A frequency-domain adaptive filtering prediction error method for[1] double-talk-robust acoustic echo cancellation[‖]

Jose M. Gil-Cacho[*], Toon van Waterschoot[‡], Marc Moonen[§], and Søren Holdt Jensen[¶]

[*]Corresponding author. KU Leuven Department of Electrical Engineering (ESAT) STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics/iMinds Future Health Department, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium, Tel. +32 16321856, Fax +32 16321970, E-mail pepe.gilcacho@esat.kuleuven.be

[‡]KU Leuven Department of Electrical Engineering (ESAT) STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics/iMinds Future Health Department, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium, Tel. +32 16321927, Fax +32 16321788, E-mail toon.vanwaterschoot@esat.kuleuven.be

[§]KU Leuven Department of Electrical Engineering (ESAT) STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics/iMinds Future Health Department, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium, Tel. +32 16321060, Fax +32 16 321970, E-mail marc.moonen@esat.kuleuven.be

[¶]Department of Electronic Systems Aalborg University, Fredrik Bajers Vej 7, DK-9220 Aalborg, Denmark, Tel. +45 9940 8654, E-mail shj@es.aau.dk

[‖]EDICS: AUD-ECHO

Abstract

In this paper, we propose a new framework to tackle the problem of double-talk (DT) in acoustic echo cancellation (AEC). It is based on a frequency domain adaptive filtering (FDAF) implementation of the so-called PEM-AFROW algorithm (FDAF-PEM-AFROW). We depart from this framework to show an improvement in performance with respect to the individual versions, i.e., PEM-AFROW and FDAF. This improvement is basically due to three reasons, (1) in case the assumed signal model or the unbiasedness constraints in the PEM-AFROW are violated we will show that a FDAF implementation ameliorates the negative effects associated with it, (2) the FDAF itself may be seen to minimize a BLUE criterion if a proper normalization factor is used during adaptation, and (3) the proposed instantaneous pseudo-correlation (IPC) measure between the near-end signal and the loudspeaker signal is significantly reduced

This research work was carried out at the ESAT Laboratory of KU Leuven, in the frame of KU Leuven Research Council CoE PFV/10/002 (OPTEC), Concerted Research Action GOA-MaNet, the Belgian Programme on Interuniversity Attraction Poles initiated by the Belgian Federal Science Policy Office IUAP P7/19 ‘Dynamical systems control and optimization’ (DYSCO) 2012-2017, Research Project FWO nr. G.0763.12 ’Wireless Acoustic Sensor Networks for Extended Auditory Communication’, FP7-PEOPLE Marie Curie Initial Training Network ’Dereverberation and Reverberation of Audio, Music, and Speech (DREAMS)’, funded by the European Commission under Grant Agreement no. 316969, EC-FP6 project ’Core Signal Processing Training Program’ (SIGNAL) and was supported by a Postdoctoral Fellowship of the Research Foundation Flanders (FWO-Vlaanderen, T. van Waterschoot). The scientific responsibility is assumed by its authors.

using the combination of FDAF and PEM. We, moreover, use the FDAF-PEM-AFROW framework to improve several state-of-the-art variable step-size (VSS) and variable regularization (VR) algorithms. The FDAF-PEM-AFROW versions significantly outperform the original versions in every simulation by at least 6 dB during DT. In terms of computational complexity the FDAF-PEM-AFROW versions are themselves 10 times cheaper than the original versions.

Index Terms

Double-talk-robust acoustic echo cancellation, acoustic echo cancellation, double-talk, frequency-domain adaptive filters, prediction error method, variable step size.

I. Introduction

Acoustic echo cancellation (AEC) is used in many speech communication application where the existence of echoes degrades the speech intelligibility and listening comfort [1], [2]. These applications range from mobile and hands-free telephony to teleconferencing and voice over IP (VoIP), and are often integrated in smartphones, tablets, notebooks, laptops, etc. The typical set-up for an acoustic echo canceler is depicted in Figure 1.

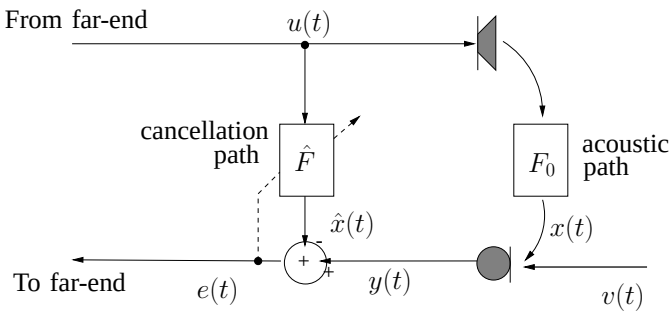


Fig. 1: Typical set-up for AEC

A far-end speech signal $u(t)$ is played back in an enclosure (i.e., the room) through a loudspeaker. In the room there is a microphone to record a near-end speech signal which is to be transmitted to the far-end side. An acoustic echo path between the loudspeaker and the microphone exists so that the microphone signal $y(t) = x(t) + v(t)$ contains an undesired echo signal $x(t)$ plus the near-end speech signal $v(t)$. The echo signal $x(t)$ can be considered as the far-end speech or loudspeaker signal $u(t)$ filtered by the echo path. An acoustic echo canceler seeks to cancel the echo signal component $x(t)$ in

the microphone signal $y(t)$, ideally leading to an *echo-free* error signal $e(t)$, which is then transmitted to the far-end side. This is done by subtracting an estimate of the echo signal $\hat{x}(t)$ from the microphone signal, i.e., $e(t) = y(t) - \hat{x}(t)$. Standard approaches to AEC rely on the assumption that the echo path can be modeled by a linear FIR filter [3]-[5]. The coefficients of the echo path are collected in the parameter vector $\mathbf{f}_0 = [f_0(0), f_0(1), f_0(2), ..., f_0(N-1)]^T \in \mathbb{R}^N$ such that $x(t) = \mathbf{f}_0^T \mathbf{u}(t)$ with $\mathbf{u}(t) = [u(t), u(t-1), \ldots, u(t-N+1)]^T$. An adaptive filter of sufficient order is used to provide an estimate $\hat{\mathbf{f}}(t) \in \mathbb{R}^N$ of $\mathbf{f}_0$, such that the echo signal estimate is $\hat{x}(t) = \hat{\mathbf{f}}^T(t)\mathbf{u}(t)$.

Practical AEC implementations often rely on computationally simple stochastic gradient algorithms (e.g., NLMS) which are very sensitive to the presence of a near-end signal [3]. In general, the presence of near-end speech signal, in a so-called double-talk (DT) scenario, makes the AEC adaptive filter converge slowly or even diverge. There are several approaches to tackle the problem of DT in AEC. We give a brief explanation of four different approaches: two that are based on a *variable step size* (VSS), one based on a *variable regularization* (VR) and one based on a *best linear unbiased estimate* (BLUE) of the echo path.

The first approach is based on the so-called gradient-based VSS algorithms. In this these algorithms, two properties of the gradient vector are used to control the step size [6]-[11] namely, that (1) the norm of the gradient vector is large initially and converges to a small value, ideally zero, at steady state, and that (2) the gradient vector direction generally shows a coherent trend during initial converge in contrast to a random trend around the optimal value at steady state. From this class of algorithms, the only one specifically designed for DT-robust AEC is the *projection-correlation* VSS (PC-VSS) which has been proposed in [11]. PC-VSS is an affine projection algorithm [12] (APA)-based VSS algorithm where the adaptation rate is controlled by a measure of the correlation between instantaneous and long-term averages of the so-called *projection vectors*, i.e., gradient vectors in APA.

The second approach is based on the *non-parametric* VSS (NPVSS) algorithm proposed in [13]. The NPVSS algorithm has been developed in a system identification context, aiming to recover the system noise (i.e., near-end noise) from the error signal of the adaptive filter when updated with the NLMS algorithm. Inspired by this idea, several other approaches have been focused on applying the NPVSS algorithm to real AEC applications where the microphone signal also contains a near-end speech signal. Consequently, different VSS-NLMS algorithms have been developed and applied successfully for DT-robust AEC, e.g., [14] and [15]. Their convergence, however, has been found to be slow in practice. Hence, also an APA version of these algorithms has been proposed to increase the convergence speed,

resulting in the *practical* VSS affine projection algorithm (PVSS-APA) [16].

The third approach is based on equipping the adaptive filter with a *variable regularization*. Several VR algorithms have been proposed in the literature based on the derivation of an optimal regularization parameter which sometimes need quantities or models that are difficult to obtain in practice [17], [18]. Practical implementations of these algorithms have also been proposed that employ easily measurable quantities. One example is the APA-based VR (VR-APA) algorithm that has been proposed in [19] which incorporates the statistics of the noise into the design of the VR.

The fourth approach is based on a minimum-variance echo path estimate, i.e., the *best linear unbiased estimate* (BLUE) [20]. This estimate depends on the near-end signal characteristics, which are in practice unknown and time-varying [3], [21]. The use of the prediction error method (PEM) [22] has been proposed to jointly estimate the echo path model and an autoregressive (AR) model of the near-end speech signal [30]. The algorithms in [3], [30] and [21] aim to whiten the near-end signal speech component in the microphone signal by using adaptive decorrelation filters that are estimated concurrently with the echo path. Among the PEM-based algorithms, the PEM-based adaptive filtering using row operations (PEM-AFROW) [23] is particularly interesting and it will be explained further on. Other algorithms, although not applied to DT-robust AEC, have been proposed for recursively minimizing a BLUE criterion. In fact, in [24], [25], a frequency-domain adaptive filtering (FDAF) algorithm has been obtained by first minimizing a BLUE criterion using the stochastic gradient method and then switching to the frequency domain to reduce the computational complexity. Moreover, it has been shown [24], [25] that by using overlap-save as well as overlap-add methods, two FDAF schemes may be derived from constructing circulant matrix extensions of Toeplitz matrices.

One particular assumption in all these algorithms, and in most AEC applications in general, is that the near-end signal is uncorrelated with the loudspeaker signal. This assumption is valid only asymptotically and applies to ergodic and stationary processes. In real AEC applications, however, the near-end signal is usually a speech signal (or a highly colored and non-stationary noise signal). Indeed, the near-end signal and the loudspeaker signal are *statistically* uncorrelated, which however does not imply that the correlation between these two signals is zero within a short-time window [29]. Therefore, we define, the *instantaneous pseudo-correlation* (IPC) measure between the near-end signal $v(t)$ and the loudspeaker signal $u(t)$ as and indication of the effect of a DT situation occurring during adaptation.

The aim of this paper is to introduce a new framework for DT-robust AEC which is based on a FDAF implementation of the PEM-AFROW (FDAF-PEM-AFROW). We depart from this framework to

show an improvement in performance with respect to the individual versions, i.e., PEM-AFROW and FDAF independently. We show this improvement based on simulation results but also we justify the use of FDAF-PEM-AFROW, basically due to three supporting elements, (1) in case the assumed signal model or the unbiasedness constraints in the PEM-AFROW are violated we will show that a FDAF implementation ameliorates the negative effects associated with it, (2) the FDAF itself may be seen to minimize a BLUE criterion if a proper normalization factor is used during adaptation, and (3) the proposed instantaneous pseudo-correlation (IPC) measure between the near-end signal and the loudspeaker signal is significantly reduced using the combination of FDAF and PEM. Moreover, we use this framework to improve the previously introduced VR-APA, PVSS-APA and PCVSS algorithms. Finally, we show that this framework is computationally much more efficient than the original versions.

The paper is organized as follows. In Section II, the BLUE is explained in two subsections: in Section II-A, the basic linear regression model is explained with unbiasedness constraints that are assumed in the derivation of typical algorithms for AEC. In Section II-B, the generalized least squares model is showed, which provides a framework to explain other related estimators. In Section III, the three element that support the choice of FDAF-PEM-AFROW are explained: In Section III-A, the BLUE achieved using the PEM is considered and in Section III-B it is explained how a BLUE based on applying the DFT can be achieved. In Section III-C, the IPC measure is derived and defined, and several results are shown evaluating the IPC measure using NLMS, DFT-NLMS and FDAF-NLMS. The recursions that are used to compute the IPC measure are given in Appendix A. In Section IV, the proposed FDAF-PEM-AFROW is derived: in Section IV-A, the algorithm description is given and in Section IV-B the IPC analysis of PEM-AFROW, DFT-PEM-AFROW and FDAF-PEM-AFROW is provided. In Section VI-B, more detail is provided about the selected state-of-the-art VSS and VR algorithms as well as an explanation of their FDAF-PEM-AFROW versions. In Section V, computer simulations are provided together with a complexity analysis comparing the following algorithms: the PEM-AFROW, APA-PEM-AFROW, DFT-PEM-AFROW, FDAF-PEM-AFROW, FDAF-NLMS, PVSS-APA, PCVSS-APA, VR-APA, PVSS-FDAF-PEM-AFROW, PCVSS-FDAF-PEM-AFROW and VR-FDAF-PEM-AFROW. Finally, Section VI concludes the paper.

II. Best linear unbiased estimate

In this section, we present the general linear regression model and the constraints for an estimator to be unbiased.

A. Linear unbiased estimator

We first assume that the echo path is time-invariant, $\mathbf{f}(t) = \mathbf{f}, \forall t$ over the estimation window $t = 1, 2, \ldots, L$ with $L >> N$. A regression data model may be constructed as

$$\mathbf{y} = \mathbf{U}\mathbf{f} + \mathbf{v} \tag{1}$$

where

$$\mathbf{y} = [y(t), y(t-1), ..., y(1))]^T \tag{2}$$

$$\mathbf{v} = [v(t), v(t-1), ..., v(1))]^T \tag{3}$$

$$\mathbf{U} = [\mathbf{u}(t), \mathbf{u}(t-1), ..., \mathbf{u}(1)]^T, \tag{4}$$

with $(\cdot)^T$ the transpose operator. Any *linear* estimate of parameter vector $\mathbf{f}$, given the data up till time $t = L$, can be written as a linear function of the data vector $\mathbf{y}$, i.e.,

$$\hat{\mathbf{f}} = \mathbf{D}^T\mathbf{y}. \tag{5}$$

For this estimate to be *unbiased*, the $L \times N$ matrix $\mathbf{D}$ should be subjected to two constraints,

$$\mathbf{D}^T\mathbf{U} = \mathbf{I}_N \tag{6}$$

$$\mathcal{E}\{\mathbf{D}^T\mathbf{v}\} = \mathbf{0}_{N\times 1} \tag{7}$$

These constraints are often implicitly assumed to hold for the derivation of the majority of AEC algorithms.

B. Generalized least squares and BLUE

Let us now consider a *generalized least squares* (GLS) estimator given as

$$\hat{\mathbf{f}}_{\text{GLS}} = \left[\mathbf{U}^T\mathbf{M}^{-1}\mathbf{U}\right]^{-1}\mathbf{U}^T\mathbf{M}^{-1}\mathbf{y} \tag{8}$$

This estimator is quite general since the structure of the weighting matrix $\mathbf{M}$ can result in different types of estimators [26], i.e., biased or unbiased, minimum-variance or not. As we are interested in an $\mathbf{M}$ that makes (8) a BLUE [3], we consider

$$\mathbf{M} = \mathbf{R}_v = \mathcal{E}\{\mathbf{v}\mathbf{v}^T\}, \tag{9}$$

where $\mathbf{R}_v$ is the near-end signal autocorrelation matrix with symmetric Toeplitz structure, such that

$$\hat{\mathbf{f}}_{\text{BLUE}} = \left[\mathbf{U}^T\mathbf{R}_v^{-1}\mathbf{U}\right]^{-1}\mathbf{U}^T\mathbf{R}_v^{-1}\mathbf{y}. \tag{10}$$

Note that, when the near-end signal is a white noise with variance σ_v^2, the weighting matrix $\mathbf{R}_v = \sigma_v^2 \mathbf{I}_L$, where $\mathbf{I}_L$ is the $L \times L$ identity matrix. If we substitute this into (10) the ordinary least squares (OLS) estimator is obtained, i.e.,

$$\hat{\mathbf{f}}_{\text{LS}} = \left[\mathbf{U}^T \mathbf{U}\right]^{-1} \mathbf{U}^T \mathbf{y} \tag{11}$$

This means that the OLS is a BLUE only if the errors, $\mathbf{v}$, in the observations, $\mathbf{y}$, are not correlated, otherwise the OLS estimate is suboptimal. Most of the existing AEC algorithms are based on the OLS estimator which is assumed to be unbiased, following (6), (7), but which is not necessarily minimum-variance. In AEC applications, the errors correspond to the near-end speech signal and the, typically colored, near-end noise signal. Hence, these errors are expected to increase the estimation variance.

III. The elements of FDAF-PEM-AFROW

In this section, we explain the three main elements that support the choice of FDAF-PEM-AFROW as a framework for DT-robust AEC. In the next two sections, we present several results to show how a BLUE of the echo path can be achieved under certain assumptions: Firstly, we consider [3] and [21] where a formal relationship between the PEM-based adaptive filtering using row operations (PEM-AFROW) [23] and the BLUE is given. Secondly, we explain some classical results given in [24] and [26] relating DFT-based estimators to the BLUE. Finally, we derive and define the IPC measure, and show several results evaluating the IPC measure using NLMS, DFT-NLMS and FDAF-NLMS.

A. PEM-based BLUE

Note that a BLUE as in (10) cannot be calculated as such, because the near-end signal autocorrelation matrix $\mathbf{R}_v$ is usually unknown. To this end, the near-end signal autocorrelation matrix $\mathbf{R}_v$ is conveniently decomposed. Let us first assume that the near-end signal is generated as $v(t) = H(q,t)w(t)$ where $w(t)$ is a white noise signal with $\mathcal{E}\{w(t)w(t-i)\} = \delta(i)\sigma_w^2(t)$, and

$$H(q,t) = \frac{1}{A(q,t)} = \frac{1}{1 + a_1(t)q^{-1} + ... + a_{n_A}(t)q^{-n_A}}. \tag{12}$$

is the near-end signal autorregresive (AR) model, expressed as an order-n_A time-varying linear filter with q^{-1} the unit delay operator, i.e., $q^{-1}u(t) = u(t-1)$. The near-end signal autocorrelation matrix may then be written as

$$\mathbf{M}_{\text{PEM}} = \mathbf{R}_v = \mathcal{E}\{\mathbf{H}\mathbf{w}\mathbf{w}^T\mathbf{H}^T\} \tag{13}$$

where $\mathbf{w} = [w(t), ..., w(1)]^T$ and the matrix $\mathbf{H}$ contains the filter coefficients $H(q,t)$, $t = 1, ..., L$. In [3], [21] it is shown that a BLUE can be realized as

$$\hat{\mathbf{f}}_{\text{BLUE}} = \left[\mathbf{U}^T \left(\mathbf{H}^{-T} \boldsymbol{W}_{\text{PEM}}^{-1} \mathbf{H}^{-1}\right) \mathbf{U}\right]^{-1} \mathbf{U}^T \left(\mathbf{H}^{-T} \boldsymbol{W}_{\text{PEM}}^{-1} \mathbf{H}^{-1}\right) \mathbf{y} \tag{14}$$

with

$$\boldsymbol{W}_{\text{PEM}} = \begin{bmatrix} \sigma_w^2(1) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_w^2(t) \end{bmatrix}, \tag{15}$$

that is, by prefiltering and weighting the kth row of $\mathbf{U}$ and $\mathbf{y}$ with the inverse near-end signal model $H^1(q,k) = A(q,k)$ and the inverse near-end excitation signal variance $\sigma_w^{-2}(k)$. In practice, the AR model parameters and $\sigma_w^{-2}(t)$ have to be estimated concurrently with the echo path $\mathbf{f}(t)$ at each time instant t. The use of the prediction error method (PEM) [22] has been proposed to jointly estimate the echo path and an autoregressive (AR) model of the near-end speech signal [30].

In [21], it has been shown that PEM-based algorithms that minimize a BLUE criterion with respect to the echo path model $\mathbf{f}(t)$ only, are possible if three assumptions are fulfilled, (1) the near-end signal $v(t)$ can at each time instant $t \in [1, L]$ be modeled as an AR process of order n_A, (2) the prefilter $\hat{A}(q,t)$ contains at each time instant the true AR coefficients and (3) the weight $\hat{\sigma}_w^{-2}(t)$ is at each time instant equal to the inverse variance of the near-end excitation signal $w(t)$, i.e., $\hat{\sigma}_w^{-2}(t) = \sigma_w^{-2}(t)$.

PEM-based algorithms that minimize a BLUE criterion lead to very simple stochastic gradient algorithms featuring a decorrelation prefilter (as shown in Figure 2), $\hat{A}(q,t)$, and a weighing scalar, $\hat{\sigma}_w^2(t)$, in the denominator of the adaptive filter update equation [3]. However, it seems that some of the previous assumptions are difficult to fulfill in practice, which will make the weighting matrix (15) to be densely populated. In this case, the use of a single weighting scalar $\hat{\sigma}_w^2(t)$ in the denominator of the adaptive filter update equation would not be possible. Even if the near-end signal autocorrelation matrix $\mathbf{R}_v$ could be estimated as such, its direct inversion, as needed in a typical stochastic gradient algorithm [24], would be prohibitive in terms of computational complexity.

B. DFT-based BLUE

In this section, we show that both relaxing the three previous assumptions (essentially, the structure of the near-end signal autocorrelation matrix) and achieving a BLUE is possible, if the adaptive filter is implemented in the frequency domain. In [26], a GLS estimator in the frequency domain is derived that optimally robustifies the OLS estimator against correlated noise. The optimal robust estimator minimizes

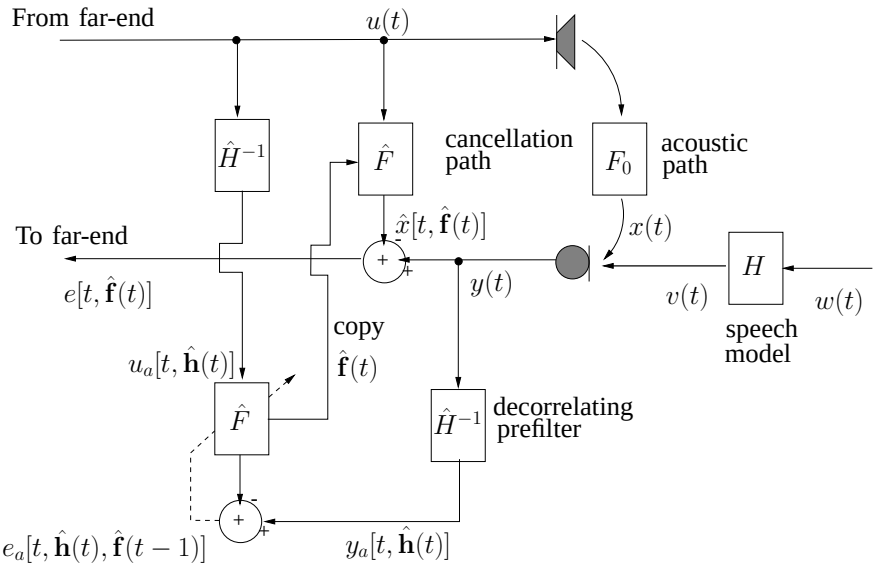


Fig. 2: Typical set-up for AEC using a decorrelation prefilter.

the asymptotic variance if the discrete Fourier transform (DFT) is applied to the observed data. Therefore, by applying the DFT to the input matrix, $\mathbf{X} = \mathbf{\Omega U}$, and to the observation vector, $\mathbf{Y} = \mathbf{\Omega y}$, where $\mathbf{\Omega}$ is the $L \times L$ unitary DFT matrix, the GLS estimator can be written as

$$\hat{\mathbf{F}}_{\text{GLS}} = \left[\mathbf{X}^H \mathbf{M}_{\text{DFT}}^{-1} \mathbf{X}\right]^{-1} \mathbf{X}^H \mathbf{M}_{\text{DFT}}^{-1} \mathbf{Y} \tag{16}$$

where $(\cdot)^H$ represents the Hermitian transpose of a matrix, and the weighting matrix

$$\mathbf{M}_{\text{DFT}} = \mathcal{E}\{\mathbf{\Omega v v}^T \mathbf{\Omega}^H\} \tag{17}$$

$$= \mathbf{\Omega R}_v \mathbf{\Omega}^H \tag{18}$$

It is known [28] that for large L, such that the Toeplitz matrix $\mathbf{R}_v$ resembles a circulant matrix, the weighting matrix $\mathbf{M}_{\text{DFT}}$ is *nearly diagonal*. Based on this observation, we approximate $\mathbf{M}_{\text{DFT}}$ by a diagonal matrix $\mathbf{W}_{\text{DFT}}$. This approximation converts (16) into,

$$\hat{\mathbf{F}}_{\text{BLUE}} = \left[\mathbf{X}^H \mathbf{W}_{\text{DFT}}^{-1} \mathbf{X}\right]^{-1} \mathbf{X}^H \mathbf{W}_{\text{DFT}}^{-1} \mathbf{Y} \tag{19}$$

where the elements of $\mathbf{W}_{\text{DFT}}$ are given as

$$\mathbf{W}_{\text{DFT}} = \begin{bmatrix} \Sigma_v^2\left(\frac{2\pi 0}{N}\right) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \Sigma_v^2\left(\frac{2\pi(N-1)}{N}\right) \end{bmatrix} \tag{20}$$

with $\Sigma_v^2\left(\frac{2\pi k}{N}\right)$, $k = 0, ..., N-1$ equal to the variance of the near-end signal in the k-th frequency bin.

From (19) a FDAF that minimizes a BLUE criterion can be easily derived. In principle, having unequal (frequency-domain) variance values would pose a problem if a (time-domain) stochastic gradient algorithm is to be derived. Conversely, in an FDAF, the near-end signal variance estimate in each frequency bin may be straightforwardly incorporated in the normalization factor of the adaptive filter update equation. In fact, in [24], and [25], it is shown that a rather simple modification of the normalization factors can significantly improve steady-state performance of an FDAF if a colored near-end noise signal is present.

C. Instantaneous pseudo-correlation in NLMS-based algorithms

In the previous sections, we have shown the unbiasedness constraints in (6), (7) must be satisfied in both the PEM-based and DFT-based cases. Since $\mathbf{D}$ in (7) is a function of the loudspeaker signal matrix $\mathbf{U}$, the constraint reduces to $\mathcal{E}\{\mathbf{U}^T\mathbf{v}\} = \mathbf{0}_{N\times 1}$. This means that the near-end signal should be uncorrelated with the loudspeaker signal. This assumption is valid only asymptotically and applies to ergodic and stationary processes [4]. In real AEC applications, however, the near-end signal is usually a speech signal (or a, typically colored and non-stationary, noise signal). Indeed, the near-end signal and the loudspeaker signal are *statistically* uncorrelated, which however does not imply that the correlation between these two signals is zero when using a short-time window [29]. It is known [30], [31] that the correlation between these two signals causes standard adaptive filtering algorithms to converge to a *biased* solution. This means that the adaptive filter does not only predict and cancel the echo signal component in the microphone signal, but also part of the near-end speech signal.

In this section, we define, the *instantaneous pseudo-correlation* (IPC) measure between the near-end signal $v(t)$ and the loudspeaker signal $u(t)$. To this end, we first consider the least mean squares (LMS) algorithm [32] update equation given as

$$e(t) = y(t) - \hat{\mathbf{f}}^T(t-1)\mathbf{u}(t) \tag{21}$$

$$\hat{\mathbf{f}}(t) = \hat{\mathbf{f}}(t-1) + \mu e(t)\mathbf{u}(t). \tag{22}$$

By repeated application of (22), it is straightforwardly shown that

$$\hat{\mathbf{f}}(t-1) = \mu \sum_{i=1}^{t-1} e(i)\mathbf{u}(i) \tag{23}$$

so that

$$e(t) = y(t) - \mu \sum_{i=1}^{t-1} e(i) < \mathbf{u}(i), \mathbf{u}(t) > \tag{24}$$

where $<,>$ represents the inner product operation and $y(t) = x(t) + v(t)$. If we express the estimated echo signal $\hat{x}(t)$, i.e., the second term of the right hand side (RHS) of (24), using the 'true error' signal $e_0(t)$ (i.e., the error signal obtained if $v(t) = 0$) and the near-end signal $v(t)$, we obtain

$$e(t) = x(t) - \mu \sum_{i=1}^{t-1} e_0(i) < \mathbf{u}(i), \mathbf{u}(t) > + v(t) - \mu \sum_{i=1}^{t-1} v(i) < \mathbf{u}(i), \mathbf{u}(t) >. \tag{25}$$

In this representation (25), we are able to identify a signal, let us call it $z(t) = \mu \sum_{i=1}^{t-1} v(i) < \mathbf{u}(i), \mathbf{u}(t) >$ which resembles a correlation function, hence the name IPC. Logically, we want $z(t)$ to be as small as possible, ideally $z(t) = 0$. This would imply that $v(t)$ and $u(t)$ are uncorrelated, and, as a consequence, the near-end signal $v(t)$ would be preserved and transmitted to the far-end side without distortion. Unfortunately, this is not the case in practice. To this end, we show that several algorithms can help to reduce the IPC between these two signals and, as a result, the signal $z(t)$ will be small.

In what follows, we show that, within their adaptation loops, the NLMS, the transform-domain NLMS using DFT (DFT-NLMS), and the frequency-domain adaptive filtering using NLMS (FDAF-NLMS) show different IPC values. We consider the recursions given in Appendix A where it is shown how the signal $z(t)$ is computed in the NLMS, the DFT-NLMS and the FDAF-NLMS algorithms. As a measure of the IPC between the near-end signal and the loudspeaker signal within an adaptation loop we consider

$$\text{IPC measure} = 10 \log_{10} \frac{\sum_{t=1}^{L} z(t)^2}{\sum_{t=1}^{L} v(t)^2}, \tag{26}$$

i.e., the variance of the signal $z(t)$ normalized w.r.t. the variance of the near-end signal with L the total length of the signals.

Table I shows the values of the IPC measure for the NLMS, the DFT-NLMS and the FDAF-NLMS. The loudspeaker signal is a female speech signal and the near-end signal is a white noise signal, a colored noise signal or male speech signal, and $N = 80$ taps which is the length of the impulse response of the system under study. The IPC measure obtained with NLMS is higher than that of DFT-NLMS and FDAF-NLMS in each case of near-end signal. The smallest IPC measure value is obtained using the FDAF-NLMS followed by the DFT-NLMS. It can also be observed that the IPC measure is smaller when the near-end signal is a white noise signal. The highest IPC measure is obtained with a near-end speech signal.

In addition, the IPC measures for different filter lengths are shown in Fig. 3. It is shown that by increasing the filter length, the IPC in FDAF-NLMS and DFT-NLMS is considerably reduced. Conversely,

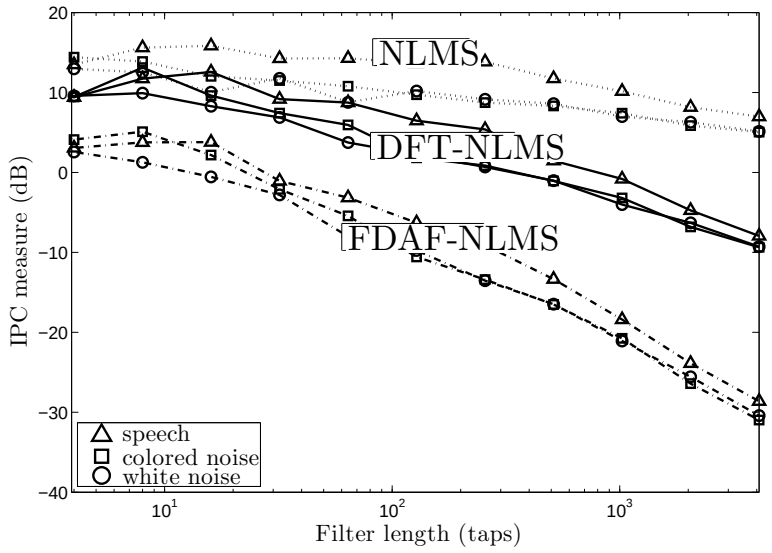


Fig. 3: IPC measure for the NLMS, the DFT-NLMS and the FDAF-NLMS. The loudspeaker signal is a female speech signal and the near-end signal is a white noise signal, a colored noise signal or male speech signal.

$N = 80$	White noise	Colored noise	Speech
NLMS	11.3 dB	11.7 dB	25.9 dB
DFT-NLMS	2.7 dB	3.5 dB	5.4 dB
FDAF-NLMS	0.18 dB	0.24 dB	0.4 dB

TABLE I: IPC measure for the NLMS, the DFT-NLMS and the FDAF-NLMS. The loudspeaker signal is a female speech signal and the near-end signal is a white noise signal, a colored noise signal or male speech signal.

in the NLMS case, the IPC measure is only slightly reduced. Although FDAF-NLMS outperforms DFT-NLMS, the latter also obtains some reduction in the IPC measure compared to NLMS. This is the reason why in [33] we also obtained good results during DT. The IPC measure evaluated using these NLMS-based algorithms is consistently smaller when the near-end signal is a white noise signal compared to a colored noise signal (middle), and finally a speech signal signal (largest).

To conclude, in this section we have reviewed two frameworks to achieve a BLUE, namely the PEM-based and the DFT-based frameworks. The PEM-based framework has been successfully applied for DT-robust AEC but it seems that some of its assumptions are difficult to fulfill in practice. On the other hand, the DFT-based framework brings some other benefits for the AEC to approach a BLUE provided

that it is implemented in the frequency domain. A common assumption in both frameworks is that the loudspeaker signal and near-end signal are uncorrelated. In this regard, we have shown that the IPC can be considerably reduced using an FDAF implementation. Based on these observations we propose the use of FDAF and PEM together in one algorithm to obtain a even better approximation to the BLUE.

IV. Proposed algorithm

In this section, we give a description of the proposed algorithm and a calculation of the IPC measure (26) for the proposed algorithm.

A. Algorithm description

Among the PEM-based algorithms proposed in [3], [21], the PEM-based adaptive filtering using row operations (PEM-AFROW) [23] is particularly interesting. In AEC applications, the near-end speech signal $v(t)$ is short-term stationary, which implies that the near-end speech model $A(q,t)$ does not need to be re-estimated at each time instant t. That is, instead of identifying the near-end signal AR coefficients recursively, these can also be identified non-recursively on a batch of loudspeaker and microphone data. This is the idea behind the PEM-AFROW, which estimates $A(q,t)$ in a frame-based manner, using a frame length that approximates the stationary interval of the speech signal. Both $\hat{\mathbf{a}}(t)$ and $\hat{\sigma}_w^2(t)$ are efficiently computed using the Levinson-Durbin recursion. The algorithm proposed in this paper implements the PEM-AFROW using the well-known frequency-domain adaptive filter (FDAF-PEM-AFROW). In particular the gradient constraint FDAF is implemented using the overlap-save method where the size of each input frame is N and so the FFT size is $M = 2N$. For a complete description of the PEM-AFROW and the FDAF the reader is referred to [23] and [34] respectively. Here, we only give details of the particular implementation of the algotihm. A complete description is provided as Algorithm 1 where $\mathcal{F}/\mathcal{F}^{-1}$ represents a DFT/IDFT operation, $(\cdot)$ represents an element-wise multiplication, $[\cdot]_{a:b}$ represents a range of samples within a vector, k is the block-time index, capital letters represent frequency-domain variables, e.g. $\mathbf{U}(k)$, while lower-case letters represent time-domain variables, e.g. $\mathbf{u}(k)$, variables in boldface, e.g., $\mathbf{U}_a(k)$ represent vectors and non-boldface variables represent scalars, e.g., $G(m,k)$.

In *line* 5, two length-N error vectors are concatenated in $\mathbf{r}(k)$ to calculate both the AR coefficients and the near-end excitation signal variance using the Levison-Durbin algorithm on $\mathbf{r}(k)$ assuming $N \leq P \leq 2N$. The order-n_A AR coefficients are used to prefilter the loudspeaker signal $u(t)$ and the microphone signal $y(t)$. With these prefiltered signals a standard FDAF algorithm with gradient constraint is then

Algorithm 1 PEM-AFROW [23] using FDAF (FDAF-PEM-AFROW)

1: Initialize: $M = 2N$, $\hat{\mathbf{F}}(0) = \mathbf{e}^T(0) = \mathbf{S}_{U_a}(0) = \mathbf{0}_{M\times 1}$.
2: Variables $\mathbf{u}(k)$ and $\mathbf{y}(k)$ are M-length vectors satisfying the overlap-save condition in FDAF, $\mathbf{u}(k) = [u(kN - N + 1), ...u(kN + N)]^T$ and $\mathbf{y}(k) = [y(kN - N + 1), ...y(kN + N)]^T$
3: **for** $k = 1, 2, ...$ **do**
4: $\quad \mathbf{e}(k) = \left[\mathbf{y}(k) - \mathcal{F}^{-1}\left\{\mathcal{F}\{\mathbf{u}(k)\} \cdot \hat{\mathbf{F}}(k-1)\right\}\right]_{N+1:M}$
5: $\quad \mathbf{r}(k) = [\mathbf{e}^T(k), \quad \mathbf{e}^T(k-1)]^T$
6: $\quad [\hat{\mathbf{a}}(k) \quad \hat{\sigma}_w^2(k)] =$ Levison-Durbin $\{[\mathbf{r}(k)]_{1:P}, n_A\}$
7: $\quad$ **for** $m = 0, ..., M-1$ **do** (Decorrelation prefilter)
8: $\quad\quad u_a(m,k) = [u(kN + 1 + m), ..., u(kN + 1 + m - n_A)]\hat{\mathbf{a}}(k)$
9: $\quad\quad y_a(m,k) = [y(kN + 1 + m), ..., y(kN + 1 + m - n_A)]\hat{\mathbf{a}}(k)$
10: $\quad$ **end for**
11: $\quad \mathbf{U}_a(k) = \mathcal{F}\{\mathbf{u}_a(k)\}$
12: $\quad \hat{\mathbf{x}}_a(k) = \mathcal{F}^{-1}\{\mathbf{U}_a(k) \cdot \hat{\mathbf{F}}(k-1)\}$
13: $\quad \mathbf{e}_a(k) = [\mathbf{y}_a(k) - \hat{\mathbf{x}}_a(k)]_{N+1:M}$ (Prediction-error signal)
14: $\quad \mathbf{E}_a(k) = \mathcal{F}\{\begin{bmatrix}\mathbf{0}_N & \mathbf{e}_a^T(k)\end{bmatrix}^T\}$
15: $\quad$ **for** $m = 0, 1, ..., M-1$ **do**
16: $\quad\quad S_{U_a}(m,k) = \lambda_0 S_{U_a}(m,k-1) + (1-\lambda_0)|U_a(m,k)|^2$ (Recursive power estimate of loudspeaker signal)
17: $\quad\quad G(m,k) = \left[S_{U_a}(m,k) + \hat{\sigma}_w^2(k) + \alpha\right]^{-1}$ (Normalization factor)
18: $\quad$ **end for**
19: $\quad \boldsymbol{\theta}(k) = \left[\mathcal{F}^{-1}\{\mathbf{G}(k) \cdot \mathbf{U}_a{}^*(k) \cdot \mathbf{E}_a(k)\}\right]_{1:N}$
20: $\quad \hat{\mathbf{F}}(k) = \hat{\mathbf{F}}(k-1) + \mu_0 \mathcal{F}\left\{\begin{bmatrix}\boldsymbol{\theta}^T(k) & \mathbf{0}_N\end{bmatrix}^T\right\}$
21: **end for**

implemented from *line* 11 to *line* 19. It is worth noticing that the normalization factor in *line* 17 contains three terms: (1) the frequency-dependent loudspeaker signal power estimation $S_{U_a}(m,k)$, (2) the non-frequency-dependent near-end signal excitation variance estimated in *line* 6 and (3) a small regularization term α to avoid division by zero. The latter term introduces some bias, and therefore we would be slightly moving away from the linear unbiased estimator framework. We, however, assume this bias to be very small.

B. Instantaneous pseudo-correlation in PEM-AFROW-based algorithms

Table II shows the values of the IPC measure (26) which is now calculated for the PEM-based algorithms, namely for the PEM-AFROW [23], the DFT-PEM-AFROW [33] and the proposed FDAF-PEM-AFROW. The reason for including these three algorithms is to show the impact that either the PEM-AFROW itself (i.e., decorrelation prefilter and weighting scalar) or the type of adaptation (i.e., time-domain, DFT or FDAF) has in the resulting IPC measure. Calculations are performed in the same scenario as before with $N = 80$ and two different AR orders, $n_A = 1$ and $n_A = 12$. The recursions are

the same as those given in Appendix A with an extra calculation for the AR coefficients and the near-end excitation signal variance.

The results when the near-end signal is a white noise signal and $n_A = 1$ are similar to those given in Table I. On the other hand, in the colored noise signal and speech signal cases the improvement with respect to the NLMS-based algorithms is significant. The fact that both a decorrelating prefilter and a weighting scalar are included in the recursions seems to strongly decrease the IPC. Increasing the near-end signal model order, from $n_A = 1$ to $n_A = 12$ also reduces the IPC measure in the colored noise signal and speech signal cases. In the white noise signal case when $n_A = 12$ the IPC measure tends to increase.

V. Selection of VSS algorithms

In this section, we explain the proposed FDAF-PEM-AFROW version of three state-of-the-art algorithms: Variable Regularization (VR) [19], Practical Variable Step Size (PVSS) [16], Projection-Correlation Variable Step Size (PCVSS) [11]. The chosen state-of-the-art VSS and VR algorithms are originally defined using APA. This is the reason for also including the simulation results of PEM-AFROW when the APA update is used in Section IV. It is important to notice that the FDAF-PEM-AFROW given in Algorithm 1 uses the inverse of the estimated variance of the near-end excitation signal $w(t)$, i.e., $\hat{\sigma}_w^2(t)$ to account for the variance in the estimation so as to obtain a BLUE. $\hat{\sigma}_w^2(t)$ is estimated directly using the Levinson-Durbin algorithm and is subsequently used in the adaptation gain $\mathbf{G}(\omega, k)$. On the other hand, in the PVSS [16] and VR [19] algorithms the near-end signal variance σ_v^2 needs to be estimated instead.

$N = 80$ and $n_A = 1$	White noise	Colored noise	Speech
PEM-AFROW	9.1 dB	-0.1 dB	-1.3 dB
DFT-PEM-AFROW	6.1 dB	-7.6 dB	-9 dB
FDAF-PEM-AFROW	-6.2 dB	-20 dB	-21.1 dB
$N = 80$ and $n_A = 12$	White noise	Colored noise	Speech
PEM-AFROW	8.6 dB	-2.4 dB	-4.2 dB
DFT-PEM-AFROW	6.1 dB	-10.9 dB	-12.4 dB
FDAF-PEM-AFROW	-5.7 dB	-21.7 dB	-25.1 dB

TABLE II: IPC measure for PEM-based algorithms. The loudspeaker signal is a female speech signal and the near-end signal is a white noise signal, a colored noise signal or male speech signal. $N = 80$, $n_A = 1$ and $n_A = 12$

- Variable Regularization [19] (Algorithm 2)

In [19] a practical algorithm to design a variable regularization factor for the APA has been proposed. The condition to derive the VR is to minimize the difference between the estimated and true adaptive filter coefficients. For this, they assumed that the l_2 norm of the a posteriori error is equal to the noise variance, similar to the condition imposed in [13], [16]. They also compared the performance with the existing techniques [17], [18] [35], and demonstrated the effectiveness of this algorithm. The implementation of the VR-FDAF-PEM-AFROW is straightforward since the estimate of the near-end signal excitation variance is already available. The VR parameter is obtained in *line* 8 and use included in the normalization factor in *line* 9 as shown in Algorithm 2.

- Practical VSS [16] (Algorithm 3)

In PVSS, the condition that is imposed to preserve the near-end component in the echo-compensated signal is to set the a posteriori error equal to the near-end signal. As this is not possible in practice, an estimate of the power of the near-end signal is calculated and compared to an estimate of the power of the error signal [13], [16]. The PVSS-FDAF-PEM-AFROW version of the original algorithm is given in Algorithm 3. The main steps of this algorithm are: (1) the near-end signal variance estimation which in PVSS-FDAF-PEM-AFROW is calculated using the Levinson-Durbin algorithm, and (2) the VSS calculation in *line* 7 which is frequency-dependent in PVSS-FDAF-PEM-AFROW.

- Projection Correlation VSS [11] (Algorithm 4)

Projection Correlation VSS (PCVSS) belongs to the gradient-based VSS algorithms. It appears that PC-VSS outperforms the algorithms given in [6] and [36] in DT situations and moreover it does not rely on any signal or system model so it is claimed to be easy to control in practice. The adaptation rate is controlled by a measure of the correlation between instantaneous and long-term averages of the so-called *projection vectors*, i.e., gradient vectors in APA. The three main features of PCVSS-FDAF-PEM-AFROW (Algorithm 4) are represented in (1) *line* 7 where the correlation of the current and past gradient estimates is calculated, (2) *line* 8 where the current step size is generated and (3) *line* 9 where the control logic to bound its value is applied.

VI. Simulation results

Simulations are performed using speech signals. The sampling frequency in every simulation is 8 kHz. In the AEC simulations the far-end speech is a female speech signal and the near-end signal is a male speech signal. The microphone signal consists of three concatenated segments of speech: the

first and third 12.5-s segments consist of echo only, the second segment is the sum of echo $+$ near-end signal generating a DT situation of 13 s. Notice that throughout the paper we assume a sufficient-order condition for the acoustic path model, i.e., $N = 80$. Furthermore, the PEM-AFROW algorithm using an APA [12] update (APA-PEM-AFROW) is also included. The tuning parameters of every algorithm are chosen to have a similar initial convergence that also leads to acceptable DT robustness. Their specific values are shown only for the FDAF-PEM-AFROW versions. The Matlab functions, and scripts with tuning parameters, that are used to generate the figures in this paper are available online [1].

The performance measure *Misadjustment* (MSD) is used to evaluate different VSS algorithms. The MSD between the estimated echo path $\hat{\mathbf{f}}(t)$ and the true echo path $\mathbf{f}$ represents the accuracy of the echo path estimation and is defined as,

$$\text{MSD}(t) = 10\log_{10}\frac{\left\|\hat{\mathbf{f}}(t)-\mathbf{f}\right\|_2^2}{\|\mathbf{f}\|_2^2} \tag{27}$$

A. Choice of the FDAF-PEM-AFROW algorithm

In this section, we show some results comparing FDAF-PEM-AFROW to FDAF-NLMS with noise power normalization [24] and to other PEM-AFROW-based algorithms. This comparison is done based on both performance and complexity results. The comparison is structured to clarify whether or not other PEM-AFROW-based algorithms, and FDAF-based algorithms, would be a better option for DT-robust AEC applications than the proposed FDAF-PEM-AFROW algorithm. In particular, Fig. 4(a) shows the performance comparison between FDAF-PEM-AFROW and the FDAF-NLMS with noise power normalization. Note that these two algorithms are constructed from a BLUE framework. In the bottom part of Fig. 4(a), it is clear to see that FDAF-PEM-AFROW generally outperforms, by $5-6$ dB, the FDAF-NLMS with noise power normalization.

Fig. 4(b) shows the comparison of the FDAF-PEM-AFROW and the other PEM-AFROW-based algorithms from Table III: PEM-AFROW, APA-PEM-AFROW, and DFT-PEM-AFROW. The parameters of these algorithms were tuned to have the same initial convergence rate and acceptable DT robustness. It can be observed that FDAF-PEM-AFROW obtains the lowest MSD values both in single talk and in DT. Compared to PEM-AFROW, FDAF-PEM-AFROW achieves $8-9$ dB improvement. APA-PEM-AFROW seems to improve PEM-AFROW by a couple of dB during DT. DFT-PEM-AFROW appears to have same MSD during single talk as FDAF-PEM-AFROW but it is outperformed by FDAF-PEM-AFROW, during

[1] $http://homes.esat.kuleuven.be/\sim pepe/abstract13-13_2.html$

Algorithm	Computational complexity	Total
PEM-AFROW	$\left(8+\frac{4P+2n_A+1}{P}\right)N+\frac{1}{P}n_A^2+\left(4+\frac{4P+2}{P}\right)n_A+\frac{P-1}{P}+10$	980
APA-PEM-AFROW	$\left(2K^2+4K+2+\frac{4P+2n_A+1}{P}\right)N+4K+\frac{1}{P}n_A^2+\left(4+\frac{4P+2}{P}\right)n_A+\frac{P-1}{P}+10$	4356
DFT-PEM-AFROW	$\left(8+\frac{4P+2n_A+1}{P}\right)N+\frac{1}{P}n_A^2+\left(4+\frac{4P+2}{P}\right)n_A+\frac{P-1}{P}$ $+10+6N\log_2 N$	4015
FDAF-PEM-AFROW	$\frac{18M\log_2 M+18M+n_A^2+n_A(4+4M)}{N}$	307

TABLE III: Complexity comparison by the number of FLOPS per recursion. One FFT/IFFT is a $3M\log_2 M$ complexity operation.

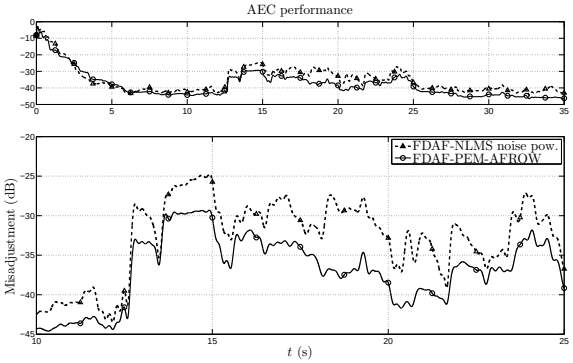
DT by $5-6$ dB. The performance of DFT-PEM-AFROW and FDAF-NLMS with power normalization seems to be comparable although the latter appears to be slightly more robust. FDAF-NLMS appears to outperform both the PEM-AFROW and the APA-PEM-AFROW. FDAF-PEM-AFROW is clearly the algorithm with the lowest MSD of all the other at each time.

Moreover, Table III shows the complexity analysis of the four PEM-AFROW-based algorithms. It is shown that, for a typical choice of the algorithm parameters, FDAF-PEM-AFROW is 3.2 times cheaper than PEM-AFROW itself for a great performance improvement, and more than 14 times cheaper than DFT-PEM-AFROW (using FFTs) and APA-PEM-AFROW. To conclude, this section has shown that FDAF-PEM-AFROW is superior to any of the presented algorithms both in performance as well as complexity-wise. Based on FDAF-PEM-AFROW, a highly robust algorithm have been proposed in [37] using a Wiener variable step size (WISE) and a gradient spectral variance smoothing (GRASS) for DT-robust AEC and acoustic feedback cancellation (AFC).

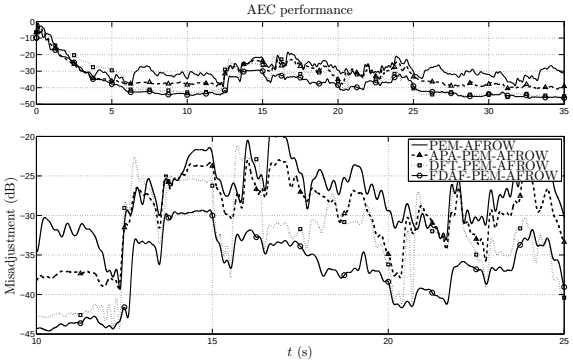
B. Results from VSS algorithms

The three selected VSS algorithms are implemented using FDAF-PEM-AFROW and show different levels of improvement as compared to the original versions using APA with projection order $K=4$. In Fig. 5(a) the upper part illustrates the microphone signal where the dark solid line is the echo and the light green is the near-end signal. In Fig. 5(b) and Fig. 5(c) the upper part shows the full-length simulation and the bottom part shows a zoom-in coinciding with the DT period.

- Variable Regularization (Algorithm 2):



(a)



(b)

Fig. 4: AEC performance using speech signals (far-end and near-end) between 12.5 and 25 s. Bursting DT occurs at -10 dB SER and white near-end noise at 30 dB SNR. The upper part shows the full-length simulation and the bottom part shows a zoom-in coinciding with the DT. Comparison among different algorithms: (a) FDAF-NLMS with noise power normalization [24] and the FDAF-PEM-AFROW (b) PEM-AFROW [23], APA-PEM-AFROW, DFT-PEM-AFROW and FDAF-PEM-AFROW.

In this particular case, shown in Fig. 5(a), the performance of VR-APA is not satisfactory at all. It was difficult to tune in practice to have a comparable initial convergence curve as that of VR-FDAF-PEM-AFROW. On the other hand, it is obvious that VR-FDAF-PEM-AFROW significantly outperforms VR-APA, turning the latter into an algorithm suitable for DT situations as well.

- Practical VSS (Algorithm 3):

The convergence, shown in Fig. 5(a), of PVSS-FDAF-PEM-AFROW seems to be as fast as for the original version during initial convergence and convergence after DT. During DT PVSS-FDAF-PEM-AFROW generally outperforms PVSS-APA by $5-7$ dB. Moreover, it seems that PVSS-FDAF-PEM-AFROW sets

a lower bound in the MSD, meaning that PVSS-APA does not perform any better than that.

- Projection Correlation VSS (Algorithm 4):

The initial convergence, shown in Fig. 5(c), of the two algorithms are similarly fast. During DT, the improvement of PCVSS-FDAF-PEM-AFROW w.r.t. PCVSS-APA becomes apparent. PCVSS-FDAF-PEM-AFROW outperforms PCVSS-APA by $6 - 8$ dB and it also shows a more stable curve. The complexity analysis in Table IV shows that, for a typical choice of the algorithm parameters, the FDAF-PEM-AFROW versions are 10 times cheaper than the original versions. Consequently, it is clear to see that the PVSS-FDAF-PEM-AFROW, PCVSS-FDAF-PEM-AFROW and VR-FDAF-PEM-AFROW algorithms are to be preferred over the original versions during DT periods.

VII. Conclusion

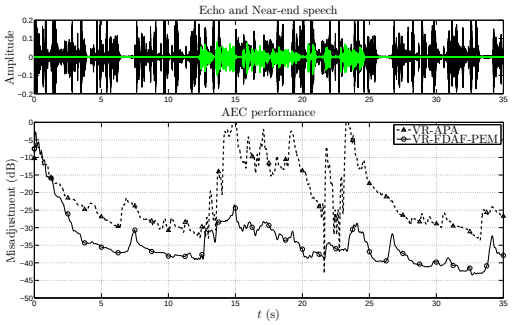
In this paper, we have proposed a new framework to tackle the problem of double-talk (DT) in acoustic echo cancellation (AEC). It is based on a frequency domain adaptive filtering (FDAF) implementation of the so-called PEM-AFROW algorithm (FDAF-PEM-AFROW). We have departed from this framework to show an improvement in performance with respect to the individual versions, i.e., PEM-AFROW and FDAF. This improvement is basically due to three reasons, (1) in case the assumed signal model or the unbiasedness constraints in the PEM-AFROW are violated we have showed that a FDAF implementation ameliorates the negative effects associated with it, (2) the FDAF itself have been seen to minimize a BLUE criterion if a proper normalization factor is used during adaptation, and (3) we have showed that the proposed instantaneous pseudo-correlation (IPC) measure between the near-end signal and the loudspeaker signal is significantly reduced using the combination of FDAF and PEM. We, moreover, have used the FDAF-PEM-AFROW framework to improve several state-of-the-art variable step-size (VSS) and variable regularization (VR) algorithms. In particular, the projection-correlation VSS (PC-VSS), the practical VSS affine projection algorithm (PVSS-APA) and the APA-based VR (VR-APA). It has been showed that the FDAF-PEM-AFROW versions significantly outperform the original versions in every simulation by at least 6 dB during DT. In terms of computational complexity the FDAF-PEM-AFROW versions are themselves 10 times cheaper than the original versions.

Algorithm 2 Variable Regularization [19], using FDAF-PEM-AFROW (VR-FDAF-PEM-AFROW)

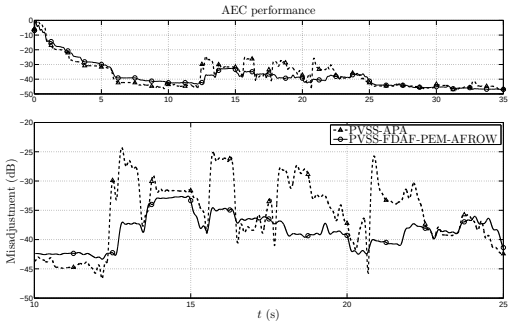
Initialize: $\boldsymbol{S}_{U_a} = \boldsymbol{S}_{E_a} = \mathbf{0}_{M\times 1}$.
for $k = 1, 2, ...$ **do**
  Perform lines $4-14$ from Algorithm 1
  $\mathbf{\Phi}(k) = \mathbf{U}_a{}^*(k) \cdot \mathbf{E}_a(k)$
  for $m = 0, 1, ..., M-1$ **do**
    $S_{U_a}(m,k) = \lambda_0 S_{U_a}(m,k-1) + (1-\lambda_0)|U_a(m,k)|^2$
    $S_{E_a}(m,k) = \lambda_1 S_{E_a}(m,k) + (1-\lambda_1)\,|E_a(m,k)|^2$
    $\alpha(m,k) = S_{U_a}(m,k)\dfrac{\sqrt{\hat{\sigma}_w^2(k)}}{\sqrt{S_{E_a}(m,k)} - \sqrt{\hat{\sigma}_w^2(k)}}$
    $G(m,k) = [S_{U_a}(m,k) + \alpha(m,k)]^{-1}$
  end for
  $\boldsymbol{\theta}(k) = \left[\mathcal{F}^{-1}\left\{\mathbf{G}(k)\cdot\mathbf{\Phi}(k)\right\}\right]_{1:N}$
  $\hat{\mathbf{F}}(k) = \hat{\mathbf{F}}(k-1) + \mu_0\mathcal{F}\left\{\begin{bmatrix}\boldsymbol{\theta}^T(k) & \mathbf{0}_N\end{bmatrix}^T\right\}$
end for
In our simulation we used the following set of parameters: $P = 160$, $N = 80$, $n_A = 1$, $\mu_0 = 0.0325$, $\lambda_0 = 0.95$, $\lambda_1 = 0.9$.

Algorithm 3 PVSS [16] using FDAF-PEM-AFROW (PVSS-FDAF-PEM-AFROW)

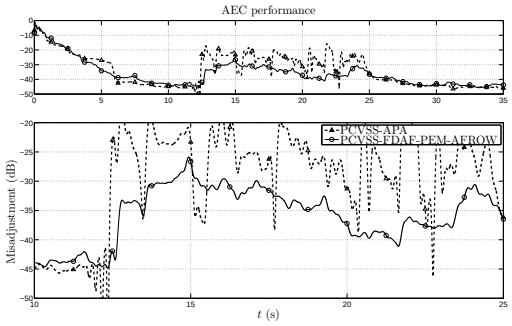
Initialize: $\boldsymbol{S}_{U_a} = \boldsymbol{S}_{E_a} = \mathbf{0}_{M\times 1}$.
for $k = 1, 2, ...$ **do**
  Perform lines $4-14$ from Algorithm 1
  for $m = 0, 1, ..., M-1$ **do**
    $S_{E_a}(m,k) = \lambda_0 S_{E_a}(m,k-1) + (1-\lambda_0)|E_a(m,k)|^2$
    $S_{U_a}(m,k) = \lambda_0 S_{U_a}(m,k-1) + (1-\lambda_0)|U_a(m,k)|^2$
    $\mu_{\text{PVSS}}(m,k) = \mu_0\left|1 - \dfrac{\sqrt{\hat{\sigma}_w^2(k)}}{\sqrt{S_{E_a}(m,k)+\alpha}}\right|$
    $G(m,k) = [S_{U_a}(m,k) + \alpha]^{-1}$
  end for
  $\boldsymbol{\theta}(k) = \left[\mathcal{F}^{-1}\left\{\boldsymbol{\mu}_{\text{PVSS}}(k)\cdot\mathbf{G}(k)\cdot\mathbf{U}_a{}^*(k)\cdot\mathbf{E}_a(k)\right\}\right]_{1:N}$
  $\hat{\mathbf{F}}(k) = \hat{\mathbf{F}}(k-1) + \mathcal{F}\left\{\begin{bmatrix}\boldsymbol{\theta}^T(k) & \mathbf{0}_N\end{bmatrix}^T\right\}$
end for
In our simulations we used the following set of parameters: $P = 160$, $N = 80$, $n_A = 1$, $\lambda_0 = 0.99$, $\alpha = 2e^{-2}$, $\mu_0 = 0.025$, $\alpha = 10^{-8}$.



(a)



(b)



(c)

Fig. 5: AEC performance using speech signals (far-end and near-end) between 12.5 and 25 s. The original APA and the FDAF-PEM-AFROW versions are compared when bursting DT occurs at -10 dB SER and white near-end noise at 30 dB SNR. In (a) the upper part illustrates the microphone signal where the dark solid line is the echo and the light green is the near-end signal. In (b) and (c) The upper part shows the full-length simulation and the bottom part shows a zoom-in coinciding with the DT. (a) Variable Regularization (VR: [19]). (b) Practical VSS (PVSS: [16]). (c) Projection Correlation VSS (PCVSS: [11])

Algorithm 4 PCVSS [11] using FDAF-PEM-AFROW (PCVSS-FDAF-PEM-AFROW)

Initialize: $\mathbf{\Phi}(k<1)=\mathbf{0}_{M\times 1}$, $\mu_{\text{PCVSS}}(k-1)=\mathbf{I}_{M\times 1}$, $\mathbf{S}_{U_a}=\mathbf{C}(k-1)=\mathbf{0}_{M\times 1}$.

for $k=1,2,...$ **do**

  Perform lines $4-14$ from Algorithm 1

  $\mathbf{\Phi}(k)=\mathbf{U}_a{}^*(k)\cdot\mathbf{E}_a(k)$

  for $m=0,1,...,M-1$ **do**

    $S_{U_a}(m,k)=\lambda_0 S_{U_a}(m,k-1)+(1-\lambda_0)|U_a(m,k)|^2$

    $C(m,k)=C(m,k-1)+\Phi(m,k)-\Phi(m,k-B)$

    $\mu'(m,k)=\lambda_1\ \mu_{\text{PCVSS}}(m,k-1)+\beta_0\ |C(m,k)|$

    $\mu_{\text{PCVSS}}(m,k)=\begin{cases}0, & \mu'(m,k)<0\\ \mu_0, & \mu'(m,k)>\mu_0\\ \mu'(m,k), & \text{otherwise}\end{cases}$

    $G(m,k)=[S_{U_a}(m,k)+\alpha]^{-1}$

  end for

  $\boldsymbol{\theta}(k)=\left[\mathcal{F}^{-1}\left\{\boldsymbol{\mu}_{\text{PCVSS}}(k)\cdot\mathbf{G}(k)\cdot\mathbf{U}_a{}^*(k)\cdot\mathbf{E}_a(k)\right\}\right]_{1:N}$

  $\hat{\mathbf{F}}(k)=\hat{\mathbf{F}}(k-1)+\mathcal{F}\left\{\begin{bmatrix}\boldsymbol{\theta}^T(k) & \mathbf{0}_N\end{bmatrix}^T\right\}$

end for

In our simulation we used the following set of parameters: $P=160$, $N=80$, $n_A=1$, $\lambda_0=0.95$, $B=50$, $\beta=\lambda_1=0.9$, $\alpha=1e^{-3}$, $\mu_0=0.03$.

Algorithm	Computation	Total
PVSS	$2K^2N+4KN+15K+8$	3908
PVSS-FDAF-PEM-AFROW	(FDAF) $\frac{21M\log_2 M+34M}{N}+$ (PEM) $+\frac{6M\log_2 M+2M+n_A^2+(4+4M)n_A}{N}$	475
PCVSS	$2K^2N+4KN+4K+4N+12$	4188
PCVSS-FDAF-PEM-AFROW	(FDAF) $\frac{15M\log_2 M+19M+BM}{N}+$ (PEM) $+\frac{6M\log_2 M+2M+n_A^2+(4+4M)n_A}{N}$	367
VR	$2K^2N+4KN+4K$	3878
VR-FDAF-PEM-AFROW	(FDAF) $\frac{15M\log_2 M+31M}{N}+$ (PEM) $+\frac{6M\log_2 M+2M+n_A^2+(4+4M)n_A}{N}$	381

TABLE IV: Complexity comparison by the number of FLOPS per recursion. One FFT/IFFT is $3M\log_2 M$ FLOPS, one $\sqrt{\cdot}$ is M FLOPS and $K=4$

APPENDIX A
IPC CALCULATION IN THE NLMS, DFT-NLMS AND FDAF-NLMS ALGORITHMS.

The following recursions show how the signal $z(t)$ for the IPC measure is calculated in different adaptive filtering algorithms, namely, the NLMS, the DFT-NLMS and the FDAF-NLMS. To this end, we consider the NLMS update represenation typically found in the literature. Using this representation makes the computation of the signal $z(t)$ more efficient and straightforward than using (23), (24).

$$e(t) = y(t) - \hat{x}(t) \tag{28}$$

$$\hat{\mathbf{f}}(t) = \hat{\mathbf{f}}(t-1) + \mu \frac{e(t)}{\alpha + \mathbf{u}^T(t)\mathbf{u}(t)} \mathbf{u}(t). \tag{29}$$

With $y(t) = x(t) + v(t)$ the error signal is

$$e(t) = x(t) - \hat{\mathbf{f}}^T(t-1)\mathbf{u}(t) + v(t) \tag{30}$$

$$= e_0(t) + v(t) \tag{31}$$

so that similarly the adaptive filter update equation becomes

$$\hat{\mathbf{f}}(t) = \hat{\mathbf{f}}_0(t-1) + \mu \frac{e_0(t)}{\alpha + \mathbf{u}^T(t)\mathbf{u}(t)} \mathbf{u}(t) + \tilde{\mathbf{f}}(t-1) + \mu \frac{v(t)}{\alpha + \mathbf{u}^T(t)\mathbf{u}(t)} \mathbf{u}(t) \tag{32}$$

$$= \hat{\mathbf{f}}_0(t) + \tilde{\mathbf{f}}(t). \tag{33}$$

where

$$\hat{\mathbf{f}}_0(t) = \hat{\mathbf{f}}_0(t-1) + \mu \frac{e_0(t)}{\alpha + \mathbf{u}^T(t)\mathbf{u}(t)} \mathbf{u}(t) \qquad \text{and} \qquad \tilde{\mathbf{f}}(t) = \tilde{\mathbf{f}}(t-1) + \mu \frac{v(t)}{\alpha + \mathbf{u}^T(t)\mathbf{u}(t)} \mathbf{u}(t) \tag{34}$$

Hence $\hat{\mathbf{f}}_0^T(t)$ is the 'true filter estimate' (i.e., when $v(t) = 0$). The error signal that is computed at each iteration would therefore become

$$e(t) = x(t) - \hat{\mathbf{f}}_0^T(t-1)\mathbf{u}(t) + v(t) - \tilde{\mathbf{f}}^T(t-1)\mathbf{u}(t) \tag{35}$$

and the signal $z(t) = \tilde{\mathbf{f}}^T(t-1)\mathbf{u}(t)$ will be used in the IPC measure.

The parameters are chosen as $\lambda = 0.99$, $\alpha_0 = N\sigma_u^2$ with σ_u^2 being the variance of the input signal. The recursion within the NLMS algorithm is given as

$$g(t) = \frac{1}{\mathbf{u}^T(t)\mathbf{u}(t) + \alpha_0} \tag{36}$$

$$\tilde{\mathbf{f}}_{\text{NLMS}}(t) = \tilde{\mathbf{f}}_{\text{NLMS}}(t-1) + g(t)\mathbf{u}(t)v(t) \tag{37}$$

$$z_{\text{NLMS}}(t) = \tilde{\mathbf{f}}_{\text{NLMS}}^T(t-1)\mathbf{u}(t), \tag{38}$$

similarly, the recursion in the DFT-NLMS is given as,

$$\mathbf{U}(t) = \frac{1}{\sqrt{N}}\mathcal{F}\{\mathbf{u}(t)\} \tag{39}$$

$$S_U(m,t) = \lambda S_U(m,t-1) + (1-\lambda)|U(m,t)|^2 \quad m = 1,...,N \tag{40}$$

$$G(m,t) = [S_U(m,t) + \alpha_0]^{-1} \quad m = 1,...,N \tag{41}$$

$$\tilde{\mathbf{F}}_{\text{DFT}}(t) = \tilde{\mathbf{F}}_{\text{DFT}}(t-1) + \mathbf{G}(t)\cdot\mathbf{U}(t)\cdot v(t) \tag{42}$$

$$z_{\text{DFT}}(t) = \text{First element of } \left[\mathcal{F}^{-1}\left\{\tilde{\mathbf{F}}_{\text{DFT}}(t-1)\cdot\mathbf{U}(t)\right\}\sqrt{M}\right] \tag{43}$$

Finally the recursion in the FDAF-NLMS is written as,

$$\mathbf{U}(k) = \frac{1}{\sqrt{M}}\mathcal{F}\left\{[\mathbf{u}(k-1) \quad \mathbf{u}(k)]\right\} \tag{44}$$

$$\mathbf{V}(k) = \frac{1}{\sqrt{M}}\mathcal{F}\left\{[\mathbf{0}_N \quad \mathbf{v}(k)]\right\} \tag{45}$$

$$S_U(m,k) = \lambda S_U(m,k-1) + (1-\lambda)|U(m,k)|^2 \quad m = 1,...,M \tag{46}$$

$$G(m,k) = [S_U(m,k) + \alpha_0]^{-1} \quad m = 1,...,M \tag{47}$$

$$\tilde{\mathbf{F}}_{\text{FDAF}}(k) = \tilde{\mathbf{F}}_{\text{FDAF}}(k-1) + \frac{1}{\sqrt{M}}\mathcal{F}\left\{[\mathbf{I}_N \quad \mathbf{0}_N]^T\mathcal{F}^{-1}\left\{\mathbf{G}(k)\cdot\mathbf{V}(k)\cdot\mathbf{U}^*(k)\right\}\sqrt{M}\right\} \tag{48}$$

$$\mathbf{z}_{\text{FDAF}}(k) = \text{First } N \text{ elements of } \left[\mathcal{F}^{-1}\left\{\tilde{\mathbf{F}}_{\text{FDAF}}(k-1)\cdot\mathbf{U}(k)\right\}\sqrt{M}\right] \tag{49}$$

where $(\cdot)^*$ represents a complex conjugation operation.

REFERENCES

[1] J. Benesty, T. Gänsler, D. R. Morgan, M. M. Sondhi, and S. L. Gay, *Advances in Network and Acoustic Echo Cancellation*. Berlin: Springer-Verlag, 2001.

[2] P. Gil-Cacho, T. van Waterschoot, M. Moonen, and S. H. Jensen, "Multi-microphone acoustic echo cancellation using multi-channel warped linear prediction of common acoustical poles," in *Proc. 18th European Signal Process. Conf. (EUSIPCO'10)*, Aalborg, Denmark, Aug. 2010, pp. 2121–2125.

[3] T. van Waterschoot, G. Rombouts, P. Verhoeve, and M. Moonen, "Double-talk-robust prediction error identification algorithms for acoustic echo cancellation," *IEEE Trans. Signal Process.*, vol. 55, no. 3, pp. 846–858, Mar. 2007.

[4] S. Haykin, *Adaptive Filter Theory*. Upper Saddle River, NY: Prentice Hall, 2002.

[5] P. S. R. Diniz, *Adaptive filtering: Algorithms and Practical Implementations*. Boston, MA: Springer, 2008.

[6] V. J. Mathews and Z. Xie, "A stochastic gradient adaptive filter with gradient adaptive step size," *IEEE Trans. Signal Process.*, vol. 41, no. 6, pp. 2075–2087, June 1993.

[7] Y. Zhang, J. A. Chambers, W. Wang, P. Kendrick, and T. J. Cox, "A new variable step-size LMS algorithm with robustness to nonstationary noise," in *Proc. 2007 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP'07)*, vol. 3, Honolulu, Hawaii, USA, Apr. 2007, pp. 1349–1352.

[8] W.-P. Ang and B. Farhang-Boroujeny, "A new class of gradient adaptive step-size LMS algorithms," *IEEE Trans. Signal Process.*, vol. 49, no. 4, pp. 805–810, Apr. 2001.

[9] H.-C. Shin, A. H. Sayed, and W.-J. Song, "Variable step-size NLMS and affine projection algorithms," *IEEE Signal Process. Lett.*, vol. 11, no. 2, pp. 132–135, Feb. 2004.

[10] Y. Zhang, N. Li, J. A. Chambers, and Y. Hao, "New gradient-based variable step size LMS algorithms," *EURASIP J. Advances Signal Process.*, vol. 2008, no. 105, Jan. 2008.

[11] T. Creasy and T. Aboulnasr, "A projection-correlation algorithm for acoustic echo cancellation in the presence of double talk," in *Proc. 2000 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP'00)*, vol. 1, Istanbul, Turkey, June 2000, pp. 436–439.

[12] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," *Electronics and Communication in Japan*, vol. 67, no. 5, pp. 19–27, Aug. 1984.

[13] J. Benesty, H. Rey, L. R. Vega, and S. Tressens, "A nonparametric VSS NLMS algorithm," *IEEE Signal Process. Lett.*, vol. 13, no. 10, pp. 581–584, Oct. 2006.

[14] C. Paleologu, S. Ciochină, and J. Benesty, "Double-talk robust VSS-NLMS algorithm for under-modeling acoustic echo cancellation," in *Proc. 2008 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP'08)*, Las Vegas, USA, Mar. 2008.

[15] M. A. Iqbal and S. L. Grant, "Novel variable step size NLMS algorithms for echo cancellation," in *Proc. 2008 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP'08)*, Las Vegas, USA, Mar. 2008, pp. 241 – 244.

[16] C. Paleologu, J. Benesty, and S. Ciochină, "A variable step-size affine projection algorithm designed for acoustic echo cancellation," *IEEE Trans. Audio Speech Lang. Process.*, vol. 16, no. 8, pp. 1466 – 1478, Nov. 2008.

[17] H. Rey, L. R. Vega, S. Tressens, and J. Benesty, "Variable explicit regularization in affine projection algorithm: Robustness issues and optimal choice," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 2096 – 2108, May 2007.

[18] Y.-S. Choi, H.-C. Shin, and W.-J. Song, "Adaptive regularization matrix for affine projection algorithm," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 54, no. 12, pp. 1087–1091, Dec. 2007.

[19] W. Yin and A. S. Mehr, "A variable regularization method for affine projection algorithm," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 57, no. 6, pp. 476 – 480, June 2010.

[20] S. M. Kay, *Fundamentals of statistical signal processing: Estimation theory*. Upper Saddle River, New Jersey: Prentice Hall, 1993.

[21] T. van Waterschoot and M. Moonen, "Double-talk robust acoustic echo cancellation with continuous near-end activity," in *Proc. 13th European Signal Process. Conf. (EUSIPCO '05)*, Antalya, Turkey, Sep. 2005.

[22] L. Ljung, *System Identification: Theory for the user*. Englewood Cliffs, New Jersey: Prentice Hall, 1987.

[23] G. Rombouts, T. van Waterschoot, K. Struyve, and M. Moonen, "Acoustic feedback cancellation for long acoustic paths using a nonstationary source model," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3426–3434, Sep. 2006.

[24] T. Trump, "A frequency domain adaptive algorithm for colored measurement noise environment," in *Proc. 1998 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP'98)*, vol. 3, Seattle, Washington, USA, May 1998, pp. 1705 – 1708.

[25] ——, "Accounting for measurement noise color in frequency domain adaptive algorithms," in *Proc. Thirty-Second Asilomar Conf. Signals Syst. Comp.*, vol. 1, Pacific Grove, CA, USA, May 1998, pp. 508 – 512.

[26] M. L. Hambaba, "Frequency domain analysis of robust signal estimators," *Circuits Syst. Signal Process.*, vol. 11, no. 3, pp. 365–376, Sep. 1992.

[27] T. van Waterschoot, G. Rombouts, K. Struyve, and M. Moonen, "Acoustic echo cancellation in the presence of continuous double-talk," in *Proc. 1st Annual IEEE Benelux/DSP Valley Signal Process. Symp. (SPS-DARTS '05)*, Antwerp, Belgium, Apr. 2004, pp. 53–57.

[28] P. J. Sherman, "Frequency domain analysis of a family of minimax signal estimators," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 36, no. 3, pp. 400–405, Mar. 1988.

[29] P. Loizou, *Speech Enhancement: Theory and Practice*. Boca Raton, Florida: Taylor and Francis., 2007.

[30] A. Spriet, I. Proudler, M. Moonen, and J. Wouters, "Adaptive feedback cancellation in hearing aids with linear prediction of the desired signal," *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 3749–3763, Oct. 2005.

[31] T. van Waterschoot and M. Moonen, "Fifty years of acoustic feedback control: state of the art and future challenges," *Proc. IEEE*, vol. 99, no. 2, pp. 288–327, Feb. 2011.

[32] B. Widrow and M. Hoff, "Adaptive switching circuits," in *Proc. WESCON Conv. Rec.*, vol. 4, 1960, pp. 96–140.

[33] J. M. Gil-Cacho, T. van Waterschoot, M. Moonen, and S. H. Jensen, "Transform domain prediction error method for improved acoustic echo and feedback cancellation," in *Proc. 20th European Signal Process. Int. Conf. (EUSIPCO'12)*, Bucharest, Rumania, Aug. 2012, pp. 2422–2426.

[34] J. J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Process. Mag.*, vol. 9, no. 1, pp. 14–37, Jan. 1992.

[35] V. Myllylä and G. Schmidt, "Pseudo-optimal regularization for affine projection algorithms," in *Proc. 2002 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP'02)*, vol. 2, Orlando, Florida, May 2002, pp. 1917–1920.

[36] C. Rohrs and R. Younce, "Double talk detector for echo canceler and method," US Patent US 4 918 727, 04 17, 1990. [Online]. Available: http://www.patentlens.net/patentlens/patent/US4918727/

[37] J. M. Gil-Cacho, T. van Waterschoot, M. Moonen, and S. H. Jensen, "Wiener variable step size and gradient spectral variance smoothing for double-talk-robust acoustic echo cancellation and acoustic feedback cancellation," *Submitted to Elsevier Signal Process.*, June 7, 2013. pp. 1–30.