COMPUTATIONALLY EFFICIENT FREQUENCY-DOMAIN ROBUST GENERALIZED SIDELOBE CANCELLER

W. Herbordt and W. Kellermann

Telecommunications Laboratory, University Erlangen-Nuremberg
Cauerstr. 7, 91058 Erlangen, Germany
{herbordt,wk}@LNT.de

ABSTRACT

This paper deals with hands-free acoustical front-ends for human/machine interaction in competing talker situations using an efficient implementation of a robust Generalized Sidelobe Canceller (GSC). A new frequency-domain scheme is devised which leads to a factor of five of computational savings and which improves tracking capability during double-talk. Robustness constraints are included using optimization with penalty functions.

1. INTRODUCTION

For hands-free acoustical human/machine interfaces, microphone arrays in conjunction with robust adaptive beamforming techniques allow extraction of desired signals from many kind of interferers, e.g. background noise, reverberation, or competing talkers [1]. Here, the time-domain GSC after [5] provides both high rejection of local interferers with sufficient quality of the signal of interest.

In practical applications, microphone arrays are often combined with acoustic echo cancellation [1, 3] or single-channel noise reduction [1], which are often realized advantageously in the frequency domain.

In [4], we have devised a computationally efficient frequency-domain combination of AECs and the GSC. In this paper, we revisit the frequency-domain GSC for adding robustness constraints and for discussing adaptation control aspects. In contrast to [5], we derive the robustness constraints by directly introducing penalty terms into the optimization criterion and do not introduce them as an add-on. We further show that considerable improvement of interference rejection is possible in double-talk situations.

In Section 2, we describe the block convolution-based GSC. Then, we compare the computational complexity of this arrangement to that of the conventional time-domain implementation in Section 3. The performance is illutrated by simulations in Section 4.

2. FREQUENCY-DOMAIN GSC

In this section, we transform the robust time-domain GSC (TGSC, see Fig. 1) into the frequency domain. For fast linear convolution, we use the overlap-save (OLS) method for partitioning and reassembling the data. Although constrained and unconstrained frequency-domain adaptive filters [6, 8] are considered in later sections, we only describe the constrained algorithm (FGSC) here. The unconstrained realization (UFGSC) can be obtained by simply omitting the constraints in the filter update equations.

In the following, uppercase symbols denote variables in the frequency domain, lowercase symbols stand for time-domain variables, and the boldface font indicates a vector or matrix quantity. Superscripts T and H represent transpose and complex conjugate transpose, respectively. The number of microphones is denoted by M, the DFT length is $2L$. $\mathbf{F}$ is the $2L \times 2L$ discrete Fourier transform (DFT) matrix. The discrete time variable is n. We further use the time index $k = n/L$ that reflects the discrete time in numbers of blocks. $\mathbf{0}$ and $\mathbf{1}$ are length L-row vectors with zeroes and ones, respectively.

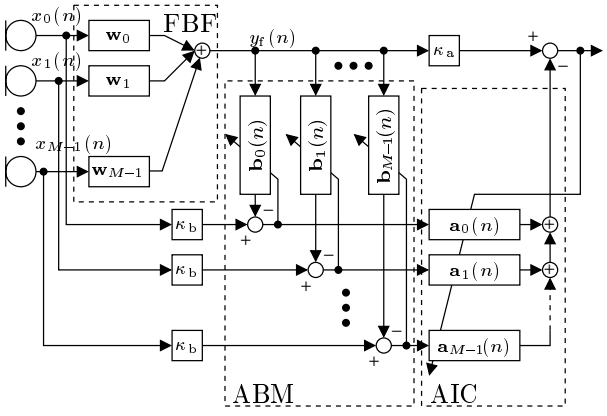

Figure 1: *Robust time-domain GSC after [5]*

In Section 2.1, we start the discussion with the Fixed Beamformer (FBF). It follows the transformation of the adaptive sidelobe cancelling path, which consists of the Adaptive Blocking Matrix (ABM) (Section 2.2) and of the Adaptive Interference Canceller (AIC) (Section 2.3). Adaptation control aspects are addressed in Section 2.4.

2.1. Fixed Beamformer (FBF)

The FBF enhances target signal components, and is used as reference for the adaptation of the adaptive sidelobe cancelling path.

The fractional time-delays that are required in the discrete time domain for the steering into the assumed tar-

This work was supported by a grant from Intel Corp. Hillsboro, OR

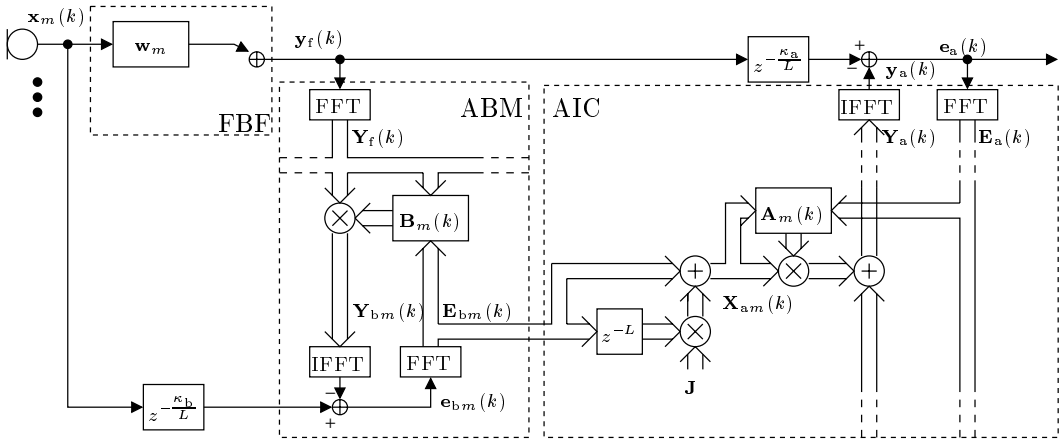

Figure 2: *Frequency-domain GSC*

get direction-of-arrival (DOA) are usually realized by short fractional delay filters. They are efficiently realized in the time domain.

Usage of a simple delay&sum beamformer with a broadside steered microphone array leads to target signal cancellation at high frequencies, if the steering direction and the actual target position do not coincide, since the (parabolic) mainlobe width decreases hyperbolically with increasing frequency.

In order to prevent target signal cancellation at high frequencies, we use a fixed filter&sum beamformer with filter weight vectors $\mathbf{w}_m$, with $m = 0,\ 1,\ \ldots,\ M-1$.

At low frequencies, where the mainlobe is sufficiently large, simple delay&sum beamforming is used. At high frequencies, the filter coefficients are obtained by a Dolph-Chebychev design [2], which minimizes the sidelobe level for a predefined first null of the array pattern (relative to the array steering direction) at a given frequency. Choosing a frequency-independent first null such that the mainlobe is sufficiently wide and applying the Dolph-Chebyshev design for each frequency bin yields an 'optimum' magnitude of the FBF filter transfer functions. These magnitude responses are then fed into a filter design algorithm for calculating the time-domain tap weight vectors $\mathbf{w}_m$.

Since low-order filters are generally sufficient, the FBF can be realized efficiently in the time domain: $y_\mathrm{f}(n) = \sum_{m=0}^{M-1} \mathbf{w}_m^T \mathbf{x}_m(n)$, where $\mathbf{x}_m(n)$ is a column vector holding the latest samples of the m-th microphone signal $x_m(n)$.

2.2. Adaptive Blocking Matrix (ABM)

The ABM consists of coefficient constrained adaptive filters between the FBF output and the sensor channels: The signal of interest is adaptively subtracted from the sidelobe cancelling path in order to prevent target signal cancellation by the AIC. The coefficient constraints assure that only signals from a predefined target tracking region are cancelled by the ABM [5].

First, we introduce some notations. Then, the robustness constraints are described. Finally, the constrained filter update equations are given.

Notations. In Fig. 2, only the reference path with one signal path of the FBF and with one signal path of the adaptive sidelobe cancelling is depicted for simplicity.

The time delay κ_b ensures causality of the adaptive filters. The DFT of the FBF output $\mathbf{Y}_\mathrm{f}(k)$ can be written as a diagonal matrix:

$$\mathbf{Y}_\mathrm{f}(k) = \mathrm{diag}\Big\{\mathbf{F} \begin{pmatrix} y_\mathrm{f}(kL-L) \\ \vdots \\ y_\mathrm{f}(kL) \\ \vdots \\ y_\mathrm{f}(kL+L-1) \end{pmatrix} \Big\}. \qquad (1)$$

The adaptive filter transfer functions $\mathbf{B}_m(k)$, with $m = 0,\ 1,\ \ldots,\ M-1$ are calculated from the time-domain L-by-1 weight vectors $\mathbf{b}_m(k)$ by

$$\mathbf{B}_m(k) = \mathbf{F}\left(\mathbf{b}_m^T(k),\ \mathbf{0}\right)^T. \qquad (2)$$

The FBF output $\mathbf{Y}_\mathrm{f}(k)$ is filtered by the adaptive filters $\mathbf{B}_m(k)$ yielding

$$\mathbf{Y}_{\mathrm{b}m}(k) = \mathbf{Y}_\mathrm{f}(k)\mathbf{B}_m(k)\,. \qquad (3)$$

For the adaptation algorithm, error signals $\mathbf{E}_{\mathrm{b}m}(k)$ are required which are free of circular convolution effects. Or, the time-domain signals $\mathbf{e}_{\mathrm{b}m}(k)$ have to be constrained in such a way that the first block of L samples is discarded and that the second block of L samples is saved. That is,

$$\mathbf{e}_{\mathrm{b}m}(k) = \mathbf{x}_m(k - \frac{\kappa_\mathrm{b}}{L}) - \mathbf{v}\,\mathbf{F}^{-1}\mathbf{Y}_\mathrm{f}(k)\mathbf{B}_m(k)\,, \qquad (4)$$

where $\mathbf{v} = \mathrm{diag}\{(\mathbf{0},\ \mathbf{1})\}$ and where the vector $\mathbf{x}_m(k)$ is defined as

$$\mathbf{x}_m(k) = \big(\mathbf{0},\ x_m(kL),\ \ldots,\ x_m(kL+L-1)\big)^T. \qquad (5)$$

Robustness Constraints. The filter update equations are obtained by minimization of the mean square of ABM output signals $J = \mathcal{E}\{\mathbf{e}_{\mathrm{b}m}^T(k)\mathbf{e}_{\mathrm{b}m}(k)\}$.

The constraints are introduced into the objective function by an additional penalty term, which is zero if the constraints are maintained, but which is equal to the squared sum of the constraints weighted by σ, if the constraints are violated [7]:

$$J + \sigma \sum_{l=0}^{L-1} \lfloor c_{\mathrm{u}l}(\mathbf{b}_m)\rfloor^2 + \lfloor c_{\mathrm{o}l}(\mathbf{b}_m)\rfloor^2 \rightarrow \min, \qquad (6)$$

The operator $\lfloor\cdot\rfloor$ returns zero, if the argument is less than zero, otherwise it returns the value of the argument. The upper coefficient constraints $c_{\mathrm{u}l}(\mathbf{b}_m)$ and the lower coefficient constraints $c_{ol}(\mathbf{b}_m)$, with $l = 0, 1, \ldots, L-1$ are defined as proposed in [5]

$$c_{\mathrm{u}l}(\mathbf{b}_m) = b_m(l) - b_{\mathrm{u}m}(l) < 0\,, \quad (7)$$
$$c_{ol}(\mathbf{b}_m) = b_{om}(l) - b_m(l) < 0\,, \quad (8)$$

respectively. The tap weights $b_{x\,m}(l)$ with $x \in \{\mathrm{u}, \mathrm{o}\}$ define the upper and the lower coefficient bounds, respectively. $b_m(l)$ denote the l-th element of the vector $\mathbf{b}_m$.

Update Equation. Deriving now an LMS adaptation step with this constrained objective function and transforming these time-domain filter update equations into a block-based FDAF update equation, we obtain

$$\mathbf{B}_m(k+1) = \mathbf{B}_m(k) + \mathbf{G}\boldsymbol{\mu}(k)\mathbf{Y}_{\mathrm{f}}^H(k)\mathbf{E}_{\mathrm{b}m}(k) + \boldsymbol{\Xi}_m(k)\,. \quad (9)$$

The matrix $\mathbf{G} = \mathbf{F}\mathbf{g}\mathbf{F}^{-1}$ with $\mathbf{g} = \mathrm{diag}\{(\mathbf{1}, \mathbf{0})\}$ constrains the gradient and ensures linear convolution.

The matrix with normalized step sizes $\boldsymbol{\mu}(k)$ is defined as

$$\boldsymbol{\mu}(k) = 2\mu\mathrm{diag}\Big\{\big(S^{-1}_{Y_{\mathrm{f}}Y_{\mathrm{f}}}(k,0), \ldots, S^{-1}_{Y_{\mathrm{f}}Y_{\mathrm{f}}}(k,2L-1)\big)\Big\}\,, \quad (10)$$

with a fixed step size parameter μ and with the power estimate of the l-th frequency bin

$$S_{Y_{\mathrm{f}}Y_{\mathrm{f}}}(k,l) = \lambda S_{Y_{\mathrm{f}}Y_{\mathrm{f}}}(k-1,l) + (1-\lambda)|Y_{\mathrm{f}}(k,l)|^2\,, \quad (11)$$

with $l = 0, 1, \ldots, 2L-1$. $Y_{\mathrm{f}}(k,l)$ denotes the l-th frequency bin of $\mathbf{Y}_{\mathrm{f}}(k)$.

The constraint vector $\boldsymbol{\Xi}_m(k)$ reads:

$$\boldsymbol{\Xi}_m(k) = \mathbf{F}\big(\boldsymbol{\xi}_m^T(k),\, \mathbf{0}\big)^T\,, \quad (12)$$

where the vector $\boldsymbol{\xi}_m(k)$ is defined as

$$\boldsymbol{\xi}_m(k) = 2\mu\sigma\sum_{l=0}^{L-1}\boldsymbol{\epsilon}_l\big(\lfloor c_{\mathrm{u}l}(\mathbf{b}_m(k))\rfloor - \lfloor c_{ol}(\mathbf{b}_m(k))\rfloor\big)\,, \quad (13)$$

and where the vector $\boldsymbol{\epsilon}_l$ is a L-by-1 zero vector and the l-th element equals one.

We note that the coefficient constraints are realized in the time domain, since they cannot be transformed efficiently into the frequency domain.

2.3. Adaptive Interference Canceller (AIC)

The AIC adaptively subtracts all signal components from the reference path which are correlated between the AIC filter inputs and the FBF output. A norm constraint that limits the norm of the adaptive filters is applied in order to improve robustness against target signal cancellation [5].

The time delay κ_{a} is introduced for causality reasons. For the frequency-domain adaptive filter inputs $\mathbf{X}_{\mathrm{a}m}(k)$, both the actual block of input samples and the previous block of input samples are required (see Eq. 1). However, we observe that the frequency-domain ABM output signals $\mathbf{E}_{\mathrm{b}m}(k)$ are obtained by the DFT of a block of zeroes in front of a block of time-domain error signals (see Eqs. 4,5). The signals $\mathbf{E}_{\mathrm{b}m}(k)$ can still be used as AIC inputs, and computational complexity can be saved when these signals are put into the required format directly in the frequency domain. This is accomplished by the following operation:

$$\mathbf{X}_{\mathrm{a}m}(k) = \mathrm{diag}\big\{\mathbf{E}_{\mathrm{b}m}(k) + \mathbf{J}\,\mathbf{E}_{\mathrm{b}m}(k-1)\big\}\,, \quad (14)$$

with $\mathbf{J} = \mathrm{diag}\{(1, -1, 1, \ldots, -1)_{1\times 2L}\}$, which realizes a circular shift of L samples in the frequency domain. This operation does not require any multiplications.

The following steps are straightforward: The AIC error signal $\mathbf{e}_{\mathrm{a}}(k)$ can be written as

$$\mathbf{e}_{\mathrm{a}}(k) = \mathbf{y}_{\mathrm{f}}(k) - \mathbf{y}_{\mathrm{a}}(k)\,, \quad (15)$$

where the vector $\mathbf{y}_{\mathrm{f}}(k)$ is now defined as

$$\mathbf{y}_{\mathrm{f}}(k) = \big(\mathbf{0},\, y_{\mathrm{f}}(kL-\kappa_{\mathrm{a}}), \ldots, y_{\mathrm{f}}(kL+L-1-\kappa_{\mathrm{a}})\big)^T\,, \quad (16)$$

and where $\mathbf{y}_{\mathrm{a}}(k)$ is given by

$$\mathbf{y}_{\mathrm{a}}(k) = \mathbf{w}\,\mathbf{F}^{-1}\sum_{m=0}^{M-1}\mathbf{X}_{\mathrm{a}m}(k)\mathbf{A}_m(k)\,. \quad (17)$$

The adaptive filters $\mathbf{A}_m(k)$ are defined as in Eq. 2.

The norm constrained filter update equations of the adaptive filters are given by

$$\mathbf{A}_m(k+1) = \mathbf{A}_m(k) + \mathbf{G}\big(\boldsymbol{\beta}(k)\mathbf{X}_{\mathrm{a}m}^H(k)\mathbf{E}_{\mathrm{a}}(k) + \boldsymbol{\Psi}_m(k)\big)\,, \quad (18)$$

where $\boldsymbol{\beta}(k)$ is defined according to Eq. 10 with $S_{Y_{\mathrm{f}}Y_{\mathrm{f}}}(k,l)$ replaced by $S_{X_{\mathrm{a}}X_{\mathrm{a}}}(k,l)$:

$$S_{X_{\mathrm{a}}X_{\mathrm{a}}}(k,l) = \lambda S_{X_{\mathrm{a}}X_{\mathrm{a}}}(k-1,l) + (1-\lambda)\sum_{m=0}^{M-1}|X_{\mathrm{a}m}(k,l)|^2\,. \quad (19)$$

$X_{\mathrm{a}m}(k,l)$ denotes the l-th frequency bin of $\mathbf{X}_{\mathrm{a}m}(k)$. The norm constraint is added by the vector $\boldsymbol{\Psi}_m(k)$. The constraint equation is given by

$$c\big(\mathbf{A}_m\big) = \sum_{n=0}^{M-1}\mathbf{A}_n^H\mathbf{A}_n - \Omega < 0\,, \quad (20)$$

where the parameter Ω is the maximum allowable filter norm. Following above procedure, we obtain for the constraint matrix $\boldsymbol{\Psi}_m(k)$:

$$\boldsymbol{\Psi}_m(k) = 4\mu\sigma\lfloor c\big(\mathbf{A}_m(k)\big)\rfloor\,\mathbf{A}_m(k)\,. \quad (21)$$

Finally, the GSC output is obtained by saving the last L samples of $\mathbf{e}_{\mathrm{a}}(k)$.

2.4. Adaptation Control

The adaptation control should ensure that the ABM (AIC) is only adapted when the SNR is high (low). Otherwise, the adaptation should be stalled. Due to the non-stationarity of the power spectral density (PSD) of speech signals, a decision based on fullband SNR estimation [5] may fail and may lead to significant target signal cancellation. SNR estimation and ABM/AIC adaptation control in frequency bins improves robustness and target signal quality.

3. COMPUTATIONAL COMPLEXITY

We compare the total number of real multiplications (NRM) per output sample and the memory requirements of FGSC and UFGSC with TGSC.

For this comparison, we disregard the adaptation control. Assuming that the DFTs are carried out by a radix-2 algorithm, we obtain for the NRM per output sample of the UFGSC $NRM_{\mathrm{u}} = \frac{1}{L}\big((4M+3)\,2L\log_2 2L + 47ML + 14L - 46M\big)$. The FGSC requires 2 additional DFTs per update equation, or $NRM_{\mathrm{c}} = NRM_{\mathrm{u}} + 8M\log_2 2L$. The TGSC using NLMS algorithms requires $NRM_{\mathrm{t}} = 13ML + 3M$ real multiplications.

The complexity ratios, $CR_x = NRM_x/NRM_{\mathrm{t}}$, with $x \in \{\mathrm{c}, \mathrm{u}\}$ are depicted for $L = 32, 64$ over the number of microphones in Fig. 3(a). An estimate of the memory requirement ratio (MR) between the FGSC, UFGSC and the TGSC is illustrated in Fig. 3(b).

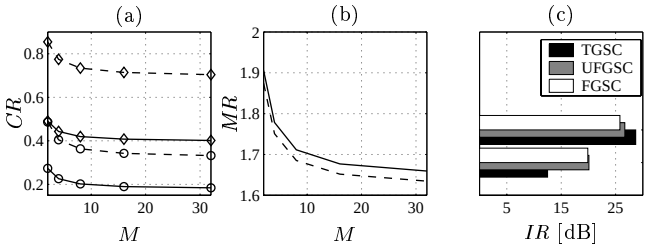

Figure 3: *(a) Complexity ratios CR_{u} ($\circ$), CR_{c} ($\diamond$) and (b) Memory requirement ratio (MR) for $L = 32$ (dashed) and for $L = 64$ (solid); (c) average IR*

The block processing reduces the computational requirements to $21\% - 41\%$ ($L = 64, M = 8$) of the nonblock time-domain processing depending on whether the unconstrained or the constrained algorithm is used. We also see that the computational complexity is reduced even with very short adaptive filters.

4. EXPERIMENTAL EVALUATION

Here, the interference rejection (IR), or the interferer-energy ratio between the microphones and the beamformer output is illustrated for double-talk and for interference only.

A linear array with 8 equally spaced ($d = 4$ cm) sensors is placed in a chamber with low reverberation ($T_{60} = 50$ ms). The female target speaker and the male interferer speaker are located broadside at a distance of 60 cm and 30 degrees off-axis at a distance of 1.3 m, respectively. The frequency band is $360 - 6000$ Hz at a sampling rate $T_s = 12$ kHz. The FBF is designed according to Section 2.1. The transition frequency f_0 between delay&sum beamformer and Dolph-Chebyshev based beamformer is $f_0 = 1700$ Hz. The mainlobe width is 45 degrees (frequency $f > f_0$). The GSC parameters are chosen such that maximum IR is obtained (FGSC, UFGSC: $L = 64$, $\mu = 1.5$, $\lambda = 0.4$ for interference-only and $\lambda = 0.95$ for double-talk; TGSC: $L = 64$, NLMS algorithm step size $\mu = 0.7$). For the interference-only case, the ABM filters are frozen after adaptation for 50000 samples with only the target present. The AIC is then adapted for 100000 samples. The IR is calculated over the last 20000 samples. For the double-talk case, ABM and AIC are adapted simultaneously according to Section 2.4 with both target signal and interference for 100000 samples. The average IR is again measured over the last 20000 samples.

For interference only, the average IR is almost identical for all GSC realizations (see Fig. 3(c)). The AIC tracks the time-variance of the interferer PSD and thus cancels the non-stationary interference efficiently. Since the block sizes of the frequency-domain arrangement are kept small relative to the short-time stationarity of speech signals ($L = 64$, corresponding to 5.3 ms with $T_s = 12$ kHz), the reduced tracking capability of block adaptive algorithms does little influence the average IR. Due to faster convergence, the IR of the UFGSC is slightly higher than the IR of the FGSC.

For double-talk, the IR is considerably improved with FGSC/UFGSC relative to TGSC due to better tracking capability of the bin-wise adaptation control.

5. CONCLUSION

In this paper, we have formulated a computationally efficient frequency-domain robust GSC. ABM constraints and AIC constraints are realized efficiently by optimization with penalty function techniques. We have further illustrated that controlling the adaptation in discrete frequency bins improves the tracking behavior and the average interference rejection considerably. Finally, we have introduced a Dolph-Chebyshef fixed beamformer realization in order to prevent target signal cancellation at high frequencies.

6. REFERENCES

[1] M. S. Brandstein, D. B. Ward, editors. *Microphone Arrays: Signal Processing Techniques and Applications*. Springer Verlag, 2001.

[2] C. L. Dolph. A current distribution for broadside arrays which optimize the relationship between beam width and side-lobe level. *Proceedings of the I.R.E.*, 34(6):335–348, June 1946.

[3] W. Herbordt, W. Kellermann. Limits for generalized sidelobe cancellers with embedded acoustic echo cancellation. *Proc. ICASSP*, Vol.5, Mai 2001.

[4] W. Herbordt and W. Kellermann. Comutationally efficient frequency-domain combination of acoustic echo cancellation and robust adaptive beamforming. *Proc. EUROSPEECH*, September 2001.

[5] O. Hoshuyama, A. Sugiyama, A. Hirano. A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters. *IEEE Trans. on SP*, 47(10):2677–2684, 1999.

[6] D. Mansour, A. H. Gray. Unconstrained frequency-domain adaptive filter. *IEEE Trans. on ASSP*, ASSP-30(5):726–734, 1982.

[7] B. Rafaely, S. J. Elliot. A computationally efficient frequency-domain LMS algorithm with constraints on the adaptive filter. *IEEE Trans. on Signal Processing*, 48(6):1649–1655, June 2000.

[8] J. Soo, K. K. Pang. Multidelay block frequency domain adaptive filter. *IEEE Trans. on ASSP*, 38(2):373–376, February 1990.