A Robust Step-Size Control Algorithm for Frequency Domain Acoustic Echo Cancellation

Chao Wu, Kaiyu Jiang, Yanmeng Guo, Qiang Fu, Yonghong Yan

Institute of Acoustics, Chinese Academy of Sciences, Beijing, China

wuchao@hccl.ioa.ac.cn

Abstract

The presence of near-end interferences and echo path changes make it essential for an adaptive filter to vary its learning rate according to corresponding conditions. In this paper, a robust step-size control algorithm which is based on the optimization of the square of the bin-wise *a posteriori error* is proposed. To prevent the adaptive filter from diverging in the presence of interferences, constraints on the filter update are applied. The learning rate expression is derived and then we extend the method to multidelay block frequency domain adaptive filter (MDF) so as to meet the demand of low delay in practical application. An updating strategy for the constraints is proposed as well. Experiments are carried out to demonstrate the superiority of the proposed approach, especially in double-talk and echo path change situations.

Index Terms: Acoustic echo cancellation, step-size control, robust filtering.

1. Introduction

Acoustic echo cancellation (AEC), which plays an important role in diverse fields including communication and speech recognition, is one of the most popular applications of adaptive filtering. Robust echo cancellation requires an adaptive filter to update slowly during the presence of interference (mainly double-talk situation) while converge fast in other cases.

Many different approaches have been proposed in the literatures to deal with this problem. Summarizing their results, one can classify these methods into two categories: Double-talk detection based methods and variable step-size (VSS) based methods. The first category attempts to detect double-talk conditions and then freezes the adaptation of the adaptive filter [1][2][3]. The second category concentrates on VSS control technique to achieve both robustness to double-talk and favorable convergence speed. Due to the intrinsic time lag of double-talk detection, VSS control methods are more popular in practical application. VSS control methods are mostly derived from the normalized least-mean-square (NLMS) algorithm [4] on account of its simplicity and stability. Jean-Marc Valin [5] and Yin Zhou [6] both derive the optimal learning rate aiming at minimizing the system misalignment changes. Unfortunately, the optimal step-size relates to the unknown residual echo. Benesty et al. [7] propose a nonparametric variable step-size (NPVSS) NLMS algorithm, assuming that the variance of near-end interference is known. With an estimation of near-end interference, Paleologu [8] and Kun Shi [9] respectively propose the time-domain and frequency-domain VSS NLMS algorithm, whose performance are both influenced by the accuracy of the estimation.

In [10], Vega et al. develop a robust VSS NLMS method based on the optimization of the square of the *a posteriori error*. Although robust to near-end perturbations, it has to be combined with a double-talk detector (DTD) to achieve favorable tracking ability. Inspired by this thought, we present a frequency domain step-size control method which based on the optimization of the square of the bin-wise *a posteriori error*. Different from [10], bin-wise constraint on the filter update is applied in our algorithm. Moreover, by updating the constraints appropriately, the proposed method obtains favorable convergence rate.

This paper is organized as follows. In Section 2, we review the robust algorithm proposed by Vega et al. and derive the step-size for the proposed algorithm. Simulated results are presented in Section 3. Section 4 concludes this paper.

2. Proposed algorithm

In this section, we first review a robust time-domain VSS control method proposed by Vega et al. Then the expression of proposed frequency domain variable step size(FDVSS) control method is derived.

2.1. Robust VSS NLMS [10]

As in [10], let $\boldsymbol{h_i} = (h_{i,0}, h_{i,1}, \ldots, h_{i,N-1})^T$ be an unknown $N \times 1$ linear finite-impulse response system and $\boldsymbol{w_i} = (w_{i,0}, w_{i,1}, \ldots, w_{i,N-1})^T$ as its estimation, the $N \times 1$ vector $\boldsymbol{x_i} = [x(i), x(i-1), \ldots, x(i-N+1)]^T$ be the input at time i. Let d_i, e_i, v_i be the microphone input, output error and interference signal. Define the misalignment vector $\tilde{\boldsymbol{w}}_{\boldsymbol{i}} = \boldsymbol{h_i} - \boldsymbol{w_i}$ and the *a posteriori error* signal $e_{p,i} = \boldsymbol{x_i}^T \tilde{\boldsymbol{w}}_{\boldsymbol{i}} + v_i$.

Vega et al. propose to find the updating strategy as

$$\boldsymbol{w_i} = \underset{\boldsymbol{w_i} \in R^N}{arg\ min}\ e_{p,i}^2 \quad s.t. \quad ||\boldsymbol{w_i} - \boldsymbol{w_{i-1}}||^2 \leq \delta_i \tag{1}$$

where δ_i is a positive sequence. This algorithm can be put in a compact way such as

$$\boldsymbol{w_i} = \boldsymbol{w_{i-1}} + min\left[\frac{|e_i|}{||\boldsymbol{x_i}||}, \sqrt{\delta_{i-1}}\right] sign(e_i) \frac{\boldsymbol{x_i}}{||\boldsymbol{x_i}||} \tag{2}$$

Here δ_i is adapted as follows:

$$\delta_i = \alpha\delta_{i-1} + (1-\alpha) min\left[e_i^2/||\boldsymbol{x_i}||^2, \delta_{i-1}\right] \tag{3}$$

where $0 < \alpha < 1$ is a memory factor. In order to improve the tracking ability, a nonstationary control method combined with a DTD is used in Vega's method to distinguish double-talk from echo path changes, which is of great difficulty.

2.2. Robust FDVSS control algorithm

2.2.1. The FD model

Following the definition in Section 2.1, the FD input signal matrix and the FD weight vector can be defined as,

$$\boldsymbol{X}(k) = diag\{\boldsymbol{F}[x(kN-N),\ldots,x(kN+N-1)]^T\} \quad (4)$$

$$\boldsymbol{W}(k) = \boldsymbol{F}\boldsymbol{A}^T\boldsymbol{w_k} \quad (5)$$

where k is the frame index, $diag\{\cdot\}$ denotes a diagonal matrix, $\boldsymbol{F}$ is the $2N \times 2N$ DFT matrix, $A = [\boldsymbol{I_N}\ \boldsymbol{0_N}]$ with $\boldsymbol{I_N}$ being the $N \times N$ identity matrix and $\boldsymbol{0_N}$ being the $N \times N$ matrix whose all elements are zero. Define the microphone input signal vector $\boldsymbol{d_k} = [d(kN),\ldots,d(kN+N-1)]^T$ and the error signal vector $\boldsymbol{e_k} = [e(kN),\ldots,e(kN+N-1)]^T$, the FD expression can be put as

$$\boldsymbol{D}(k) = \boldsymbol{F}\boldsymbol{K}\boldsymbol{d_k} \quad (6)$$

$$\boldsymbol{E}(k) = \boldsymbol{D}(k) - \boldsymbol{F}\boldsymbol{Q}\boldsymbol{F^{-1}}\boldsymbol{X}(k)\boldsymbol{W}(k) \quad (7)$$

where $\boldsymbol{K} = [\boldsymbol{0_N}\ \boldsymbol{I_N}]$ and $\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{0_N} & \boldsymbol{0_N} \\ \boldsymbol{0_N} & \boldsymbol{I_N} \end{bmatrix}$. $\boldsymbol{E}(k)$ is the FD *a priori error*. Here, we define the FD *a posteriori error* as

$$\boldsymbol{E_P}(k) = \boldsymbol{D}(k) - \boldsymbol{F}\boldsymbol{Q}\boldsymbol{F^{-1}}\boldsymbol{X}(k)\boldsymbol{W}(k+1) \quad (8)$$

By using the approximation in [11]

$$\boldsymbol{F}\boldsymbol{Q}\boldsymbol{F^{-1}} \approx 1/2\boldsymbol{I_{2N}} \quad (9)$$

the FD representation of the *a priori error* and the *a posteriori error* are given as

$$\boldsymbol{E}(k) = \boldsymbol{D}(k) - 1/2\boldsymbol{X}(k)\boldsymbol{W}(k) \quad (10)$$

$$\boldsymbol{E_P}(k) = \boldsymbol{D}(k) - 1/2\boldsymbol{X}(k)\boldsymbol{W}(k+1) \quad (11)$$

2.2.2. Derivation of the robust FDVSS

Let $E_p(k,m), W(k,m), X(k,m)$ be the *mth* frequency bin of $\boldsymbol{E_P}(k), \boldsymbol{W}(k), \boldsymbol{X}(k)$, we propose to find the updating strategy in such a way that

$$W(k,m) = \underset{W(k,m)\in Z}{arg\ min}\ |E_p(k,m)|^2$$
$$s.t.\ |W(k+1,m) - W(k,m)| \le \delta_{k,m}|X(k,m)| \quad (12)$$

With $\delta_{k,m}|X(k,m)|$ in equation (12) being replaced by δ_0, the strategy turns to be the frequency domain implementation of Vega's method in Section 2.1. However, we can utilize more information of major frequency bins of the reference signal for the adaptation with $|X(k,m)|$ being used as a bin-wise weight of the constraints, by which means the convergence rate can be improved.

Mathematically, to perform the optimization, we can divide the problem into two cases:(a) the solution is one of the stationary points of the objective function.(b) the solution is one of the boundary points of the constraint.

For simplicity, we omit the frequency bin label *m* in the following discussion of this part. Combining equation (10) and (11) in the aspect of frequency bin, we have

$$E_p(k) = E(k) + 1/2X(k)[W(k) - W(k+1)] \quad (13)$$

In the first case, the derivation of $|E_p(k)|^2$ should be zero. By assuming $W(k+1) = A(k) + jB(k)$, we perform the derivation in a way as Haykin do in [12],

$$\frac{\partial|E_p(k)|^2}{\partial W(k+1)} = \frac{\partial|E_p(k)|^2}{\partial A(k)} + j\frac{\partial|E_p(k)|^2}{\partial B(k)} = 0 \quad (14)$$

substituting $|E_p(k)|^2 = E_p(k)E_p^*(k)$ into (14) results in

$$E_p(k)X^*(k) = 0 \quad (15)$$

where $*$ denotes the conjugate form. By combining (11) with (15), we have

$$W(k+1) = W(k) + \frac{2E(k)X^*(k)}{|X(k)|^2} \quad (16)$$

Noticing that the variation of $W(k)$ should meet the constraint (12), the following inequality must be satisfied.

$$\frac{|E(k)|}{|X(k)|^2} \le \frac{\delta_k}{2} \quad (17)$$

Substituting (16) into (13), we get $E_p(k) = 0$.

In the second case, inequality (17) will not be satisfied. The objective function reaches its minimum on the boundary where

$$|W(k+1) - W(k)| = \delta_k|X(k)| \quad (18)$$

According to equation (13), the unfolded expression of FD *a posteriori error* can be given as

$$|E_p(k)|^2 = |E(k)|^2 + (1/4)|X(k)|^2|W(k) - W(k+1)|^2 + Re\{E(k)X^*(k)[W(k) - W(k+1)]^*\} \quad (19)$$

where $Re\{\cdot\}$ indicates the real part of a complex number. Apparently, minimizing $|E_p(k)|^2$ requires to find the minimum of the third item because the former two items are both constants. Going a step further, by representing $E(k)X^*(k)$ and $[W(k) - W(k+1)]^*$ as two vectors $\vec{A}(k) = C(k) + jD(k)$ and $\vec{B}(k) = E(k) + jF(k)$ in the 2-D complex plane we get

$$Re\{\vec{A}(k)\vec{B}(k)\} =< \vec{A}(k), \vec{B}^*(k) > \quad (20)$$

where $< \cdot >$ refers to the vectors inner product. Obviously, this item will obtain its minimum when $\vec{B}^*(k) = -a\vec{A}(k)$. So equation (21) has to be satisfied to minimize $|E_p(k)|^2$,

$$W(k) - W(k+1) = -aE(k)X^*(k) \quad (21)$$

Combining (18) with (21), we have

$$W(k+1) = W(k) + \frac{\delta_k E(x)X^*(k)}{|E(k)|} \quad (22)$$

For all above, the representation of our algorithm is given as

$$W(k+1) = W(k) + min\{\frac{\delta_k}{2}, \frac{|E(k)|}{|X(k)|^2}\}\frac{2E(k)X^*(k)}{|E(k)|} \quad (23)$$

2.2.3. Extending to the MDF

FD adaptive filter is a suitable option for AEC due to its low computational complexity and decorrelating property. However, the long delay it introduces because of the implementation of block processing would be intolerant in many practical applications. In this section, we will extend the derived step-size control expression to MDF to tackle this problem. Soo [13] propose the MDF method and show its priority over the conventional FD filters. Partition the input signal $x(n)$ and the adaptive filter $w(n)$ into L blocks of length M with $N = LM$, we have

$$\boldsymbol{x}_{k,l} = [x(kM - lM - M + 1),\ldots,x(kM - lM + M)]^T \quad (24)$$

$$\boldsymbol{w}_{k,l} = [w_{k,(l-1)M+1},\ldots,w_{k,lM}]^T \quad (25)$$

for $l = 1, 2, \ldots, L$. Define $\boldsymbol{X}_l(k)$ and $\boldsymbol{W}_l(k)$ as

$$\begin{aligned} \boldsymbol{W}_l(k) &= \boldsymbol{F}_M \boldsymbol{A}_M^T \boldsymbol{w}_{k,l} && (26) \\ \boldsymbol{X}_l(k) &= diag\{\boldsymbol{F}_M \boldsymbol{x}_{k,l}\} && (27) \end{aligned}$$

where $\boldsymbol{F}_M$ is the $2M \times 2M$ DFT matrix, $\boldsymbol{A}_M = [\boldsymbol{I}_M \ \boldsymbol{0}_M]$. Based on these partitioned FD model, we extend the adaptation expressed by (23) to MDF as Soo do in [13] and obtain

$$W_l(k+1, m) = W_l(k, m) + \mu_l(k, m) \frac{2E(k, m) X_l^*(k, m)}{|E(k, m)|} \quad (28)$$

where

$$\mu_l(k, m) = min(\frac{\delta_l(k, m)}{2}, \frac{|E(k, m)|}{\sum_{l=1}^{L} |X_l(k, m)|^2}) \quad (29)$$

2.2.4. Updating of $\delta_l(k, m)$

In principle, in order to achieve good initial speed of convergence, δ_i is desired to have large values at the beginning of the adaptation. However, when the algorithm is close to its steady state, smaller values of δ_i will lead to a lower final error. In this part, we will develop a method to update $\delta_l(k, m)$.

In [10], a way to update δ_i is expressed by equation (3). The major drawback of this method lies in its poor tracking ability because of the decreasing property of δ_i. In this paper, we present a new updating method which make use of the sparsity of acoustic echo path as proportionate NLMS [14]. Basically, $\delta_l(k, m)$ is designed to be a block-varying value. We have $\delta_l(k, m) = \delta_k c_l(k)$ where $c_l(k)$ is the block-varying coefficient. These coefficients are computed according to the previous filter taps. Specific equations are

$$\begin{aligned} S_l(k) &= \sum_{m=1}^{M} |W_l(k, m)| && (30) \\ l_\infty(k) &= max\ \{S_1(k), \ldots, S_L(k)\} && (31) \\ l_\infty^{'}(k) &= max\ \{\epsilon, l_\infty(k)\} && (32) \\ P_l(k) &= max\ \{\kappa l_\infty^{'}(k), S_l(k)\} && (33) \\ \bar{P}(k) &= mean\ \{P_1(k), \ldots, P_L(k)\} && (34) \\ c^{'}{}_l(k) &= min\ \{\rho, P_l(k)/\bar{P}(k)\} && (35) \\ c_l(k) &= \begin{cases} 1 & k\ is\ an\ odd\ number, \\ c^{'}{}_l(k) & k\ is\ an\ even\ number. \end{cases} && (36) \end{aligned}$$

where ϵ and κ are small positive numbers. Equation (36) is used to avoid the disadvantage that the remaining small coefficients adapt at a slow rate especially when the bigger taps have converged. In addition, the way to update δ_k used in [10] is introduced in this paper.

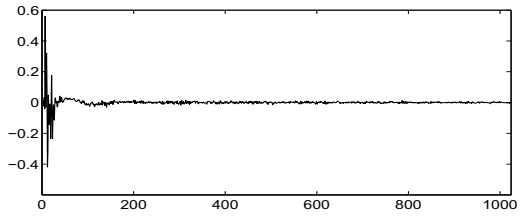
$$\begin{aligned} err(k) &= (\sum_{m=0}^{M_0} |E(k, m)|/|X(k, m)|^2)/M_0 && (37) \\ \delta_k &= \alpha \delta_{k-1} + (1-\alpha)\, min\ (\delta_{k-1}, err(k)) && (38) \\ \delta_k &= max(\delta_k, \delta_{min}) && (39) \end{aligned}$$

Finally, we look into a sliding-window with length L of value $err(k)$ considering the double-talk robustness. δ_k will be set to a smaller value $\delta_{min}^{'}$ than δ_{min} if K elements of L are bigger than δ_{thd}, otherwise, δ_k will roll back to δ_{min}.

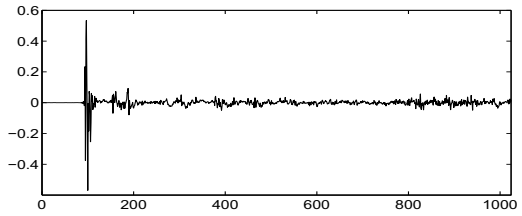
3. Simulation results

The systems used in the simulation are taken from Aachen Impulse Response (AIR) Database [15] and truncated to $N = 1024$. The length of adaptive filter is set to be N in all cases. We plot them in Fig.1. Echo Return Loss Enhancement (ERLE) and Misalignment are used as two different measures of performance for single-talk and double-talk respectively.

$$\begin{aligned} ERLE &= E\{d^2\}/E\{e^2\} && (40) \\ Misalignment &= 10\ log_{10}\{||\boldsymbol{h} - \boldsymbol{w}||^2/||\boldsymbol{h}||^2\} && (41) \end{aligned}$$



(a) Impulse response 1



(b) Impulse response 2

Figure 1: Impulse response(IR) from AIR database .

The behavior of the proposed method is compared with Vega's robust VSS NLMS [10] and the FD VSS by Kun Shi [9]. To keep the comparison fair, both algorithms are realized in FD using the multidelay block structure. As mentioned above, in order to achieve favorable tracking ability, Vega's VSS method requires a DTD. In our simulations, the performance of this algorithm is evaluated in two different ways. On one hand, we declared double-talk only when it really occurs so as to guarantee that it works with a perfect DTD. On the other hand, we set a minimum for δ_i and evaluate the method without DTD. Correspondingly, we mark these two implementations as Vega (DTD) and FD Vega in the following. Moreover, the performance of the proposed algotithm is evaluated with and without updating of $\delta_l(k, m)$, which are marked Proposed2 and Proposed1 accordingly. Parameters of the proposed algorithm are set experimentally as follows: $\delta_0 = 2 \times 10^{-4}, \delta_{min} = 3.0 \times 10^{-6}, \delta_{min}^{'} = 1.0 \times 10^{-6}, \delta_{thd} = 1.2 \times 10^{-5}, M = 64, \epsilon = 0.01, \kappa = 0.01, \rho = 4, \alpha = 0.995, M_0 = 16, K = 80, L = 150$.

3.1. Convergence Speed During Single-Talk

To begin with, we compare the performance of different algorithms without double-talk.

An independent zero-mean white Gaussian noise signal is added to the input first, then we evaluate the performance of the proposed method with speech input signal. Echo path change takes place after 5s from IR1 to IR2. The behaviors of the simulated methods are presented in Fig.2 and Fig.3 separately. As demonstrated by these two figures, the proposed method

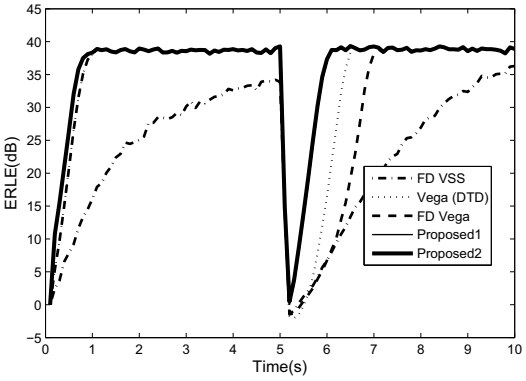


Figure 2: ERLE of Vega(DTD), Vega and Proposed method. The input signal is a white Gaussion noise. Echo path change takes place after 5s.

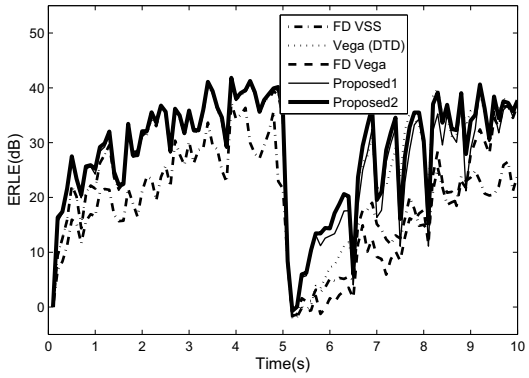


Figure 3: ERLE of Vega(DTD), Vega and Proposed method. The input signal is speech. Echo path change takes place after 5s.

has almost the same convergence rate with Vega's method and outperforms [9]. However, the tracking behavior of proposed method shows significant advantage over Vega's method, especially when it performs without a perfect DTD. It is worth mentioning that the better tracking ability of the proposed1 method over FD Vega confirms the superiority of the proposed strategy (equation (12)) because δ_k maintain constants after 5s for both algorithms.

3.2. Misalignment During Double-Talk

Another possible situation in AEC is the presence of the ambient noise. This part we examine the behavior of these methods with speech and ambient noise as near-end input.

It is obviously shown by Fig.4 that the lower misalignment than Vega's method during double-talk periods shows the good robustness of the proposed method, even compared with Vega's method with a perfect DTD. The same trend is observed with ambient noise being the near-end input in Fig.5. In addition, a lower misalignment achieved by Proposed2 method than Proposed1 demonstrate a faster convergence rate which is attributed to the proposed updating strategy of $\delta_l(k,m)$. It is worthy mentioning that the misalignment of Vega's method with DTD

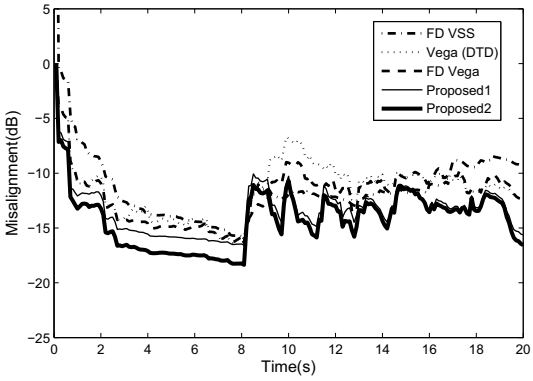


Figure 4: Misalignment of Vega(DTD), Vega and Proposed method. Far-end input signal is speech. Double-talk occurs after 8s. The near-end signal to far-end echo ratio is 0dB.

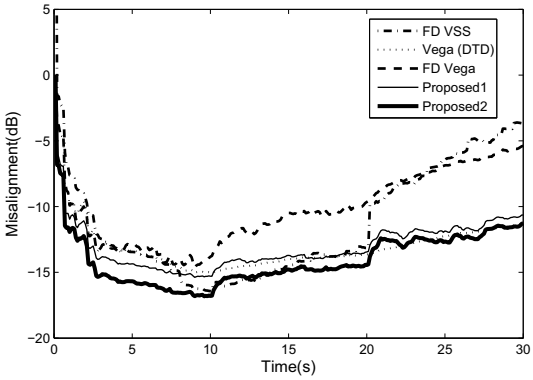


Figure 5: Misalignment of Vega(DTD), Vega and Proposed method. Far-end input signal is speech. The far-end echo to near-end noise ratio (ENR) decreases from 30dB to 20dB after 10s and then to 10dB after 20s.

increase in double-talk due to its nonstationary control method.

4. Conclusions

The presence of the near-end interference and echo path change are two factors mainly influence the performance of adaptive filter. In this paper, a robust frequency domain step-size control method based on the optimization of the square of the bin-wise *a posteriori error* is proposed. Time varying constraints on the filter update are applied. Simulation results confirm that the proposed algorithm shows up favorable initial convergence rate, good tracking ability and excellent robustness to double-talk.

5. Acknowledgements

This work has been supported by the National Natural Science Foundation of China (Nos. 10925419, 90920302, 61072124, 11074275, 11161140319, 91120001, 61271426), the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant Nos. XDA06030100, XDA06030500), the National 863 Program (No. 2012AA012503) and the CAS Priority Deployment Project.

6. References

[1] D.L. Duttweiler, "A twelve-channel digital echo canceler," *IEEE Transactions on Communications*, vol. 26, pp. 647–653, May 1978.

[2] J. Benesty, D.R. Morgan, and J.H. Cho, "A new class of doubletalk detectors based on cross-correlation," *IEEE Transactions on Speech and Audio Processing*, vol. 8, pp. 168–172, March 2000.

[3] A. Sugiyama, J. Berclaz, and M. Sato, "Noise-robust double-talk detection based on normalized cross correlation and a noise offset," in *International Conference on Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05).* IEEE, 2005, vol. III, pp. 153–156.

[4] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Englewood Cliffs, NJ: Prentice-Hall, 1985.

[5] J.M. Valin, "On adjusting the learning rate in frequency domain echo cancellation with double-talk," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, pp. 1030–1034, March 2007.

[6] Y. Zhou and X.D. Li, "A variable step-size for frequency-domain acoustic echo cancellation," IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2007, pp. 303–306.

[7] J. Benesty, H. Rey, L.R. Vega, and S. Tressens, "A nonparametric vss nlms algorithm," *IEEE Signal Processing Letters*, vol. 13, pp. 581–584, October 2006.

[8] C. Paleologu, S. Ciochina, and J. Benesty, "Variable step-size nlms algorithm for under-modeling acoustic echo cancellation," *IEEE Signal Processing Letters*, vol. 15, pp. 5–8, 2008.

[9] K. Shi and X.L. Ma, "A frequency domain step-size control method for lms algorithms," *IEEE Signal Processing Letters*, vol. 17, pp. 125–128, February 2010.

[10] L.R. Vega, H. Rey, J. Benesty, and S. Tressens, "A new robust variable step-size nlms algorithm," *IEEE Transactions on Signal Processing*, vol. 56, pp. 1878–1893, May 2008.

[11] Y. Huang and J. Benssty, *Audio Signal Processing for Next Generation Multimedia Communication Systems*, Norwell, MA: Kluwer, 2004.

[12] Simon S Haykin, *Adaptive filter theory*, Pearson Education India, 2005.

[13] J.S. Soo and K.K. Pang, "Multidelay block frequency domain adaptive filter," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, pp. 373–376, February 1990.

[14] Donald L Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 5, pp. 508–518, 2000.

[15] http://www.ind.rwth-aachen.de/de/forschung/tools-downloads/aachen-impulse-response-database/