ \frac{1}{2}\left[\frac{1}{\mathbf{F}^H(m)\mathbf{F}(m)} - 1\right]\mathbf{F}(m) \\ &+ \mu\left[\mathbf{X}(m)\mathbf{X}^H(m)\mathbf{F}(m) - \frac{\mathbf{F}^H(m)\mathbf{X}(m)\mathbf{X}^H(m)\mathbf{F}(m)}{\mathbf{F}^H(m)\mathbf{F}(m)}\mathbf{F}(m)\right]. \end{aligned} \tag{14}$$

Employing (3) and rearranging terms we finally obtain

$$\begin{aligned} \mathbf{F}(m+1) = \mathbf{F}(m)&\frac{1 + \mathbf{F}^H(m)\mathbf{F}(m)}{2\mathbf{F}^H(m)\mathbf{F}(m)} \\ &+ \mu Y^*(m)\left[\mathbf{X}(m) - \frac{\mathbf{F}(m)Y(m)}{\mathbf{F}^H(m)\mathbf{F}(m)}\right]. \end{aligned} \tag{15}$$

Note that in this and the last section we omitted the frequency index k. However it has to be kept in mind that the aforementioned iteration has to be carried out for every frequency bin separately. Due to the block frequency nature of the algorithm it is easy to use frequency-dependent step sizes that are inversely proportional to the power levels in the DFT frequency bins, as proposed e.g. in [9], for improved convergence speed.

If we assume that the constraint is met, i.e. $\mathbf{F}^H(m)\mathbf{F}(m) = 1$ then the equation can be simplified to

$$\mathbf{F}(m+1) = \mathbf{F}(m) + \mu Y^*(m)\left[\mathbf{X}(m) - \mathbf{F}(m)Y(m)\right]. \tag{16}$$

This result is actually well known in the neural networks literature under the name *Hebbian-based maximum eigenfilter* or *Oja's rule* [6, 10].

5. EXPERIMENTAL RESULTS

While a complete theoretical stability and convergence analysis has yet to be conducted, we are going to evaluate convergence rate and robustness by an experimental study in the following sections.

5.1. Comparison of FSB with DSB

In a first set of experiments the performance advantage of Filter-and-Sum beamforming (FSB) over conventional Delay-and-Sum beamforming (DSB) was quantified.

We simulated various room reverberation times RT_{60} and different sensor constellations with the image method [11] and computed the *Energy Decay Curve* $EDC(j)$

$$EDC(j) = \sum_{n=j}^{\infty} h^2(n) \; / \; \sum_{n=0}^{\infty} h^2(n) \tag{17}$$

where the impulse response $h(n)$ has the following interpretation in the three setups investigated:

- *Single Mic.*: $h(n)$ is the room impulse response (RIR) from the source to a single microphone.
- *DSB*: $h(n)$ is the sum of the properly delayed (according to the direct path length differences) RIRs from the source to M sensors. This setup corresponds to a perfectly operating Delay-and-Sum beamformer.
- *FSB*: $h(n)$ is the overall impulse response from the source to the Filter-and-Sum (FSB) beamformer output: the RIRs from the source to the M sensors are convolved with the respective FSB filter impulse responses and then summed. The FIR filter lengths of the FSB was set to 64 (at sampling rate of 8kHz) and serve to align the direct paths from source to sensors and, in addition, capture early reflections.

While the exact results depend on various parameters, we observed in all experiments the same trends as exemplified in Fig. 1. It is the result of the simulation for a room of size (6m)x(5m)x(3m) with four linearly arranged microphones (distance of 10cm) and a room reverberation time RT_{60} of 0.32s. The acoustical source was placed 3m away from the sensors at an angle of 60 degrees relative to broadside.

As can be seen, the ratio between the power in the direct to the power in the diffuse part of $h(n)$, which determines the clarity of the output signal, is increased from a single microphone to a DSB and even more increased by employing an FSB.

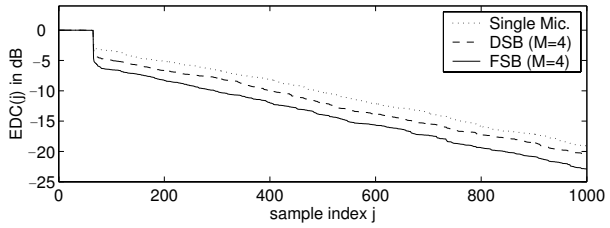

Fig. 1. *Energy decay curves for a single RIR, the DSB system and the FSB system with reverberation* $RT_{60} = 0.32s$.

5.2. Stability of Deterministic Gradient Ascent Rule

The step size parameter μ is crucial for the stability and rate of convergence of the adaptation rule. For fast adaptation of the FSB μ should be set to the largest value μ_{max} which still guarantees stability. Since an analytic computation of μ_{max} from eq. (12) is rather complicated we adopted the following simulation approach.

Any coefficient vector $\mathbf{F}(\kappa)$ can be expressed as a linear combination of the eigenvectors $\mathbf{e}_1, \ldots, \mathbf{e}_M$ of $\boldsymbol{\phi}_{\mathbf{XX}}$:

$$\mathbf{F}(\kappa) = \sum_{i=1}^{M} \alpha_i(\kappa)\mathbf{e}_i \tag{18}$$

where $\alpha_i(\kappa)$ denotes the weight of $\mathbf{e}_i$ at iteration κ. Using eq. (18) the deterministic update rule (12) leads to

$$\begin{aligned}\boldsymbol{\alpha}(\kappa+1) = \boldsymbol{\alpha}(\kappa)\frac{1+\boldsymbol{\alpha}^T(\kappa)\boldsymbol{\alpha}(\kappa)}{2\boldsymbol{\alpha}^T(\kappa)\boldsymbol{\alpha}(\kappa)} \\ + \mu_L\left[\boldsymbol{\Lambda} - diag\left\{\frac{\boldsymbol{\alpha}^T(\kappa)\boldsymbol{\Lambda}\boldsymbol{\alpha}(\kappa)}{\boldsymbol{\alpha}^T(\kappa)\boldsymbol{\alpha}(\kappa)}\right\}\right]\boldsymbol{\alpha}(\kappa)\end{aligned} \tag{19}$$

with $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_M)^T$. $\boldsymbol{\Lambda} = diag\{\lambda_i\}$ denotes the diagonal matrix of eigenvalues arranged in descending order $\lambda_1 > \lambda_2 \geq \ldots \geq \lambda_M > 0$.

Similarly, a recursion in the weight vector $\boldsymbol{\alpha}$ can be derived for the deterministic version of Oja's rule (16):

$$\boldsymbol{\alpha}(\kappa+1) = \boldsymbol{\alpha}(\kappa) + \mu_O\left[\boldsymbol{\Lambda} - diag\left\{\boldsymbol{\alpha}^T(\kappa)\boldsymbol{\Lambda}\boldsymbol{\alpha}(\kappa)\right\}\right]\boldsymbol{\alpha}(\kappa). \tag{20}$$

The step size parameter μ has been given the indices "O" in **O**ja's rule and "L" in the novel rule derived from **L**agrange approach. Using these recursions we studied the convergence of the weight vector $\boldsymbol{\alpha}$ from its initial value $\boldsymbol{\alpha}(0)$ to its final value $\boldsymbol{\alpha}(\infty) = (1, 0, \ldots, 0)^T$ as a function of the initial deviation $K := \| \boldsymbol{\alpha}(0) \|^2$ from the constraint $\| \boldsymbol{\alpha}(0) \|^2 = 1$, the eigenvalue spread $\chi = \lambda_{max}/\lambda_{min}$, and of the step size parameter μ.

We found by simulations that the maximum step size μ_{max} for eq. (19) and (20) can be written as

$$\mu_{max} = \frac{2}{\xi_{min} \cdot \lambda_{max}}. \tag{21}$$

From experiments we derived the following upper bounds which guarantee stability of the deterministic gradient ascent rules:

$$\xi_{Omin} < \xi_O(\chi, K) = 1 + \frac{K-1}{2}(1 + \frac{1}{\chi}) < K \tag{22}$$

$$\xi_{Nmin} < \xi_L(\chi) = 1 - \frac{1}{\chi} < 1. \tag{23}$$

Fig. 2 compares the experimental results for $K = 50$ (cross markers) and $K = 100$ (square markers) with the upper bounds $\xi_O(\chi, K)$ and $\xi_L(\chi)$, respectively. Note, that unlike $\xi_O(\chi, K)$ for Oja's rule, $\xi_L(\chi)$ is *independent* of the initial deviation K! This is a very desirable property, since K is very hard to predict for a moving speaker. In experiments we observed temporary constraint mismatches K due to fast speaker movement, which can assume values of a few hundred. Accordingly μ_O, the step size parameter in Oja's rule, has to be chosen very conservatively, whereas μ_L can be chosen much larger:

$$0 < \mu_O < \frac{2}{K\lambda_{max}}$$

$$0 < \mu_L < \frac{2}{\lambda_{max}}.$$

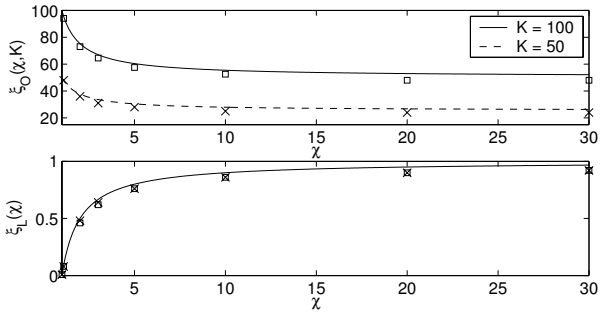

Fig. 2. *Comparison of experimental results with upper bound for Oja's rule (upper figure) and the novel rule (lower figure).*

5.3. Learning Curves

In another set of experiments we studied the evolution of the squared error

$$e^2(\kappa) = \| \boldsymbol{\alpha}(\infty) - \boldsymbol{\alpha}(\kappa) \|^2 \tag{24}$$

over the number of iterations κ.

Fig. 3 shows the squared error of Oja's rule $e_O^2(\kappa)$ and of the Lagrange method $e_L^2(\kappa)$, both normalized to $e^2(0)$. The initial constraint mismatch was set to $K = 10$, and the number of channels was $M = 4$. The squared error is shown for the same normalized step sizes γ for both update rules, where

$$\gamma = \frac{\mu}{2/(K\lambda_{max})}. \tag{25}$$

Note, that $\gamma = 1$ is the maximum value that still guarantees stability of Oja's rule, whereas γ could have been chosen larger (and thus enabling even faster convergence) for the novel update rule.

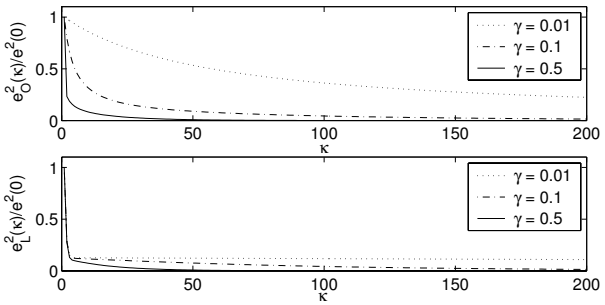

Fig. 3. *Convergence rate in terms of the squared error with varying step size parameter.*

The increased convergence rate of the proposed algorithm may be attributed to the fact, that $\boldsymbol{\alpha}^T(\kappa)\boldsymbol{\alpha}(\kappa)$ is compliant with the constraint already after a few iterations.

In Fig. 4 we have plotted the values of $\alpha_i(\kappa), i = 1, \ldots, 4$ and $\boldsymbol{\alpha}^T(\kappa)\boldsymbol{\alpha}(\kappa)$ versus iteration index κ for $K = 10$ and $\gamma = 0.01$. It can be seen that α_1, the weight of the eigenvector corresponding to the largest eigenvalue, converges to one and the other α_i converge to zero for both rules. However, the constraint (dashed line) is met much faster with the novel rule as compared to Oja's rule.

6. CONCLUSIONS

In this paper we have proposed a novel adaptation scheme for acoustic Filter-and-Sum beamforming, which is based on adaptive principal component analysis. The well-known Oja's rule is obtained as a special case. The experimental results show the superiority of Filter-and-Sum beamforming compared to Delay-and-Sum beamforming and demonstrate the excellent convergence rate and robustness of the proposed method. The range of allowable step sizes and the convergence rate does not depend on temporary constraint mismatches which can be caused by fast speaker movement.

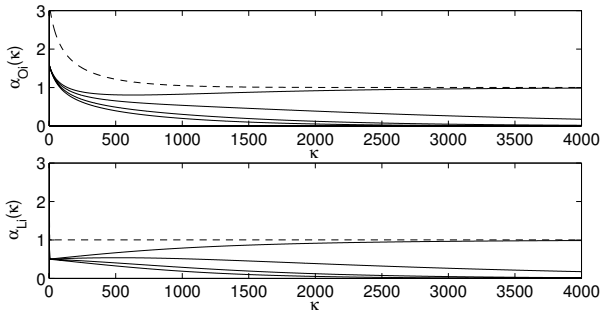

Fig. 4. *Convergence rate of $\boldsymbol{\alpha}^T(\kappa)\boldsymbol{\alpha}(\kappa)$ (dashed) and α_i (solid) for Oja's rule in the upper figure and the novel Lagrange rule in the lower figure for $K = 10$ and $\gamma = 0.01$.*

7. REFERENCES

[1] J. H. DiBiase, H. F. Siverman, and M.S. Brandstein, "Robust localization in reverberant rooms", in *Microphone Arrays: Signal Processing Techniques and Applications*, Springer Verlag, 2001.

[2] D. N. Zotkin and R. Duraiswami, "Accelerated speech source localization via a hierarchical search of steered response power", *IEEE Trans. on Speech and Audio Processing*, vol. 12, no. 5, pp. 499-508,Sep. 2004.

[3] E. Warsitz, R. Haeb-Umbach, and S. Peschke, "Adaptive beamforming combined with particle filtering for acoustic source localization", in *Proc. ICSLP*, Jeju, Corea, Oct. 2004.

[4] O. Frost, "An algorithm for linearly constrained adaptive array processing", *Proceedings of the IEEE*, vol. 60, no. 8, pp. 927-935, Aug. 1972.

[5] 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, Jan. 1982.

[6] E. Oja, "A simplified neuron model as a principal component analyzer", *J. of Mathematical Biology*, 15:267-273, 1982.

[7] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, 1996.

[8] K. I. Diamantaras, S. Y. Kung, *Principal Component Neural Networks: Theory and Applications*, John Wiley & Sons Inc., 1996.

[9] J. J. Shynk, "Frequency-Domain and Multirate Adaptive Filtering", *IEEE Signal Processing Magazine*, vol. 9, no. 1, pp. 14-39, Jan. 1992.

[10] S. Haykin, *Neural Networks*, Prentice Hall, 1999.

[11] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics", *J. Acoust. Soc. Amer.*, vol. 107, no. 4, pp. 943-950, 1979.