IEEE Asilomar Conference on Signals, Systems, and Computers, November 9-12, 2003

Wideband Algorithms versus Narrowband Algorithms for Adaptive Filtering in the DFT Domain

Walter Kellermann and Herbert Buchner
University of Erlangen-Nuremberg
Cauerstr. 7, 91058 Erlangen, Germany*
{wk,buchner}@LNT.de

Abstract

Adaptive filtering in the DFT domain is popular for its computational efficiency and its attractive convergence properties resulting from the applicability of the FFT and separate adaptation of individual DFT bins. Narrowband algorithms assume a complete decoupling of different frequency bins, which corresponds to assuming a circulant structure for the input data matrix. Wideband designs account for the difference between the actual Töplitz structure and the circulant structure by introducing additional constraints. In this contribution, we show that a wideband approach with rigorous implementation of appropriate constraints leads to highly efficient algorithms with excellent convergence properties. As examples we consider multichannel acoustic echo cancellation (MC-AEC) and blind source separation (BSS) of convolutive mixtures.

1 Introduction

Adaptive filtering in the DFT domain not only exploits the computational efficiency of linear filtering in the DFT domain [1], but it also promises superior convergence behavior compared to most time-domain adaptation algorithms [2].

In many cases, such DFT-domain adaptive systems are designed under the assumption that individual frequencies of the input signal(s) can be considered independently from each other. This implicitly assumes that the input can be described by a set of countable complex exponentials as eigenfunctions of the linear system. This "narrowband signal model" is especially popular in adaptive beamforming and blind source separation (BSS). However, especially for nonstationary signals, time-limitation implying continuous spectra has to be accounted for by a "wideband signal model". In this paper we show that

- the difference between narrowband and wideband signal model reduces to the difference between circulant and Töplitz matrices,
- this difference is acknowledged in some practically important cases already in a more or less empirical manner,
- a systematical extension of Töplitz matrices to circulant matrices leads to new algorithms for wideband signals that include the already known remedies as special cases and exhibit truly optimum behavior,
- the practical impact of a wideband signal model for MC-AEC and convolutive BSS is significant.

2 Wideband vs. narrowband signals

In the discrete time domain, the output signal $\hat{y}(n)$ of an FIR filter at time n is

$$\hat{y}(n) = \mathbf{x}^T(n)\hat{\mathbf{h}}, \tag{1}$$

where the input signal vector $\mathbf{x}$ and the impulse response $\hat{\mathbf{h}}$ of length L are given by

$$\mathbf{x}(n) = [x(n), x(n-1), \cdots, x(n-L+1)]^T, \tag{2}$$

$$\hat{\mathbf{h}} = [\hat{h}_0, \hat{h}_1, \cdots, \hat{h}_{L-1}]^T. \tag{3}$$

A block of L output samples $\hat{\mathbf{y}}(m)$ reads

$$\hat{\mathbf{y}}(m) = \mathbf{U}^T(m)\hat{\mathbf{h}}, \tag{4}$$

where m is the block time index, and

$$\hat{\mathbf{y}}(m) = [\hat{y}(mL), \cdots, \hat{y}(mL+L-1)]^T, \tag{5}$$

$$\mathbf{U}(m) = [\mathbf{x}(mL), \cdots, \mathbf{x}(mL+L-1)]. \tag{6}$$

Obviously, $\mathbf{U}$ is a Töplitz matrix of size $L \times L$:

$$\mathbf{U}^T(m) = \begin{bmatrix} x(mL) & \cdots & \cdots & x(mL-L+1) \\ x(mL+1) & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ x(mL+L-1) & \cdots & \cdots & x(mL) \end{bmatrix}. \tag{7}$$

We observe that $\mathbf{U}^T(m)$ is circulant only if $x(mL+k) = x(mL-L+k)$, i.e., if $x(n)$ is a periodic signal with period L/κ, $\kappa \in \{1, 2, \ldots, L\}$. We note that such signals $x(n)$ are described by a finite Fourier series

$$x(n) = \sum_{i=0}^{L-1} c_i e^{j\frac{2\pi in}{L}}, \tag{8}$$

*This work was partly supported by the ANITA project funded by the European Commission under contract IST-2001-34327.

i.e., by at most L complex exponentials. Applying the $L \times L$ DFT matrix $\mathbf{F}_L$ with elements $f_{ik} = e^{-j2\pi ik/L}$ to (4) we write

$$\mathbf{F}_L \hat{\mathbf{y}}(m) = \mathbf{F}_L \mathbf{U}^T(m)\hat{\mathbf{h}} = \mathbf{F}_L \mathbf{U}^T(m)\mathbf{F}_L^{-1}\mathbf{F}_L\hat{\mathbf{h}}, \quad (9)$$

and note that iff $\mathbf{U}^T(m)$ is a circulant fulfilling (8), then $\mathbf{U}^T(m)$ is diagonalized by the DFT matrix:

$$\begin{aligned} \mathbf{F}_L \mathbf{U}^T(m)\mathbf{F}_L^{-1} &= \operatorname{diag}\{\mathbf{F}_L \mathbf{x}_{*\mathbf{1}}(mL)\} \\ &=: \operatorname{diag}\{\underline{\mathbf{x}}(m)\} =: \underline{\mathbf{X}}(m), \quad (10) \end{aligned}$$

where $\mathbf{x}_{*\mathbf{1}}(mL)$ denotes the first column of $\mathbf{U}^T$. Thus, the linear filtering operation is equivalent to a circular convolution and can be written as

$$\underline{\hat{\mathbf{y}}}(m) = \underline{\mathbf{X}}(m)\underline{\hat{\mathbf{h}}}, \quad (11)$$

where $\underline{\hat{\mathbf{y}}}(m), \underline{\mathbf{x}}(m), \underline{\hat{\mathbf{h}}}$ are the DFTs of $\hat{\mathbf{y}}(m), \mathbf{x}_{*1}(m), \hat{\mathbf{h}}$, respectively. If (8) is not fulfilled, $\underline{\mathbf{x}}(m)$ cannot capture the differences between the actual Töplitz matrix and the assumed circulant matrix and (11) does not describe linear filtering [1].

To exploit the efficiency of FFTs for linear filtering of arbitrary signals $x(n)$, the evaluation of (4) in the DFT domain should be carried out so that the difference to a circulant $\mathbf{U}^T(m)$ does not affect the output. The two popular approaches coincide with the respective signal models:

Wideband signal model: Exploit the fact that for arbitrary $x(n)$ (4) yields equivalent results for circular and Töplitz matrices if only the first $K < L$ elements of $\hat{\mathbf{h}}$ are nonzero and only the first $L-K+1$ elements of $\hat{\mathbf{y}}(m)$ are considered as useful output. This leads directly to the 'overlap-add' and 'overlap-save' algorithms [1].

Narrowband signal model: Choose L sufficiently large so that $x(n)$ seems to be reasonably well represented by (8), which still implies that $x(n)$ has a discrete Fourier spectrum.

3 A general approach to adaptive linear filtering in the DFT domain

The distinction between narrowband and wideband signal models from Section 2 is now carried over to the design of adaptive filtering for arbitrary signals. By way of a simple and well-known example, we introduce a very general design method for wideband DFT-domain adaptive filtering algorithms that recently led to powerful algorithms for very challenging applications such as MC-AEC and convolutive BSS [11, 12, 16, 17, 18]. This concept is applicable to a broad class of adaptive filtering problems: Supervised and unsupervised systems with single or multiple inputs and outputs, respectively, with optimization criteria based on second-order or higher-order statistics, with according gradients of first order or higher order for minimization of the respective cost functions.

For simplicity, the main idea is illustrated for the case of a linear single-channel supervised adaptive filter: According to Fig. 1, the error signal at time n between the output of the adaptive filter $\hat{y}(n)$ and the desired output signal $y(n)$ is given by

$$e(n) = y(n) - \hat{y}(n). \quad (12)$$

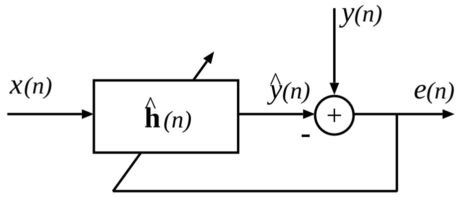

Figure 1: Single-channel supervised adaptive filter.

For the block error signal of length L we write based on (12) and (1)

$$\mathbf{e}(m) = \mathbf{y}(m) - \hat{\mathbf{y}}(m), \quad (13)$$

where

$$\begin{aligned} \mathbf{e}(m) &= [e(mL), \cdots, e(mL+L-1)]^T, \quad (14) \\ \mathbf{y}(m) &= [y(mL), \cdots, y(mL+L-1)]^T. \quad (15) \end{aligned}$$

This error definition and its DFT form the basis for the adaptive filtering algorithms below.

3.1 Narrowband signal model

Adaptive filtering in the DFT domain based on a narrowband signal model (see Section 2) is characterized by the following steps [4] as illustrated in Fig. 2:

Step 1: Running or (possibly overlapping) block DFTs for each input channel (short-time Fourier transformation (STFT))

Step 2: Independent scalar adaptive filtering for each frequency bin

Step 3: Synthesis of output signal(s) by inverse DFT(s)

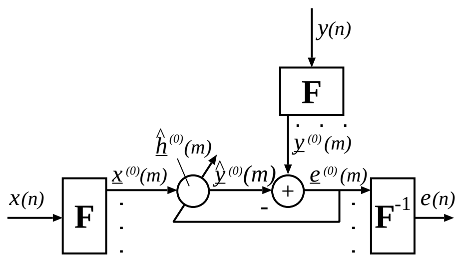

Figure 2: Illustration of narrowband approach.

As an example we consider an RLS-type adaptation algorithm[1], which minimizes the exponentially weighted

[1]RLS-type algorithms are chosen because of their optimum convergence behavior w.r.t. a given quadratic cost function

squared error in the k-th DFT bin, $|\underline{e}^{(k)}(i)|^2$, for $k = 0, \dots, N-1$:

$$J_{\mathrm{NB},k}(\underline{\hat{h}}^{(k)}(m)) = (1-\lambda)\sum_{i=0}^{m} \lambda^{m-i} |\underline{e}^{(k)}(i)|^2, \quad (16)$$

This narrowband DFT-domain RLS algorithm operates independently in each DFT bin and can be summarized (based on (11) and (13)) as follows:

$$\begin{aligned}
\mathbf{S}(m) &= \lambda \mathbf{S}(m-1) + (1-\lambda)\underline{\mathbf{X}}^H(m)\underline{\mathbf{X}}(m), && (17)\\
\underline{\mathbf{e}}(m) &= \underline{\mathbf{y}}(m) - \underline{\mathbf{X}}(m)\underline{\hat{\mathbf{h}}}(m-1), && (18)\\
\underline{\hat{\mathbf{h}}}(m) &= \underline{\hat{\mathbf{h}}}(m-1) \\
&\quad + (1-\lambda)\mathbf{S}^{-1}(m)\underline{\mathbf{X}}^H(m)\underline{\mathbf{e}}(m). && (19)
\end{aligned}$$

As its main advantages, the low computational complexity and the structural simplicity of the concept make it seemingly attractive for implementation in many application scenarios. However, major disadvantages when used for arbitrary signals $x(n)$ are:

Sub-optimum convergence properties. In the case of supervised adaptive filtering, convergence to the Wiener solution is not assured due to coupling of the DFT bins.

Large transform lengths L are needed to justify the narrowband signal model (see Section 2).

Signal distortion due to artifacts caused by circular filtering effects (i.e., insufficient approximation of (4) in the DFT domain).

3.2 Wideband signal model

To avoid errors caused by differences between the desired circulant and the actual Töplitz matrix for wideband signals, linear filtering in the DFT domain (i.e. overlap add/save) requires windowing/zero-padding of data, i.e., the introduction of constraints. The crucial question we want to answer here is: How can these constraints be systematically introduced into adaptive filtering algorithms?

As a key observation, we note that any Töplitz matrix $\mathbf{U}$ can be expanded to a circulant [3, 11, 12]

$$\mathbf{C}(m) = \begin{bmatrix} \mathbf{U}'^T(m) & \mathbf{U}^T(m) \\ \mathbf{U}^T(m) & \mathbf{U}'^T(m) \end{bmatrix}, \quad (20)$$

where $\mathbf{U}'$ is also a Töplitz matrix using the same data samples as $\mathbf{U}(m)$ (except for an arbitrary element along the main diagonal). This allows for a compact and exact matrix formulation of the Töplitz matrix $\mathbf{U}^T(m)$ in terms of the DFT of the input signal $x(n)$ [12]:

$$\begin{aligned}
\mathbf{U}^T(m) &= \mathbf{W}^{01}_{L\times 2L}\mathbf{C}(m)\mathbf{W}^{10}_{2L\times L} && (21)\\
&= \mathbf{W}^{01}_{L\times 2L}\mathbf{F}^{-1}_{2L}\underline{\mathbf{X}}_{2L}(m)\mathbf{F}_{2L}\mathbf{W}^{10}_{2L\times L}. && (22)
\end{aligned}$$

Here, we introduced the windowing matrices

$$\begin{aligned}
\mathbf{W}^{01}_{L\times 2L} &= [\mathbf{0}_{L\times L}, \mathbf{I}_{L\times L}],\\
\mathbf{W}^{10}_{2L\times L} &= [\mathbf{I}_{L\times L}, \mathbf{0}_{L\times L}]^T,
\end{aligned}$$

and the diagonal matrix $\underline{\mathbf{X}}_{2L}(m)$, which can be written in terms of the elements of the first columns of $\mathbf{C}(m)$,

$$\underline{\mathbf{X}}_{2L}(m) = \mathrm{diag}\{\mathbf{F}_{2L}[x(mL-L), \cdots, x(mL+L-1)]^T\}. \quad (23)$$

Based on this, a general method for designing wideband adaptive filtering algorithms involves:

Step 1: Formulate an exact wideband cost function in terms of DFT matrices and suitable windows based on (22).

Step 2: Minimize the cost function w.r.t. the DFT-domain filter coefficients in order to obtain an "exact" algorithm.

Step 3: Examine the various constraints appearing in the algorithm resulting from Step 2 and *selectively approximate* them for more efficient algorithms.

Analogously to the narrowband case we consider as an example an RLS-type DFT-domain supervised single-input/single-output adaptive filter which minimizes the cost function

$$J(m) = (1-\lambda)\sum_{i=0}^{m}\lambda^{m-i}\underline{\mathbf{e}}^H(i)\underline{\mathbf{e}}(i). \quad (24)$$

A rigorous derivation [12] yields the following generic algorithm (as illustrated in Fig. 3):

$$\begin{aligned}
\mathbf{S}_{2L}(m) &= \lambda\mathbf{S}_{2L}(m-1) \\
&\quad + (1-\lambda)\underline{\mathbf{X}}^H_{2L}(m)\mathbf{G}_1\underline{\mathbf{X}}_{2L}(m) && (25)\\
\underline{\mathbf{e}}_{2L}(m) &= \underline{\mathbf{y}}_{2L}(m) - \mathbf{G}_2\underline{\mathbf{X}}_{2L}(m)\underline{\hat{\mathbf{h}}}_{2L}(m-1) && (26)\\
\underline{\hat{\mathbf{h}}}_{2L}(m) &= \underline{\hat{\mathbf{h}}}_{2L}(m-1) + (1-\lambda)\mathbf{G}_3\mathbf{S}^{-1}_{2L}(m) \\
&\quad \cdot \underline{\mathbf{X}}^H_{2L}(m)\underline{\mathbf{e}}_{2L}(m), && (27)
\end{aligned}$$

where

$$\begin{aligned}
\underline{\mathbf{e}}_{2L}(m) &= \mathbf{F}_{2L}\begin{bmatrix}\mathbf{0}_{L\times 1}\\ \mathbf{e}(m)\end{bmatrix},\\
\underline{\mathbf{y}}_{2L}(m) &= \mathbf{F}_{2L}\begin{bmatrix}\mathbf{0}_{L\times 1}\\ \mathbf{y}(m)\end{bmatrix},\\
\underline{\hat{\mathbf{h}}}_{2L}(m) &= \mathbf{F}_{2L}\begin{bmatrix}\hat{\mathbf{h}}(m)\\ \mathbf{0}_{L\times 1}\end{bmatrix},\\
\mathbf{G}_1 &= \mathbf{G}_2 = \mathbf{F}_{2L}\mathbf{W}^{01}_{2L\times 2L}\mathbf{F}^{-1}_{2L},\\
\mathbf{G}_3 &= \mathbf{F}_{2L}\mathbf{W}^{10}_{2L\times 2L}\mathbf{F}^{-1}_{2L},\\
\mathbf{W}^{01}_{2L\times 2L} &= \begin{bmatrix}\mathbf{0}_{L\times L} & \mathbf{0}_{L\times L}\\ \mathbf{0}_{L\times L} & \mathbf{I}_{L\times L}\end{bmatrix},\\
\mathbf{W}^{10}_{2L\times 2L} &= \begin{bmatrix}\mathbf{I}_{L\times L} & \mathbf{0}_{L\times L}\\ \mathbf{0}_{L\times L} & \mathbf{0}_{L\times L}\end{bmatrix}.
\end{aligned}$$

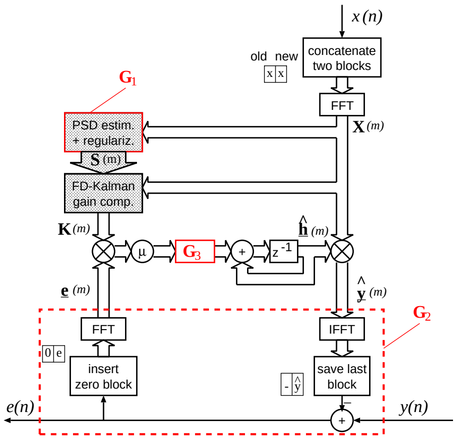

Figure 3: Generic SISO supervised FDAF as an example for the wideband approach.

This example highlights the structural similarity between wideband and narrowband adaptive filtering, but also the decisive differences: The wideband structure strictly prevents circular filtering artifacts and accounts for the coupling of different DFT bins by the constraints $\mathbf{G}_i$. Some of the constraints in this example have been previously proposed: $\mathbf{G}_2$ in [5, 6, 7] and $\mathbf{G}_3$ in [5] and the resulting algorithms can be seen as efficient versions of the generic algorithm. Unlike the narrowband algorithm, the generic wideband algorithm and the mentioned approximations can be shown to converge to the Wiener solution for arbitrary wideband signals [12]. Moreover, the generic wideband algorithm can be straightforwardly extended to partitioned FIR filters [8, 9, 10, 12] and can also be generalized to the multichannel and MIMO cases [11, 12]. Beyond that, the concept is also applicable to non-quadratic cost functions [13] which are useful to obtain robust algorithms in the presence of nongaussian distortions.

4 Applications

To demonstrate the superiority of rigorously derived wideband algorithms over narrowband algorithms, we consider two challenging adaptive filtering tasks.

4.1 MC-AEC

Multichannel acoustic echo cancellation (MC-AEC) differs from single-channel AEC mainly in that - by use of a single error signal - multiple systems with correlated inputs must be identified instead of a single system [11]. For RLS-type algorithms this implies a condition number of the autocorrelation matrix which worsens with increasing channel number and, accordingly, requires better adaptation algorithms [12]. The impact of a rigorously derived wideband DFT-domain algorithm becomes obvious from simulation results as shown in Fig.4. For $P = 1, 2, 5$ loudspeaker channels emitting filtered speech, P adaptive FIR filters of length $L = 1024$ were used to model impulse responses of length 4096. While the wideband algorithm according to [12] was used with DFT-length $2L$ and block overlap factor of 4 (for $P = 1, 2$) and 16 (for $P = 5$), for the narrowband algorithm a tenfold DFT length ($20L$) and overlap factor (40 and 160 respectively) had to be chosen to achieve the convergence behavior of Fig.4. Clearly, this choice reduces the potential computational advantage of the narrowband approach while the convergence of the wideband algorithm to higher echo suppression (ERLE: echo return loss enhancement) is by far superior.

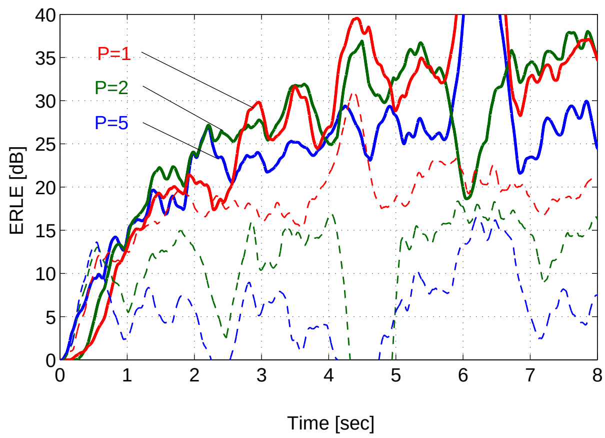

Figure 4: MC-AEC performance of the wideband approach (solid) and the narrowband approach (dashed).

4.2 BSS for convolutive mixtures

Blind source separation aims at separating P sources from Q sensor signals containing convolutive mixtures of the desired source signals, which are assumed to be statistically independent. By transforming the signals into the DFT domain, the convolutive mixtures in the time domain become scalar mixtures for each DFT bin. So far, most DFT-domain BSS algorithms are based on the narrowband signal model, which implies decoupled scalar BSS problems in each DFT bin. However, this inevitably causes the so-called internal permutation problem: E.g., for $P = Q = 2$, the separated DFT bins for sources A and B cannot be aligned so that all bins with components of A appear at one output of the BSS system, while all bins for B appear at the other. Whereas most current DFT-domain BSS algorithms include empirical repair mechanisms for the internal permutation problem, the circular filtering problem remains [15]. For $P, Q > 2$ these difficulties are known to become even more serious. A strictly wideband approach, however, inherently solves both problems [16, 18]. This is illustrated in Fig.5 for $P = Q = 2$: The mixtures of two speech signals of 10s duration, sampled at 16kHz, were recorded in a room with reverberation time $T_{60} = 150$ms. A wideband off-line BSS algorithm according to [16] and the corresponding narrowband algorithm were applied with demixing filter length $L = 512$. Obviously, for the wideband algorithm the two outputs converge to about 15dB separation gain, whereas, as to be expected, the narrowband algorithm converges to no separation at all because the internal permutation problem remains unsolved.

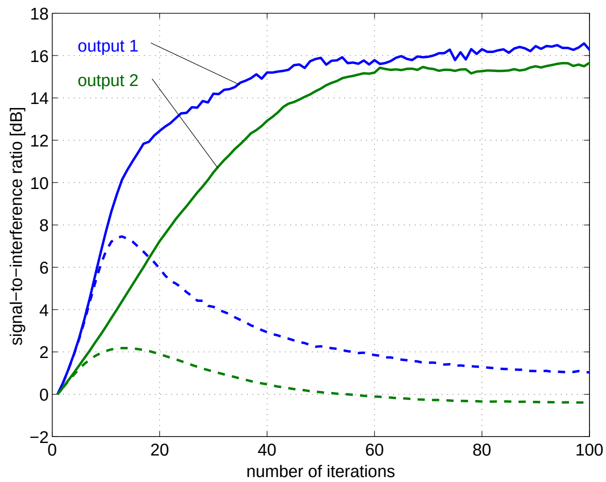

Figure 5: BSS performance of the wideband approach (solid) and the narrowband approach (dashed).

5 Summary and Conclusions

In this paper, we highlight the importance of strictly observing the wideband nature of the input signals when designing adaptive filtering algorithms if the narrowband signal model is not perfectly valid. Identifying the expansion of a Töplitz matrix to a circulant as a key for a rigorous derivation of a broad class of optimum wideband algorithms, we demonstrated the impact of the wideband signal model onto the resulting algorithm by way of a simple example. The importance of the constraints introduced by the wideband signal model becomes even more evident in the simulation results as presented for challenging adaptive filtering tasks such as multichannel acoustic echo cancellation and blind source separation for convolutive mixtures: Here we found recently published wideband algorithms to be far superior compared to corresponding narrowband algorithms.

Acknowledgements

The authors wish to thank Jacob Benesty of the Université du Québec, Montréal, Canada, for numerous stimulating discussions.

References

[1] A.V. Oppenheim, R.W. Schafer, *Digital Signal Processing,* Prentice Hall, Englewood Cliffs, NJ, 1975.

[2] S. Haykin, *Adaptive Filter Theory,* 3rd ed., Prentice Hall, Englewood Cliffs, NJ, 1996.

[3] G.H. Golub, C.F. Van Loan, *Matrix Computations*, 3rd ed., Johns Hopkins, Baltimore, MD, 1996.

[4] M. Dentino, J. McCool, and B. Widrow, "Adaptive filtering in the frequency domain," *Proc. IEEE,* vol. 66, pp.1658-1659, 1978.

[5] E.R. Ferrara, Jr., "Fast implementation of the LMS adaptive filter," *IEEE Trans. Acoust., Speech, Signal Processing (TR-ASSP),* vol. 28, pp. 474-475, 1980.

[6] D. Mansour, A.H. Gray, "Unconstrained Frequency-Domain Adaptive Filter," *IEEE TR-ASSP*, vol.30, no.5, 1982.

[7] J.C.Lee, C.K.Un, "Performance analysis of frequency-domain block LMS adaptive digital filters," *IEEE Trans. Circuits Syst.,* vol.36, pp. 173-189, 1989.

[8] J.-S. Soo, K.K. Pang, "Multidelay block frequency domain adaptive filter," *IEEE TR-ASSP*, vol.38, pp. 373-376, 1990.

[9] E. Moulines, O. Ait Amrane, and Y. Grenier, "The generalized multidelay adaptive filter: structure and convergence analysis," *IEEE Trans. Signal Processing,* vol. 43, pp. 14-28, 1995.

[10] J. Prado, E. Moulines, "Frequency-domain adaptive filtering with applications to acoustic echo cancellation," *Ann. Télécommun.,* vol. 49, pp. 414-428, 1994.

[11] J. Benesty, A. Gilloire, and Y. Grenier, "A frequency-domain stereophonic acoustic echo canceller exploiting the coherence between the channels," *J. Acoust. Soc. Am.,* vol. 106, pp. L30-L35, Sept. 1999.

[12] H. Buchner, J. Benesty, and W. Kellermann, "Multichannel Frequency-Domain Adaptive Algorithms with Application to Acoustic Echo Cancellation," in J.Benesty, Y.Huang (eds.), *Adaptive signal processing: Application to real-world problems*, Springer, Berlin, 2003.

[13] H. Buchner, J. Benesty, T. Gänsler, and W. Kellermann, "An outlier-robust extended multidelay filter with application to acoustic echo cancellation," *Proc. Intl. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Kyoto, 2003.

[14] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, Wiley, New York, 2001.

[15] H. Sawada et al., "Spectral smoothing for frequency-domain blind source separation," *Proc. IWAENC*, Kyoto, 2003.

[16] H. Buchner, R. Aichner, and W. Kellermann, "A generalization of a class of blind source separation algorithms for convolutive mixtures," *Proc. IEEE Int. Symp. on Independent Component Analysis (ICA)*, Nara, Japan, 2003, pp. 945-950.

[17] H. Buchner, R. Aichner, and W. Kellermann, "Blind Source Separation for Convolutive Mixtures Exploiting Nongaussianity, Nonwhiteness, and Nonstationarity," *Proc. IWAENC*, Kyoto, 2003.

[18] H. Buchner, R. Aichner, and W. Kellermann, "Blind Source Separation for Convolutive Mixtures: A Unified Treatment," in Y.Huang and J. Benesty (eds.), *Audio Signal Processing*, Kluwer, Boston, to appear 2004.