Katholieke Universiteit Leuven

Departement Elektrotechniek

ESAT-SISTA/TR 05-161

Double-Talk-Robust Prediction Error Identification Algorithms for Acoustic Echo Cancellation[1]

Toon van Waterschoot[2] [3], Geert Rombouts[2], Piet Verhoeve[4] and Marc Moonen[2]

October 2005
IEEE Transactions on Signal Processing, vol. 55, no.3, Mar. 2007, pp. 846–858

[1]This report is available by anonymous ftp from *ftp.esat.kuleuven.be* in the directory *pub/sista/vanwaterschoot/reports/05-161.pdf*

[2]K.U.Leuven, Dept. of Electrical Engineering (ESAT), Research group SCD(SISTA), Kasteelpark Arenberg 10, B-3001 Leuven, Belgium, Tel. +32 16 321927, Fax +32 16 321970, WWW: *http://www.esat.kuleuven.be/sista-cosic-docarch.* E-mail: *toon.vanwaterschoot@esat.kuleuven.be.*

[3]Toon van Waterschoot is a Research Assistant with the I.W.T. (Flemish Institute for Scientific and Technological Research in Industry). This research work was carried out at the ESAT laboratory of the Katholieke Universiteit Leuven, in the frame of the Belgian Programme on Interuniversity Attraction Poles, initiated by the Belgian Federal Science Policy Office IUAP P5/22 ('Dynamical Systems and Control: Computation, Identification and Modelling'), the Concerted Research Action GOA-AMBioRICS and IWT project 040803: 'SMS4PA-II: Sound Management System for Public Address Systems'. The scientific responsibility is assumed by its authors.

[4]Televic N.V., Leo Bekaertlaan 1, B-8870 Izegem, Belgium, Tel. +32 51 303045, Fax. +32 51 310670, WWW: *http://www.televic.com* E-mail *p.verhoeve@televic.com*

Double-Talk-Robust Prediction Error Identification Algorithms for Acoustic Echo Cancellation

Toon van Waterschoot, *Student Member, IEEE*, Geert Rombouts, Piet Verhoeve, *Senior Member, IEEE*, and Marc Moonen, *Fellow, IEEE*

***Abstract*—The performance of an acoustic echo canceller may be severely degraded by the presence of a near-end signal. In such a double-talk situation, the variance of the echo path estimate typically increases, resulting in slow convergence or even divergence of the adaptive filter. This problem is usually tackled by equipping the echo canceller with a double-talk detector that freezes adaptation during near-end activity. Nevertheless, there is a need for more robust adaptive algorithms since the adaptive filter's convergence may be affected considerably in the time interval needed to detect double-talk. Moreover, in some applications, near-end noise may be continuously present and then the use of a double-talk detector becomes futile. Robustness to double-talk may be established by taking into account the near-end signal characteristics, which are, however, unknown and time varying. In this paper, we show how concurrent estimation of the echo path and an autoregressive near-end signal model can be performed using prediction error (PE) identification techniques. We develop a general recursive prediction error (RPE) identification algorithm and compare it to three existing algorithms from adaptive feedback cancellation. The potential benefit of the algorithms in a double-talk situation is illustrated by means of computer simulations. It appears that especially in the stochastic gradient case a huge improvement in convergence behavior can be obtained.**

***Index Terms*—Acoustic echo cancellation (AEC), adaptive filtering, double-talk, hands-free communication, near-end signal model, prediction error (PE) identification, robustness.**

I. Introduction

ACOUSTIC echo cancellation (AEC) has been a popular research topic in acoustic signal processing, motivated mainly by the increasing demand for hands-free speech communication. A general AEC scenario is shown in Fig. 1. A speech signal $u(t)$ from the far-end side is broadcast in an acoustic enclosure (the "room") by means of a loudspeaker. A microphone is present in the room for recording a local signal $v(t)$ (the "near-end signal"), which is to be transmitted back to the far-end side. An acoustic echo path exists between the loudspeaker and the microphone such that the recorded microphone signal $y(t) = x(t) + v(t)$ contains an undesired echo component $x(t)$ in addition to the near-end signal component $v(t)$. If the echo path transfer function is modeled as a finite-impulse-response (FIR) filter $F(q,t)$, then the echo component can be considered as a filtered version of the loudspeaker signal: $x(t) = F(q,t)u(t)$. Here, q denotes the time shift operator, e.g., $q^{-k}u(t) = u(t-k)$. The main objective in AEC is to identify the unknown room impulse response (RIR) $F(q,t)$, and hence to subtract an estimate of the echo component from the microphone signal. In this way, an echo-compensated signal $d(t) = y(t) - \hat{F}(q,t)u(t)$ is sent to the far-end side, with $\hat{F}(q,t)$ an estimate of $F(q,t)$.

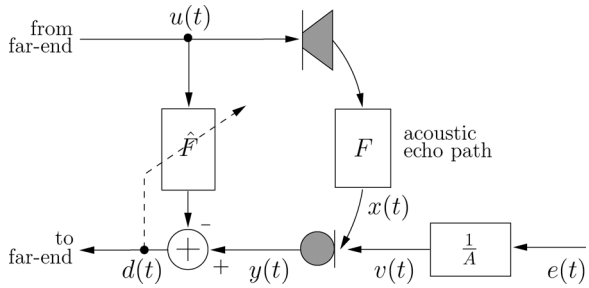


Fig. 1. AEC scenario with autoregressive modeling of the near-end signal.

Since $F(q,t)$ may be time varying (e.g., due to people moving around the room), an adaptive algorithm is usually applied for the estimation of the RIR. It is well known that the convergence speed, and hence the tracking capabilities of standard adaptive algorithms [like recursive least squares (RLS), normalized least mean squares (NLMS), and the affine projection algorithm (APA)] may decrease severely when a near-end signal is present ("double-talk" periods). Therefore, efficient double-talk detectors (DTDs) have been developed, which are used to slow down or switch off the adaptation during double-talk periods. Typically, in a DTD, some decision statistic is calculated using the available signals and signal estimates and compared with a treshold for deciding whether or not a double-talk situation occurs. Since the popular Geigel DTD, which was originally developed for line echo cancellation (LEC) [1], does not always perform satisfactorily in an AEC application, several alternative DTD algorithms have been proposed [2]–[5]. Probably the most widespread DTD for AEC is based on the cross-correlation method described in [2], [3]. Recently also some DTD algorithms have been developed that are specifically suited for subband [6] and stereo AEC [7]. It has been noted, however, that at the onset of a double-talk period, the adaptive algorithm may

Manuscript received October 28, 2005; revised April 18, 2006. The associate editor coordinating the review of this paper and approving it for publication was Prof. Fredrik Gustafsson. This research work was funded by the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen) and carried out at the ESAT laboratory of the Katholieke Universiteit Leuven, in the frame of the Belgian Programme on Interuniversity Attraction Poles, initiated by the Belgian Federal Science Policy Office IUAP P5/22 (''Dynamical Systems and Control: Computation, Identification and Modelling''), the Concerted Research Action GOA-AMBioRICS, and IWT project 040803: ''SMS4PA-II: Sound Management System for Public Address Systems." The scientific responsibility is assumed by its authors.

T. van Waterschoot, G. Rombouts, and M. Moonen are with the Department of Electrical Engineering, ESAT-SCD, Katholieke Universiteit Leuven, B-3001 Leuven, Belgium (e-mail: toon.vanwaterschoot@esat.kuleuven.be; geert.rombouts@esat.kuleuven.be; marc.moonen@esat.kuleuven.be).

P. Verhoeve is with Televic N.V., B-8870 Izegem, Belgium (e-mail: p.verhoeve@televic.com).

Digital Object Identifier 10.1109/TSP.2006.887155

1053-587X/$25.00 © 2007 IEEE

diverge considerably before double-talk is detected. Hence, it is desirable to improve the robustness of the adaptive algorithm to compensate for the detection lag as well as for other DTD imperfections. In [8], the robustness of several adaptive algorithms with regard to so-called near-end noise bursts in LEC is improved by introducing a scaled nonlinearity in the LEC cost function. This method, inspired by the theory of robust statistics [9], is parametrized by a scale factor that governs the tradeoff between robustness and convergence rate. When combined with a cross-correlation-based DTD, the method shows improved performance also in AEC double-talk situations, as compared with the standard adaptive algorithms combined with the same DTD [10, Ch. 6], [11]. Recently, the robust statistics approach has also proven successful in adaptive beamforming applications [12]. However, the increase in robustness comes at the expense of a decreased convergence rate when no double-talk is present. Moreover, the quality of the scale factor estimate depends heavily on the DTD performance.

In a normal telephone conversation, a double-talk situation will occur during approximately 20% of the time [13]. However, in some AEC scenarios, near-end noise will be continuously present and then the use of a DTD becomes futile. This may be the case, for example, in a noisy teleconferencing application, in an automatic gain adjustment system equipped with an echo canceller, or in adaptive feedback cancellation (AFC) [14], [15]. These applications have motivated the design of robust adaptive algorithms that allow for continuous adaptation during double-talk. A first class of these double-talk-robust algorithms is based on the interplay between the echo canceller and a postfilter that is added to suppress the residual echo which remains after cancellation [16]. Such a joint echo canceller and postfilter system may be extended even further with a microphone array [17] or a blind signal separation algorithm [18]. A second approach to double-talk-robust adaptation is to switch to a different adaptation strategy during double-talk. In [19], a so-called maximum-length correlation estimate replaces the stochastic gradient RIR estimate whenever a double-talk situation occurs. In [20] and [21], an adaptive cross-spectral technique is applied instead of a standard adaptive algorithm, and it is claimed to be insensitive to double-talk. Finally, a third method to obtain double-talk-robustness is to remove some correlation in the microphone signal or echo-compensated signal. In [22], an attempt is made to extract the pitch of the residual echo componenent in the echo-compensated signal, whereas in [23], the microphone signal is processed by an adaptive prediction filter to remove the near-end signal correlation during double-talk.

In this paper, we aim at developing a recursive identification algorithm that behaves in a robust way in a continuous double-talk situation (without a DTD) as well as in a bursting double-talk scenario (with, or even without a DTD). In contrast to the method described in [10, Ch. 6], we investigate how robustness may be improved by linear modifications to the standard adaptive algorithms. In our classification of continuously adapting double-talk-robust algorithms, the proposed algorithm fits in the third class described above, since it involves prediction of the near-end signal component in the microphone signal. From linear estimation theory [24], we know that the best (i.e., minimum-variance) linear unbiased estimator (BLUE) for an unknown system depends on the characteristics of the noise acting upon the system. In the AEC context, it is the near-end signal that acts as a noise signal to the RIR identification. Therefore, we expect that by using knowledge of the near-end signal characteristics, the convergence properties of the RIR identification algorithm can be improved. Moreover, in contrast to many existing double-talk-robust algorithms, the increase in robustness using the proposed method does not come at the expense of slower convergence in a single-talk situation. However, the near-end signal characteristics are typically unknown and time varying. Therefore, they have to be estimated concurrently with the unknown RIR. Inspired by prediction error (PE) identification theory [25], we propose a new recursive prediction error (RPE) identification algorithm for concurrent estimation of the RIR and of an autoregressive model of the near-end signal.

The problem of concurrently estimating the RIR and the near-end signal characteristics has also been studied in the context of AFC [14], [15], [26]. It appears that several algorithms proposed for AFC can actually be derived starting from the proposed RPE algorithm. We will explore three algorithms developed for AFC, namely the two-channel adaptive filtering (2ch-AF) algorithm and the prediction-error-method-based adaptive filtering (PEM-AF) algorithm, both developed for hearing aid applications [14], and finally the so-called PEM-AFROW algorithm, proposed in [15] as a PEM-AF variant with only row operations on the loudspeaker data matrix. In contrast to the RPE algorithm, the first two algorithms both rely on the assumption that the near-end signal characteristics do not change significantly on a time scale comparable to the echo path length. As pointed out in [26], this assumption does not hold for room acoustic applications, where the RIR length may be several orders of magnitude larger than the period over which a typical audio signal can be considered stationary. To tackle this particular problem, the PEM-AF algorithm has been modified, resulting in the PEM-AFROW algorithm [15] for room acoustic applications. Whereas in the RPE algorithm, the autoregressive near-end signal model is identified recursively, the near-end signal model identification is performed nonrecursively and in a frame-based manner in the PEM-AFROW algorithm. The frame-based representation is well suited for speech signals, but less appropriate for near-end signals that do not exhibit short-term stationary behavior such as music or babble noise.

The outline of the paper is as follows. We start by defining the data model and reviewing some results from linear estimation theory in Section II, which have motivated the work described in the rest of the paper. In Section III, we describe four PE identification algorithms for concurrent estimation of a RIR and an autoregressive near-end signal model: the new RPE algorithm and the PEM-AF, PEM-AFROW, and 2ch-AF algorithms from adaptive feedback cancellation. Section IV deals with some implementation issues, and we compare the computational complexity of the PE identification algorithms with RLS and NLMS complexity. In Section V, simulation results are shown in which the convergence properties of the proposed algorithms are compared with those of the standard RLS and NLMS algorithms, as well as the double-talk-robust algorithm described in [10, Ch. 6]. Finally, Section VI concludes the paper.

II. Linear Estimation Background

We first review some results from linear estimation theory [24] to indicate how the variance of a RIR estimator can be decreased. Let us assume that a data record $\{u(k), y(k)\}_{k=1}^{t}$ of microphone and loudspeaker samples is available, as well as initial conditions $\{u(k)\}_{k=1-n_F-n_A}^{0}$ and $\{y(k)\}_{k=1-n_A}^{0}$ (with model orders n_F and n_A to be defined later on). For now, we also assume that $F(q,t)$ is time invariant over the observed time interval so no data windowing is applied (we will consider time-varying RIRs when we derive a recursive estimation procedure in Section III). The data model can then be written as

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

where the data matrices are defined as follows:

$$\mathbf{U} = \begin{bmatrix} u(t) & u(t-1) & \dots & u(t-n_F) \\ u(t-1) & u(t-2) & \dots & u(t-n_F-1) \\ \vdots & \vdots & \ddots & \vdots \\ u(1) & u(0) & \dots & u(1-n_F) \end{bmatrix} \tag{2}$$

$$\mathbf{y} = [y(t) \quad y(t-1) \quad \dots \quad y(1)]^T \tag{3}$$

$$\mathbf{v} = [v(t) \quad v(t-1) \quad \dots \quad v(1)]^T \tag{4}$$

and

$$\mathbf{f} = [f_0 \quad f_1 \quad \dots \quad f_{n_F}]^T \tag{5}$$

is the $(n_F+1)\times 1$ parameter vector containing the true coefficients of $F(q,t)$ that have to be estimated. The loudspeaker signal Hankel matrix $\mathbf{U}$ is considered to be deterministic, whereas the near-end signal vector $\mathbf{v}$ and hence also the microphone signal vector $\mathbf{y}$ are random vectors. Note that no additional term is added for representing possible background noise, since we assume that all near-end disturbances are represented together with the useful near-end signal in $\mathbf{v}$.

Typically in AEC, an estimator for the RIR $F(q,t)$ is based on an ordinary least-squares (LS) criterion

$$\min_{\hat{\mathbf{f}}} V_{\mathrm{LS}}(t,\hat{\mathbf{f}}) = \min_{\hat{\mathbf{f}}} (\mathbf{y}-\mathbf{U}\hat{\mathbf{f}})^T(\mathbf{y}-\mathbf{U}\hat{\mathbf{f}}) \tag{6}$$

which leads to the well-known estimator

$$\hat{\mathbf{f}}_{\mathrm{LS}} = (\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{y}. \tag{7}$$

The LS estimator $\hat{\mathbf{f}}_{\mathrm{LS}}$ is unbiased, i.e., $E\{\hat{\mathbf{f}}_{\mathrm{LS}}\} = \mathbf{f}$ (with E the expectation operator), if the near-end signal $v(t)$ is uncorrelated with the loudspeaker signal $u(t)$, i.e., if

$$E\{\mathbf{U}^T\mathbf{v}\} = \mathbf{0}_{(n_F+1)\times 1}. \tag{8}$$

In AEC, this means that the signal $d(t)$ sent to the far-end side is not reflected back to the near-end side due to an acoustic echo path in the far-end room. In the sequel, we will assume (8) to hold. Except when the near-end signal $v(t)$ is a white noise signal, the least-squares estimator (7) is suboptimal. Indeed, a linear unbiased estimator with a smaller variance can be obtained by taking into account the characteristics of the near-end signal.

Minimizing the variance $E(\hat{\mathbf{f}}-E\{\hat{\mathbf{f}}\})(\hat{\mathbf{f}}-E\{\hat{\mathbf{f}}\})^T$ of a linear estimator $\hat{\mathbf{f}}$ under an unbiasedness constraint yields the best linear unbiased estimator (BLUE)

$$\hat{\mathbf{f}}_{\mathrm{BLUE}} = (\mathbf{U}^T\mathbf{R}^{-1}\mathbf{U})^{-1}\mathbf{U}^T\mathbf{R}^{-1}\mathbf{y} \tag{9}$$

which depends on the near-end signal covariance matrix $\mathbf{R}$, defined as

$$\mathbf{R} \triangleq E\left\{(\mathbf{v}-E\{\mathbf{v}\})(\mathbf{v}-E\{\mathbf{v}\})^T\right\} = E\{\mathbf{v}\mathbf{v}^T\} \tag{10}$$

which is usually unknown. Nevertheless, this estimator will be of great importance in the sequel since in Section III-B we will derive a recursive algorithm with which the RIR and the near-end signal characteristics can be estimated concurrently.

III. Prediction Error Identification Algorithms

Obviously, the BLUE as presented in Section II is unrealizable in an AEC scenario because the near-end signal characteristics (and hence $\mathbf{R}^{-1}$) are unknown. Moreover, the estimators from the previous section are nonrecursive and do not allow for tracking of time-varying RIRs. In this section, we will derive a recursive algorithm, which incorporates a RIR estimator that approaches the BLUE, by concurrently estimating the near-end signal characteristics.

We will start this section by establishing a formal link between the BLUE and the PE estimate of the RIR $F(q,t)$. We will then propose a new recursive prediction error (RPE) identification algorithm for concurrent estimation of the RIR and an autoregressive model of the near-end signal. Finally, we will review three PE identification algorithms (the PEM-AF, PEM-AFROW, and 2ch-AF algorithms), which were proposed in an AFC framework, and show how they can be applied to the AEC problem.

A. Prediction Error Criterion

Throughout this paper, we will employ a particular model structure to the system under study, as follows:

$$y(t) = F(q,t)u(t) + \frac{1}{A(q,t)}e(t) \tag{11}$$

with $F(q,t)$ and $A(q,t)$ representing finite-order polynomials in q with possibly time-varying coefficients:

$$F(q,t) = f_0(t) + f_1(t)q^{-1} + \ldots + f_{n_F}(t)q^{-n_F} \tag{12}$$

$$A(q,t) = 1 + a_1(t)q^{-1} + \ldots + a_{n_A}(t)q^{-n_A}. \tag{13}$$

The above choice of model structure will allow us to derive a truly recursive prediction error identification algorithm, which would not have been possible if either the room model were an infinite-impulse-response (IIR) filter or the near-end signal model would contain a moving average (MA) part (see Section III-C-1). The signal $e(t)$ represents the excitation signal generating the near-end signal and is assumed to be a white noise signal with a time-dependent variance σ_t^2:

$$E\{e(t)e(t-k)\} = \sigma_t^2\delta(k). \tag{14}$$

The best one-step-ahead predictor for $y(t)$ is given by [27, Ch. 7]

$$\hat{y}\left[t|\boldsymbol{\theta}(t)\right] = \left[1 - A(q,t)\right] y(t) + A(q,t)F(q,t)u(t) \tag{15}$$

which depends on the unknown coefficients of the polynomials $F(q,t)$ and $A(q,t)$ that are collected in the $(n_F + n_A + 1) \times 1$ parameter vector $\boldsymbol{\theta}(t)$, defined as

$$\boldsymbol{\theta}(t) \triangleq \begin{bmatrix} \mathbf{f}(t) \\ \mathbf{a}(t) \end{bmatrix} \tag{16}$$

with

$$\mathbf{f}(t) \triangleq \begin{bmatrix} f_0(t) & f_1(t) & \dots & f_{n_F}(t) \end{bmatrix}^T \tag{17}$$

$$\mathbf{a}(t) \triangleq \begin{bmatrix} a_1(t) & a_2(t) & \dots & a_{n_A}(t) \end{bmatrix}^T. \tag{18}$$

The prediction error, defined as

$$\varepsilon\left[t,\boldsymbol{\theta}(t)\right] \triangleq y(t) - \hat{y}\left[t|\boldsymbol{\theta}(t)\right] \tag{19}$$

can be calculated with (15) as

$$\varepsilon\left[t,\boldsymbol{\theta}(t)\right] = A(q,t)y(t) - A(q,t)F(q,t)u(t) \tag{20}$$

$$= A(q,t)\left[y(t) - F(q,t)u(t)\right]. \tag{21}$$

Hence, a PE criterion for estimating $\boldsymbol{\theta}(t)$ is given by [27, Ch. 7]

$$\min_{\hat{\boldsymbol{\theta}}(t)} V_{\mathrm{PE}}\left[t,\hat{\boldsymbol{\theta}}(t)\right] = \min_{\hat{\boldsymbol{\theta}}(t)} \frac{1}{2} E\left\{\varepsilon^2\left[t,\hat{\boldsymbol{\theta}}(t)\right]\right\} \tag{22}$$

with $\hat{\boldsymbol{\theta}}(t)$ denoting an estimate of $\boldsymbol{\theta}(t)$. When the expectation is replaced by a summation over the entire time interval on which data are available and appropriate weighting is included, the criterion can be rewritten, using (21), as

$$\min_{\hat{\boldsymbol{\theta}}(t)} \frac{1}{2N} \sum_{k=1}^{t} \frac{\lambda^{t-k}}{\hat{\sigma}_k^2} \left\{\hat{A}(q,t)\left[y(k) - \hat{F}(q,t)u(k)\right]\right\}^2 \tag{23}$$

where $\hat{A}(q,t)$ and $\hat{F}(q,t)$ represent estimates of $A(q,t)$ and $F(q,t)$, respectively, and the weighting factors $\hat{\sigma}_k^{-2}$ and λ^{t-k} are defined as follows. The weighting with the inverse of the estimated PE variance $\hat{\sigma}_k^2$ accounts for energy variations in the near-end excitation signal $e(k)$. On the other hand, an exponential forgetting profile with forgetting factor λ is included to discount old data, and hence to allow for tracking of $F(q,t)$ and $A(q,t)$. The effective window length $N = 1/(1-\lambda)$ is then determined by the forgetting factor λ. The rest of this section will be devoted to the recursive minimization of the PE criterion (23).

A formal relationship between the BLUE and the PE estimate of the room impulse response $F(q,t)$ can be derived only if the near-end signal characteristics are assumed to be known. In [28], it was shown that minimization of the PE criterion (23) with respect to $\hat{\mathbf{f}}(t)$ only, and with $\lambda = 1$, yields the best (minimum-variance) linear unbiased estimate of $\mathbf{f}(t)$ if the near-end signal $v(k)$ can at each time instant $k \in [1,t]$ be modeled as an AR process of order n_A, if the prefilter $\hat{A}(q,k)$ contains at each time instant the true AR coefficients $a_i(k)$, $i = 1,\dots,n_A$, and if the weight $\hat{\sigma}_k^{-2}$ is at each time instant equal to the inverse variance of the near-end excitation signal $e(k)$, i.e., $\hat{\sigma}_k^2 = \sigma_k^2$.

B. Recursive Prediction Error Identification Algorithm

We now derive a RPE identification algorithm for estimating the parameter vector $\boldsymbol{\theta}(t)$ defined in (16), which contains the coefficients of $F(q,t)$ and $A(q,t)$. A general update equation for recursive minimization of the PE criterion in (22) with respect to $\boldsymbol{\theta}(t)$ is given by [25, Ch. 3]

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{N}{\hat{\sigma}_t^2}\gamma(t)\mathbf{R}^{-1}(t)\boldsymbol{\psi}\left[t,\hat{\boldsymbol{\theta}}(t-1)\right]\varepsilon\left[t,\hat{\boldsymbol{\theta}}(t-1)\right] \tag{24}$$

where the terms are defined as follows. In the update term on the right-hand side of (24), $\mathbf{R}(t)$ should be an approximation of the Hessian $\partial^2/\partial\boldsymbol{\theta}^2(t)V_{\mathrm{PE}}[t,\boldsymbol{\theta}(t)]$ of the criterion (22), evaluated in $\boldsymbol{\theta}(t) = \hat{\boldsymbol{\theta}}(t-1)$. The $(n_F + n_A + 1) \times 1$ gradient vector $\boldsymbol{\psi}[t,\boldsymbol{\theta}(t)]$ is defined as

$$\boldsymbol{\psi}\left[t,\boldsymbol{\theta}(t)\right] \triangleq -\frac{\partial}{\partial\boldsymbol{\theta}(t)}\varepsilon\left[t,\boldsymbol{\theta}(t)\right] = \frac{\partial}{\partial\boldsymbol{\theta}(t)}\hat{y}\left[t|\boldsymbol{\theta}(t)\right] \tag{25}$$

where the second equality is obtained by use of (19), in which the term $y(t)$ is independent of $\boldsymbol{\theta}(t)$. The gain sequence $\gamma(t)$ depends on the type of data windowing that is applied. In the sequel, we will choose $\gamma(t) = 1/N$ to compensate for the factor N in (24), consistent with the definition of $\mathbf{R}(t)$ in (32) below.

As pointed out in [26], it is highly desirable to identify $F(q,t)$ and $A(q,t)$ on data windows of different sizes. The RIR $F(q,t)$ will typically be a high-order polynomial requiring a large data window for accurate estimation. On the other hand, the near-end signal AR polynomial $A(q,t)$ may be highly time varying; hence, a small data window is desirable for estimation of the AR coefficients. It will be shown below that such a decoupling of the data windows for estimating $F(q,t)$ and $A(q,t)$ can be achieved by block-diagonalizing the matrix $\mathbf{R}(t)$. First, however, we explore the calculation of the gradient vector $\boldsymbol{\psi}[t,\boldsymbol{\theta}(t)]$.

1) Calculation of Gradient Vector $\boldsymbol{\psi}[t,\boldsymbol{\theta}(t)]$: The gradient vector $\boldsymbol{\psi}[t,\hat{\boldsymbol{\theta}}(t-1)]$, as it appears in (24), can be calculated by differentiating the one-step-ahead predictor (15) with respect to the coefficients of $F(q,t)$ and $A(q,t)$, and replacing $A(q,t)$ and $F(q,t)$ by their estimates $\hat{A}(q,t-1)$ and $\hat{F}(q,t-1)$ at time $t-1$, i.e.,

$$\boldsymbol{\psi}\left[t,\hat{\boldsymbol{\theta}}(t-1)\right] = \begin{bmatrix} \hat{A}(q,t-1)u(t) \\ \vdots \\ \hat{A}(q,t-1)u(t-n_F) \\ \hat{F}(q,t-1)u(t-1) - y(t-1) \\ \vdots \\ \hat{F}(q,t-1)u(t-n_A) - y(t-n_A) \end{bmatrix} \tag{26}$$

or, in partitioned matrix notation

$$\boldsymbol{\psi}\left[t,\hat{\boldsymbol{\theta}}(t-1)\right] \triangleq \begin{bmatrix} \boldsymbol{\psi}_F\left[t,\hat{\mathbf{a}}(t-1)\right] \\ \boldsymbol{\psi}_A\left[t,\hat{\mathbf{f}}(t-1)\right] \end{bmatrix} \tag{27}$$

$$= \begin{bmatrix} \mathbf{0}_{(n_F+1)\times(n_F+1)} & \mathbf{U}(t) \\ \mathbf{U}^T(t) & \mathbf{0}_{n_A\times n_A} \end{bmatrix} \times \begin{bmatrix} \hat{\mathbf{f}}(t-1) \\ \hat{\mathbf{a}}(t-1) \end{bmatrix} + \begin{bmatrix} \mathbf{u}(t) \\ -\mathbf{y}(t) \end{bmatrix} \quad (28)$$

with

$$\mathbf{y}(t) \triangleq [y(t-1) \quad \dots \quad y(t-n_A)]^T_{n_A\times 1} \quad (29)$$

$$\mathbf{u}(t) \triangleq [u(t) \quad \dots \quad u(t-n_F)]^T_{(n_F+1)\times 1} \quad (30)$$

$$\mathbf{U}(t) \triangleq \begin{bmatrix} u(t-1) & \dots & u(t-n_A) \\ \vdots & \ddots & \vdots \\ u(t-n_F-1) & \dots & u(t-n_F-n_A) \end{bmatrix}_{(n_F+1)\times n_A} \quad (31)$$

and $\hat{\mathbf{f}}(t-1)$ and $\hat{\mathbf{a}}(t-1)$ denoting the most recent estimates of parameter vectors $\mathbf{f}(t)$ and $\mathbf{a}(t)$ defined in (17) and (18), respectively.

2) Block-Diagonal Approximate Gauss–Newton Update of $\mathbf{R}(t)$*:* The matrix $\mathbf{R}(t)$ determines the search direction of the recursive identification algorithm (24). It should be chosen positive definite and ideally, i.e., to obtain a Newton update direction, equal to the Hessian of the PE criterion (22), evaluated at the most recent parameter vector estimate. A common approximation is to retain only the term made up of gradient vector outer products in the calculation of the Hessian, which results in a Gauss–Newton update instead of a Newton update [25, Ch. 3]. Defining

$$\mathbf{R}(t) \triangleq N \left\{ \frac{\partial^2}{\partial \boldsymbol{\theta}^2(t)} V_{\mathrm{PE}}[t, \boldsymbol{\theta}(t)] \right\}\Big|_{\boldsymbol{\theta}(t)=\hat{\boldsymbol{\theta}}(t-1)} \quad (32)$$

a Gauss–Newton update equation for $\mathbf{R}(t)$ can be obtained

$$\mathbf{R}(t) = \lambda \mathbf{R}(t-1) + \left\{ \frac{1}{\hat{\sigma}_t^2} \boldsymbol{\psi}[t, \boldsymbol{\theta}(t)] \boldsymbol{\psi}^T[t, \boldsymbol{\theta}(t)] \right\}\Big|_{\boldsymbol{\theta}(t)=\hat{\boldsymbol{\theta}}(t-1)}. \quad (33)$$

As suggested in [14], decoupling of the data windows on which $\mathbf{f}(t)$ and $\mathbf{a}(t)$ are estimated can be achieved by forcing the off-diagonal blocks of $\mathbf{R}(t)$ to zero and introducing two different exponential forgetting factors λ_F and λ_A. The resulting update equation is shown in (34) at the bottom of the page.

When applied to the update (24), this block-diagonalization of $\mathbf{R}(t)$ results in an approximate Gauss–Newton update direction for $\hat{\boldsymbol{\theta}}(t)$. The block-diagonalization was also suggested earlier for the recursive generalized least-squares (RGLS) algorithm described in [29] and [30], which is a recursive implementation of the GLS estimator proposed in [31]. Setting the off-diagonal blocks to zero may also be justified by noting that if $\hat{\mathbf{f}}(t-1)$ and $\hat{\mathbf{a}}(t-1)$ have converged to their true values $\mathbf{f}(t)$ and $\mathbf{a}(t)$, respectively, then

$$E\left\{ \boldsymbol{\psi}_F[t, \mathbf{a}(t)] \boldsymbol{\psi}_A^T[t, \mathbf{f}(t)] \right\} \quad (35)$$

$$= E\left\{ \boldsymbol{\psi}_F[t, \mathbf{a}(t)] [v(t-1) \quad \dots \quad v(t-n_A)] \right\} \quad (36)$$

$$= \mathbf{0}_{(n_F+1)\times n_A} \quad (37)$$

since $\boldsymbol{\psi}_F[t, \mathbf{a}(t)]$ depends only on the loudspeaker signal $u(t)$, which was assumed to be uncorrelated with the near-end signal $v(t)$ in (8), and

$$E\{\varepsilon[t, \boldsymbol{\theta}(t)] \mathbf{U}(t)\} = E\{e(t)\mathbf{U}(t)\} = \mathbf{0}_{(n_F+1)\times n_A} \quad (38)$$

since $e(t)$ is a white noise signal uncorrelated with the loudspeaker signal $u(t)$.

Finally, note that the block-diagonal approximation (34) of $\mathbf{R}(t)$ is positive semidefinite by construction and will hence always lead to a recursive estimator $\hat{\boldsymbol{\theta}}(t)$ that converges to a local minimum of the PE criterion (23), but not necessarily to the global minimum. The proposed RPE identification algorithm is summarized in Table I, making use of (21), (24), (28), and (34), and explicitly decoupling all expressions for estimation of $\mathbf{f}(t)$ and $\mathbf{a}(t)$. Note that the PE variance $\hat{\sigma}_t^2$ is estimated by means of a single update equation and on the same data window as the AR polynomial $A(q,t)$, since both depend on the near-end signal characteristics.

C. Prediction Error Identification Algorithms From AFC

1) Prediction-Error-Method-Based Adaptive Filtering Algorithm: A common approximation in recursive system identification [25, Ch. 3] consists of replacing the estimates $\hat{A}(q,t-1)$ and $\hat{F}(q,t-1)$ in (26) by previously obtained estimates, e.g.,

$$\boldsymbol{\psi}(t) = \begin{bmatrix} \hat{A}(q,t-1)u(t) \\ \vdots \\ \hat{A}(q,t-1-n_F)u(t-n_F) \\ \hat{F}(q,t-2)u(t-1) - y(t-1) \\ \vdots \\ \hat{F}(q,t-1-n_A)u(t-n_A) - y(t-n_A) \end{bmatrix}. \quad (39)$$

The motivation behind this approximation is that in a generic recursive identification scenario, the system model is an IIR filter, and the noise model contains a MA part:

$$y(t) = \frac{F(q,t)}{D(q,t)} u(t) + \frac{C(q,t)}{A(q,t)} e(t). \quad (40)$$

If either one of the polynomials $D(q,t)$ and $C(q,t)$ is included in the model structure, the one-step-ahead predictor $\hat{y}[t|\boldsymbol{\theta}(t)]$,

$$\begin{bmatrix} \mathbf{R}_F(t) & \mathbf{0}_{(n_F+1)\times n_A} \\ \mathbf{0}_{n_A\times(n_F+1)} & \mathbf{R}_A(t) \end{bmatrix} = \begin{bmatrix} \lambda_F \mathbf{R}_F(t-1) & \mathbf{0}_{(n_F+1)\times n_A} \\ \mathbf{0}_{n_A\times(n_F+1)} & \lambda_A \mathbf{R}_A(t-1) \end{bmatrix} + \frac{1}{\hat{\sigma}_t^2} \begin{bmatrix} \boldsymbol{\psi}_F[t, \hat{\mathbf{a}}(t-1)] \boldsymbol{\psi}_F^T[t, \hat{\mathbf{a}}(t-1)] & \mathbf{0}_{(n_F+1)\times n_A} \\ \mathbf{0}_{n_A\times(n_F+1)} & \boldsymbol{\psi}_A\left[t, \hat{\mathbf{f}}(t-1)\right] \boldsymbol{\psi}_A^T\left[t, \hat{\mathbf{f}}(t-1)\right] \end{bmatrix}. \quad (34)$$

TABLE I
SUMMARY OF THE RPE ALGORITHM

RPE - Recursive prediction error identification algorithm

$$\hat{\mathbf{f}}(t) = \hat{\mathbf{f}}(t-1) + \frac{1}{\hat{\sigma}_t^2}\mathbf{R}_F^{-1}(t)\boldsymbol{\psi}_F[t,\hat{\mathbf{a}}(t-1)]\varepsilon[t,\hat{\boldsymbol{\theta}}(t-1)]$$

$$\hat{\mathbf{a}}(t) = \hat{\mathbf{a}}(t-1) + \frac{1}{\hat{\sigma}_t^2}\mathbf{R}_A^{-1}(t)\boldsymbol{\psi}_A[t,\hat{\mathbf{f}}(t-1)]\varepsilon[t,\hat{\boldsymbol{\theta}}(t-1)]$$

$$\varepsilon[t,\hat{\boldsymbol{\theta}}(t-1)] = \begin{bmatrix}1 & \hat{\mathbf{a}}^T(t-1)\end{bmatrix}\left(\begin{bmatrix}y(t)\\ \mathbf{y}(t)\end{bmatrix} - \begin{bmatrix}\mathbf{u}^T(t)\\ \mathbf{U}^T(t)\end{bmatrix}\hat{\mathbf{f}}(t-1)\right)$$

$$\hat{\sigma}_t^2 = \lambda_A\hat{\sigma}_{t-1}^2 + (1-\lambda_A)\varepsilon^2[t,\hat{\boldsymbol{\theta}}(t-1)]$$

$$\boldsymbol{\psi}_F[t,\hat{\mathbf{a}}(t-1)] = \begin{bmatrix}\mathbf{u}(t) & \mathbf{U}(t)\end{bmatrix}\begin{bmatrix}1\\ \hat{\mathbf{a}}(t-1)\end{bmatrix}$$

$$\boldsymbol{\psi}_A[t,\hat{\mathbf{f}}(t-1)] = \mathbf{U}^T(t)\hat{\mathbf{f}}(t-1) - \mathbf{y}(t)$$

$$\mathbf{R}_F(t) = \lambda_F\mathbf{R}_F(t-1) + \frac{1}{\hat{\sigma}_t^2}\boldsymbol{\psi}_F[t,\hat{\mathbf{a}}(t-1)]\boldsymbol{\psi}_F^T[t,\hat{\mathbf{a}}(t-1)]$$

$$\mathbf{R}_A(t) = \lambda_A\mathbf{R}_A(t-1) + \frac{1}{\hat{\sigma}_t^2}\boldsymbol{\psi}_A[t,\hat{\mathbf{f}}(t-1)]\boldsymbol{\psi}_A^T[t,\hat{\mathbf{f}}(t-1)]$$

Refer to (16), (17), (18), (29), (30), and (31) for definitions of the parameter vectors and data matrices.

and hence the gradient vector $\boldsymbol{\psi}[t,\boldsymbol{\theta}(t)]$, will be made up of data filtered by IIR filters. An exact calculation of $\boldsymbol{\psi}[t,\hat{\boldsymbol{\theta}}(t-1)]$ as in (26) would then require the complete data record to be filtered with the most recent estimates $\hat{D}(q,t-1)$ and/or $\hat{C}(q,t-1)$, and hence would not allow for a truly recursive identification algorithm. However, this problem does not occur for the model structure (11) under study, as pointed out earlier in [32].

Nevertheless, even with the model structure (11), the approximation in (39) is appealing from a complexity point of view. As pointed out in [14], using (39) instead of (26) saves n_F+n_A-1 filtering operations in each recursion step, leading to a cheaper, so-called PEM-AF algorithm. However, this approximation is fully justified only when

$$\begin{cases}\hat{A}(q,t-1) = \hat{A}(q,t-1-i), & i\in[0,n_F] \quad (41)\\ \hat{F}(q,t-1) = \hat{F}(q,t-1-j), & j\in[1,n_A]. \quad (42)\end{cases}$$

As mentioned before, in an AEC scenario, $F(q,t)$ will typically be of very high order and so $A(q,t)$ may be highly time varying over a data window of length n_F. Hence, the assumption in (41) will not hold for the current application, in contrast to the hearing aid AFC application considered in [14]. Also, in the case of a time-varying RIR, the assumption in (42) may not hold.

2) Prediction-Error-Method-Based Adaptive Filtering Algorithm With Row Operations: Instead of identifying the near-end signal AR coefficients recursively, as in the RPE and PEM-AF algorithms described above, the AR model can also be identified non-recursively on a batch of loudspeaker and microphone data. This may be interesting from a computational point of view if the near-end signal exhibits short-term stationary behavior, because then the AR model does not have to be recalculated at each time instant. This is the idea behind the PEM-AFROW algorithm, which was originally developed in an AFC framework [15] and applied to a continuous double-talk AEC scenario in [28]. This algorithm, like the RPE algorithm, performs only row operations on the loudspeaker data matrix, hence the name PEM-AFROW. The PEM-AFROW algorithm is best suited for applications involving a near-end speech signal, since speech exhibits short-term stationary behavior.

3) Two-Channel Adaptive Filtering Algorithm: All three PE identification algorithms described above aim at direct minimization of the criterion in (23), which is nonlinear in the coefficients of $F(q,t)$ and $A(q,t)$; hence they may suffer from convergence to local minima. In the 2ch-AF algorithm proposed in [14] it is noted that expression (20) for the PE can be written as

$$\varepsilon\left[t,\boldsymbol{\theta}_{2\mathrm{ch}}(t)\right] = A(q,t)y(t) + B(q,t)u(t) \quad (43)$$

with

$$B(q,t) \triangleq -A(q,t)F(q,t) \quad (44)$$

and

$$\boldsymbol{\theta}_{2\mathrm{ch}}(t) \triangleq \begin{bmatrix}b_0(t) & \dots & b_{n_B}(t) & a_1(t) & \dots & a_{n_A}(t)\end{bmatrix}^T \quad (45)$$

with $n_B = n_A + n_F$. This interpretation leads to a criterion that is quadratic in the parameter vector $\boldsymbol{\theta}_{2\mathrm{ch}}(t)$, so that an ordinary RLS algorithm can be applied.

The main drawback of the 2ch-AF algorithm is that the coefficients of $A(q,t)$ and $F(q,t)$, which are combined into $B(q,t)$, are estimated on the same data window. As pointed out in [26] for the AFC case, this makes the algorithm, which was originally derived for hearing aid applications, unsuitable when dealing with room acoustics. Another drawback is that $F(q,t)$ can only be obtained by deconvolution of $A(q,t)$ and $B(q,t)$, which adds a complexity of $O(n_F^2)$ to the algorithm. Finally, we note that the interpretation of the PE as in (43) and (44) was made earlier in [33] as an alternative to the generalized least-squares (GLS) algorithm proposed in [31].

IV. Implementation and Complexity

A. Implementation Issues

The RPE, PEM-AF, and PEM-AFROW algorithms described in Section III are implemented both with a Gauss–Newton update and a cheaper stochastic gradient update of the RIR estimate $\hat{\mathbf{f}}(t)$. The same is done for the update of the coefficients of $B(q,t)$ in the 2ch-AF algorithm. The estimate $\hat{\mathbf{a}}(t)$ is always updated according to the Gauss–Newton method in the RPE, PEM-AF, and 2ch-AF algorithms. In the PEM-AFROW algorithm, $\hat{\mathbf{a}}(t)$ and $\hat{\sigma}_t^2$ are calculated using the Levinson–Durbin recursion [34, Ch. 6].

1) Gauss–Newton Algorithms: For computational efficiency, all Gauss–Newton updates are performed by using the Kalman gain vectors and applying the matrix inversion lemma [34, Ch. 13]. Typically, the inverse correlation matrix will be initialized as $\mathbf{R}_F^{-1}(0) = 10^6 \cdot \mathbf{I}_{n_F+1}$ and the RIR estimate as $\hat{\mathbf{f}}(0) = \mathbf{0}_{(n_F+1)\times 1}$.

2) Stochastic Gradient Algorithms: The stochastic gradient implementation of the algorithms features a normalization with estimates of both the loudspeaker signal variance (as in the NLMS algorithm) and the near-end signal variance (as in the algorithms described in Section III). Simulations have shown that the sum method proposed in [35] for combining these two normalization factors yields the best convergence behavior. As an example, a stochastic gradient implementation of the RPE update equations in Table I can be formulated as

$$\hat{\mathbf{f}}(t) = \hat{\mathbf{f}}(t-1) + \mu_F \times \frac{\boldsymbol{\psi}_F\left[t,\hat{\mathbf{a}}(t-1)\right]\varepsilon\left[t,\hat{\boldsymbol{\theta}}(t-1)\right]}{\delta_F + (n_F+1)\hat{\sigma}_t^2 + \boldsymbol{\psi}_F^T\left[t,\hat{\mathbf{a}}(t-1)\right]\boldsymbol{\psi}_F\left[t,\hat{\mathbf{a}}(t-1)\right]} \quad (46)$$

where μ_F denotes the step size and δ_F the regularization parameter. Note that the weighted sum method proposed in [35] was applied in [36] to improve double-talk-robustness in an acoustic echo canceller.

3) Choice of Near-End Model Order: Simulations have shown that the choice of the near-end model order n_A has a profound influence on the performance of the PE identification algorithms described in Section III. The algorithms were derived under a sufficient-order assumption, yet in practice this assumption will not be fulfilled. Due to the non-stationary nature of speech and audio signals, the true model order is not well defined, let alone that it would be known to the user. Moreover, the larger n_A is chosen, the larger the number of local minima in the PE cost function (23).

We have evaluated four different choices of n_A by computer simulations at $f_s = 8$ kHz: a high-order model ($n_A = 55$) to capture all near-end signal dynamics, a low-order model ($n_A = 12$) as used for formant prediction in speech coding, a low-order model ($n_A = 12$) cascaded with a 3-tap pitch predictor as used in speech coding [37], and a very-low-order model ($n_A = 1$) as used in the AFC algorithm proposed in [38]. It appears that the very-low-order model ($n_A = 1$) yields the best performance in all scenarios that were tested. In other words, even if the near-end signal component $v(t)$ contained in the microphone signal $y(t)$ is undermodeled, convergence speed is gained due to the more convex shape of the PE cost function when n_A is small.

B. Computational Complexity

To compare the computational complexity of the PE identification algorithms described in Section III with the complexity of standard RLS and NLMS adaptive algorithms, we show the number of floating point operations (FLOPS) per recursion step in Table II. The complexity equations in Table II are evaluated with the values of the parameters used in the simulations described in Section V: $L_F = n_F + 1 = 1000$, $n_A = 1$, $P = 160$ and $M = P + n_A = 161$, with M the frame length and P the frame hop size of the PEM-AFROW algorithm.

Note that the Gauss–Newton algorithms are of comparable complexity if n_A is small, except for the 2ch-AF algorithm, which requires additional computations for the deconvolution of the filters $A(q,t)$ and $B(q,t)$. The stochastic gradient PEM-AF algorithm has a complexity comparable to the NLMS algorithm, whereas the RPE and PEM-AFROW algorithms require somewhat more computations due to the exact calculation of the PE and gradient vectors (in contrast to the approximate calculation in the PEM-AF algorithm). The stochastic gradient 2ch-AF algorithm is computationally not interesting due to the deconvolution with complexity $O(L_F^2)$. Since no assumptions on the short-term stationarity of the near-end signal are made in the RPE algorithm, its complexity rises somewhat faster than that of the PEM-AF and PEM-AFROW algorithms as n_A is increased.

V. Simulation Results

MATLAB simulations were performed to compare the convergence behavior of the four PE identification algorithms described in Section III with some reference algorithms in different scenarios. The Gauss-Newton reference algorithms are the standard RLS algorithm and the double-talk-robust nonlinear recursive least squares (NL-RLS) algorithm proposed in [10, Ch. 6]. The stochastic gradient reference algorithms are the standard NLMS algorithm and a stochastic gradient implementation of the NL-RLS algorithm. It should be noted that the NL-RLS algorithm was designed to be used in combination with a DTD, in contrast to the PE identification algorithms. To focus on the double-talk robustness of the adaptive algorithms, we have chosen not to include a DTD in any of the simulations because the results would then depend heavily on the DTD quality.

All simulations are performed at a sampling rate $f_s = 8$ kHz. The room impulse response $F(q,t)$ of length

TABLE II
COMPLEXITY COMPARISON BY NUMBER OF FLOPS PER RECURSION

Algorithm	Number of FLOPS per recursion	
Gauss-Newton algorithms		
RLS	$5L_F^2 + 5L_F$	5005000
RPE	$5L_F^2 + (4n_A + 5)L_F + 5n_A^2 + 5n_A + 6$	5009016
PEM-AF	$5L_F^2 + 5L_F + 5n_A^2 + 7n_A + 6$	5005018
PEM-AFROW	$5L_F^2 + \left(5 + \frac{2M + 2n_A + 1}{P}\right)L_F + \frac{1}{P}n_A^2 + \left(4 + \frac{2M + 2}{P}\right)n_A + \frac{2M}{P} + \frac{P-1}{P} + 2$	5007000
2ch-AF	$6L_F^2 + (20n_A + 4)L_F + 20n_A^2 + 10n_A + 5$	6024035
Stochastic gradient algorithms		
NLMS	$6L_F + 2$	6002
RPE	$(4n_A + 6)L_F + 5n_A^2 + 5n_A + 10$	10020
PEM-AF	$6L_F + 5n_A^2 + 7n_A + 10$	6022
PEM-AFROW	$\left(6 + \frac{2M + 2n_A + 1}{P}\right)L_F + \frac{1}{P}n_A^2 + \left(4 + \frac{2M + 2}{P}\right)n_A + \frac{2M}{P} + \frac{P-1}{P} + 6$	8046
2ch-AF	$L_F^2 + 5L_F + 5n_A^2 + 11n_A + 10$	1005026

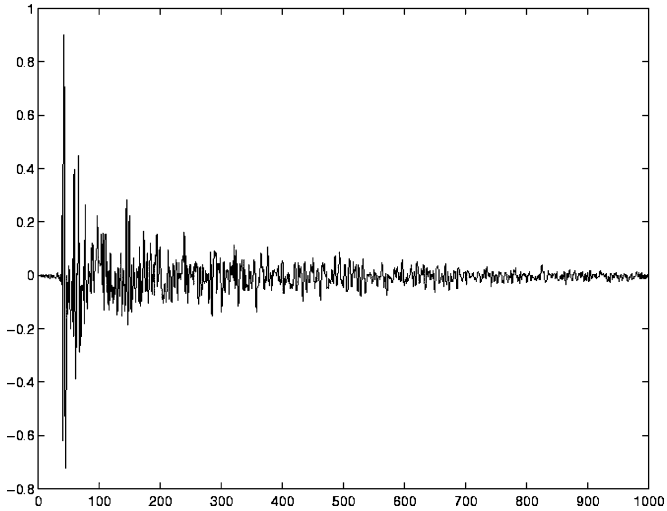


Fig. 2. Room impulse response measured in a typical office.

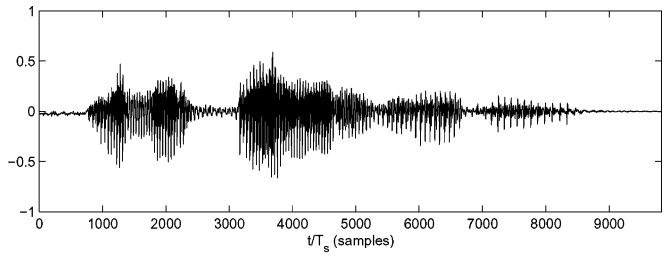
$L_F = n_F + 1 = 1000$ (corresponding to 125 ms) was measured in a typical office and is plotted in Fig. 2. The microphone and loudspeaker are assumed to be linear devices with a flat frequency response. In all simulations, some low-level (-50 dB) Gaussian white noise is added to the near-end signal to account for background noise, which is present in any realistic setup. The near-end signal model order is set to $n_A = 1$ in most of the experiments, as motivated in Section IV-A-3. The exponential forgetting factors for estimation of $\mathbf{f}(t)$ in the Gauss–Newton RPE, PEM-AF, and PEM-AFROW algorithms, and of $\boldsymbol{\theta}_{2\text{ch}}(t)$ in the 2ch-AF algorithm, are chosen such that the effective data window length is three times the parameter vector length, i.e., $\lambda_F = \lambda_{2\text{ch}} = 0.9997$. The exponential forgetting factor for recursive estimation of $\mathbf{a}(t)$ in the RPE, PEM-AF, and stochastic gradient 2ch-AF algorithms, and of σ_t^2 in the RPE, PEM-AF, and 2ch-AF algorithms is $\lambda_A = 0.9938$, such that the effective data window length is 160 samples or 20 ms (which is the average window length in which a speech signal can be considered stationary). The PEM-AFROW window hop size is set to $P = 160$ samples and the window length to $M = P + n_A$, as suggested in [15]. The step size for all stochastic gradient algorithms is equal to $\mu_F = 0.5$ (which was found to give the best convergence behavior on average) and the regularization parameter is $\delta_F = 10^{-6}$. In all the different scenarios, the near-end signal is scaled such that the echo-to-near-end ratio (ENR) equals 10 dB:

$$\text{ENR} \triangleq 10\log_{10}\frac{\sum_{k=1}^{N}|x(k)|^2}{\sum_{k=1}^{N}|v(k)|^2} = 10\text{ dB} \tag{47}$$

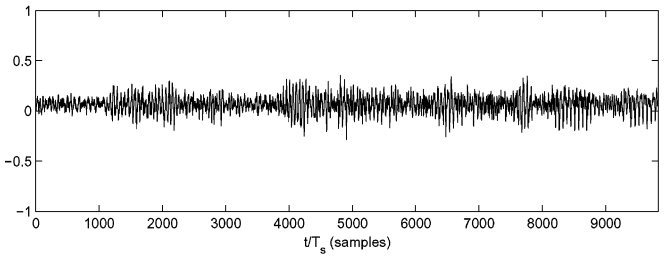
with N the number of data points used. The simulation length N varies for different simulations, and will typically be much smaller for the Gauss–Newton simulations than for the stochastic gradient simulations. The parameters of the NL-RLS algorithm are chosen as suggested in [10, Ch. 6]: the saturation level is $k_0 = 1.5$ and the exponential forgetting factor for estimation of the scale factor is $\lambda_s = 0.9975$. All algorithms, impulse responses, and sound signals are available for download [39].

The performance measure for evaluating the convergence behavior is the misalignment, defined as the logarithmic normalized Euclidian distance between the true and estimated impulse response at each time instant:

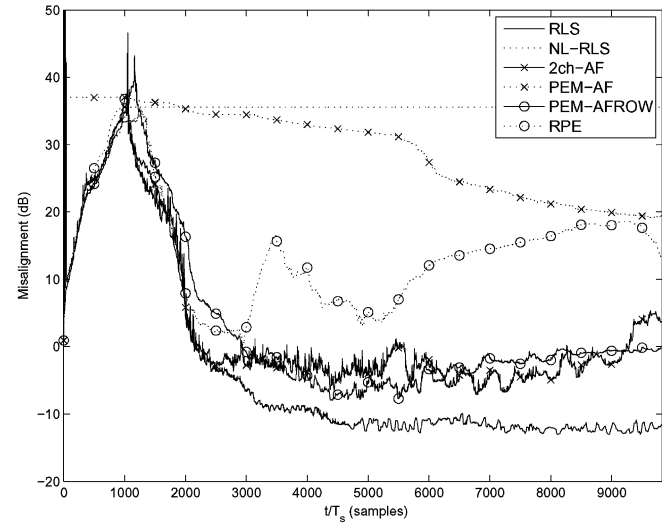
$$\text{misalignment (dB)} = 20\log_{10}\frac{\left\|\hat{\mathbf{f}}(t) - \mathbf{f}(t)\right\|}{\|\mathbf{f}(t)\|}. \tag{48}$$



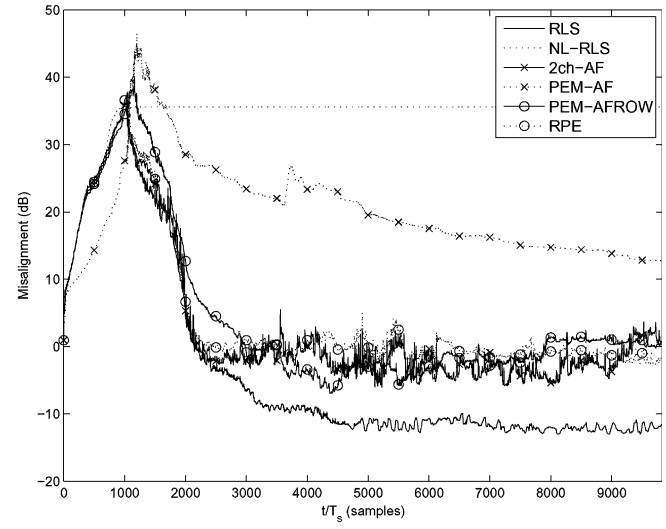
(a)



(b)

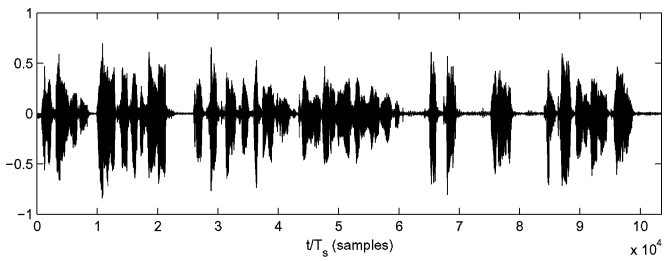


(c)

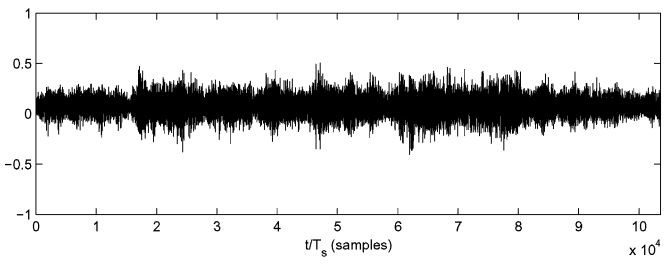


(d)

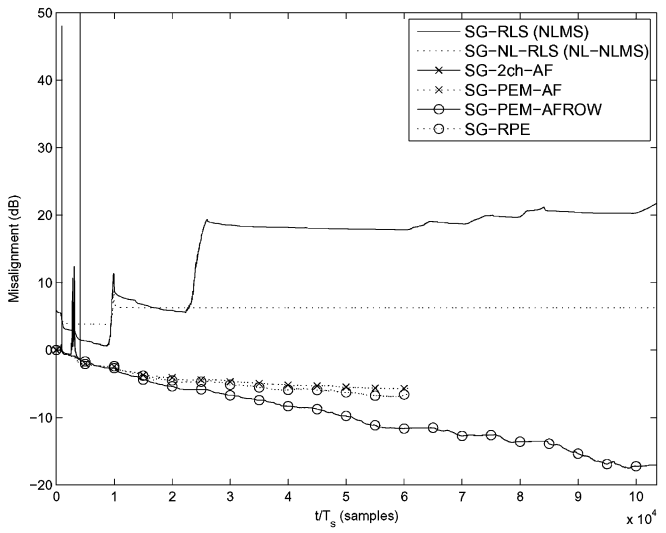
Fig. 3. Continuous double-talk scenario: Gauss–Newton simulations. (a) Far-end speech signal. (b) Near-end babble noise signal. (c) Convergence curves for $n_A = 12$. (d) Convergence curves for $n_A = 1$.



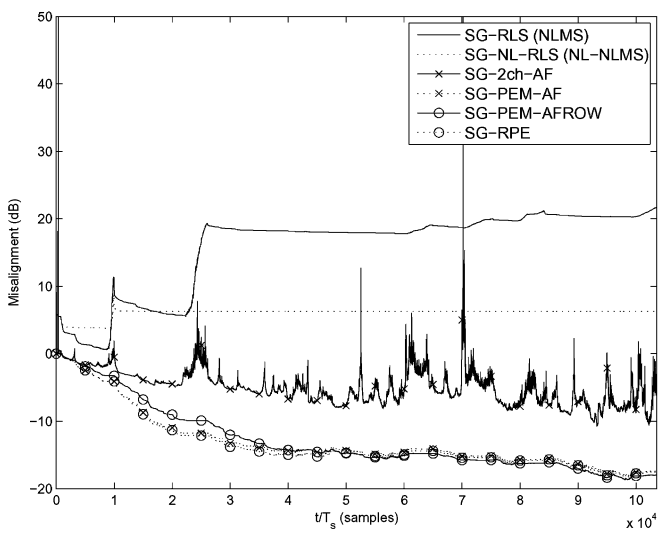
(a)



(b)



(c)



(d)

Fig. 4. Continuous double-talk scenario: stochastic gradient simulations. (a) Far-end speech signal. (b) Near-end babble noise signal. (c) Convergence curves for $n_A = 12$. (d) Convergence curves for $n_A = 1$.

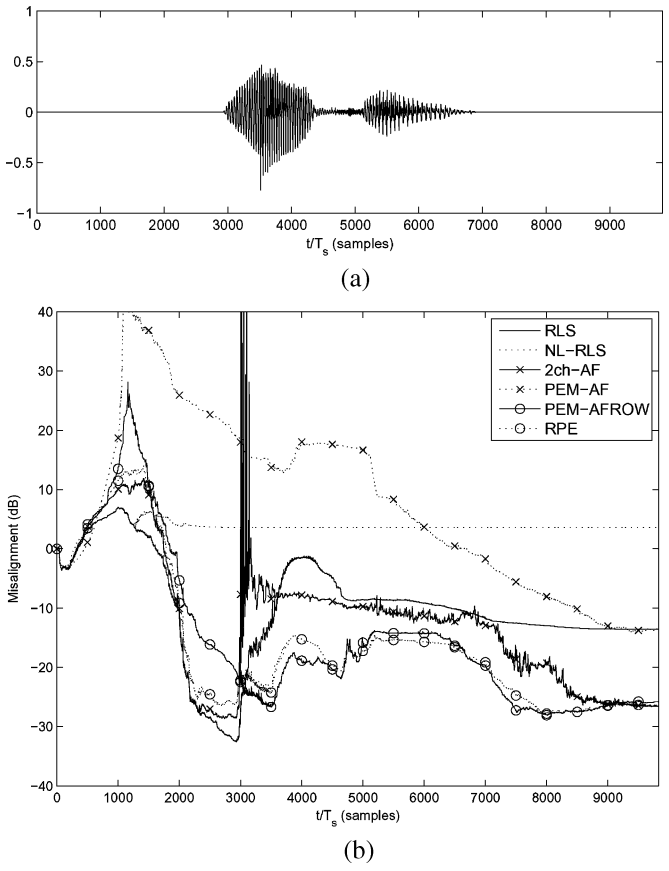


Fig. 5. Bursting double-talk scenario: Gauss–Newton simulations. (a) Near-end speech signal. (b) Convergence curves for $n_A = 1$.

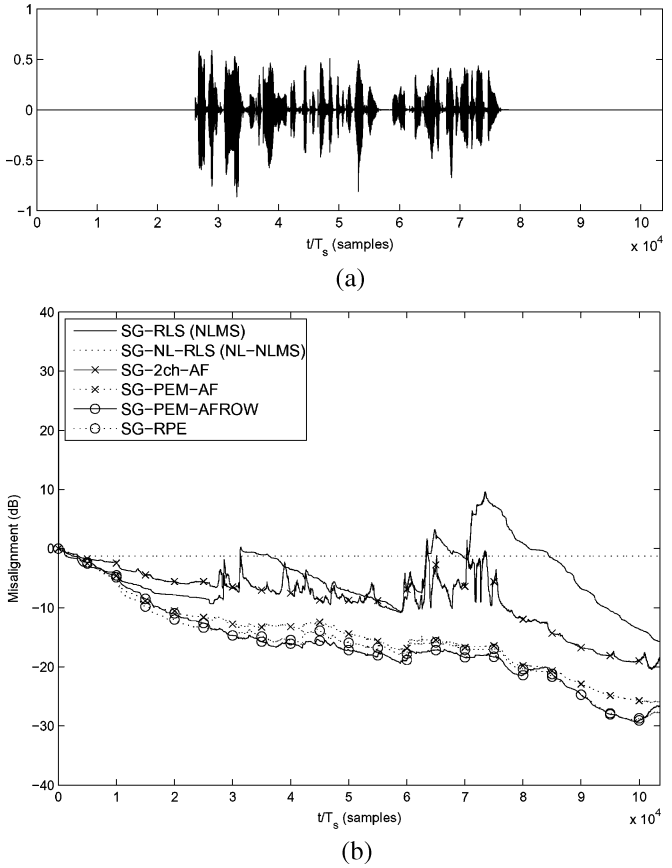


Fig. 6. Bursting double-talk scenario: stochastic gradient simulations. (a) Near-end speech signal. (b) Convergence curves for $n_A = 1$.

In a first scenario, double-talk is continuously present. The far-end signal is a male speech signal plotted in Figs. 3(a) and 4(a) for the Gauss–Newton and stochastic gradient experiments, respectively. Note that the signals are drawn from the same speech fragment, but on different time scales. With the continuous double-talk applications described in Section I in mind, the near-end signal is a babble noise signal plotted in Figs. 3(b) and 4(b). Continuous double-talk simulations with a near-end speech signal are described in [28] and [40]. The Gauss–Newton and stochastic gradient convergence curves are drawn in Figs. 3(c) and 4(c), respectively, for a low-order near-end signal model $(n_A = 12)$ and in Figs. 3(d) and 4(d), respectively, for a very-low-order near-end signal model $(n_A = 1)$. A first, somewhat surprising, observation is that the standard RLS algorithm performs in a relatively robust way in a continuous double-talk situation: the RLS algorithm outperforms all other algorithms. The standard NLMS algorithm, on the other hand, appears to be extremely sensitive to double-talk for $\mu = 0.5$, in contrast to the stochastic gradient RPE, PEM-AF, and PEM-AFROW algorithms. All PE identification algorithms exhibit improved convergence and more stable behavior with the very-low-order near-end signal model $(n_A = 1)$. Finally, in all simulations, the NL-RLS algorithms stalls at an estimate far from the true impulse response. This is because the NL-RLS scale factor converges to zero during double-talk, thereby slowing down and eventually freezing the adaptation. This effect can be overcome by limiting the scale factor to a minimum value during double-talk, which is, however, impossible when no DTD is applied.

In a second scenario, double-talk only occurs after the adaptive algorithms have initially converged to some extent. The same male far-end speech signals as in the first scenario are used, whereas the near-end signal is now a short female speech signal, occuring in the middle of the experiment, as shown in Figs. 5(a) and 6(a). In this scenario, we only use a very-low-order near-end signal model $(n_A = 1)$, which yields the convergence curves in Figs. 5(b) and 6(b). The Gauss–Newton RPE, PEM-AFROW, and 2ch-AF algorithms show approximately the same initial convergence during single-talk as the standard RLS algorithm, while the Gauss–Newton PEM-AF and NL-RLS algorithms perform worse. At the onset of double-talk (around $t/T_s = 3000$ samples), the 2ch-AF and standard RLS algorithms diverge considerably. The PEM-AFROW and RPE algorithms behave remarkably robust, although they show some divergence when the near-end signal power increases. The standard RLS algorithm clearly reacts in a less robust way to a sudden near-end signal onset than to the continuous double-talk situation in the first scenario. None of the Gauss–Newton algorithms returns to its initial convergence speed after the double-talk situation has passed, which could however be solved by resetting the adaptive filter memory at the end of such a double-talk period. The stochastic gradient RPE, PEM-AFROW, and

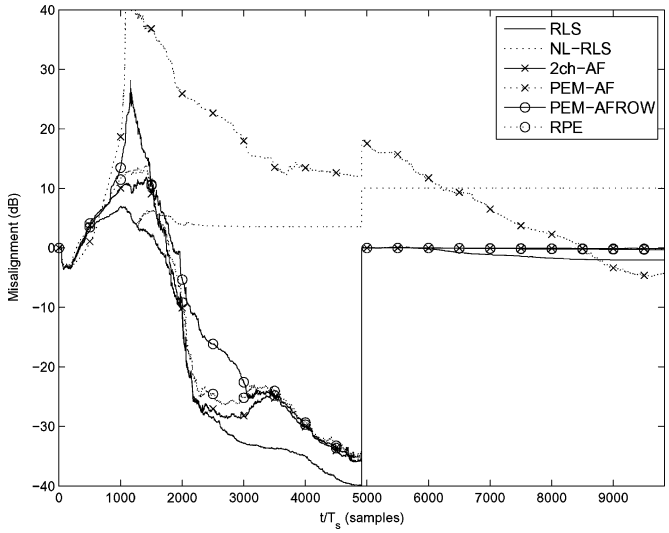


Fig. 7. Echo path change in single-talk: Gauss–Newton simulations. Convergence curves for $n_A = 1$.

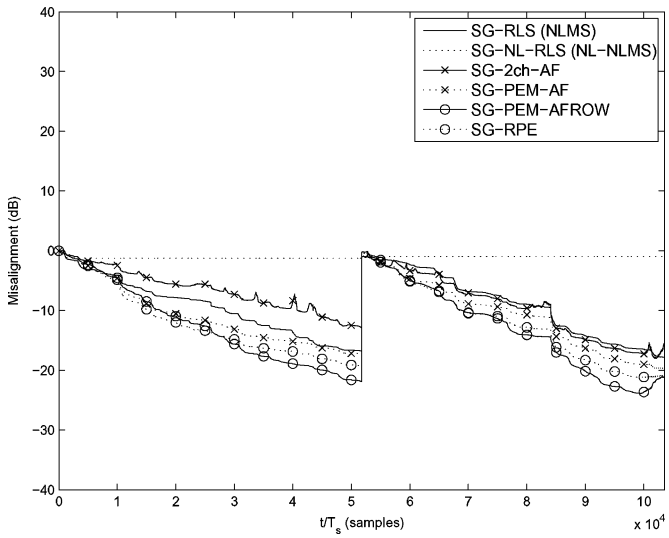


Fig. 8. Echo path change in single-talk: stochastic gradient simulations. Convergence curves for $n_A = 1$.

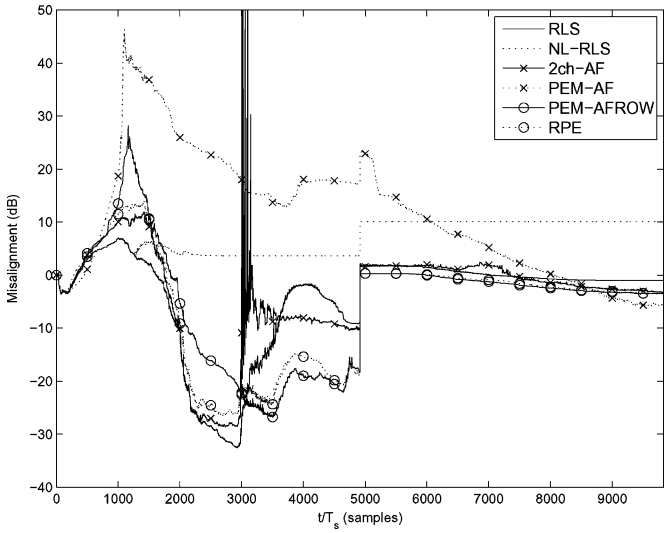


Fig. 9. Echo path change in double-talk: Gauss–Newton simulations. Convergence curves for $n_A = 1$.

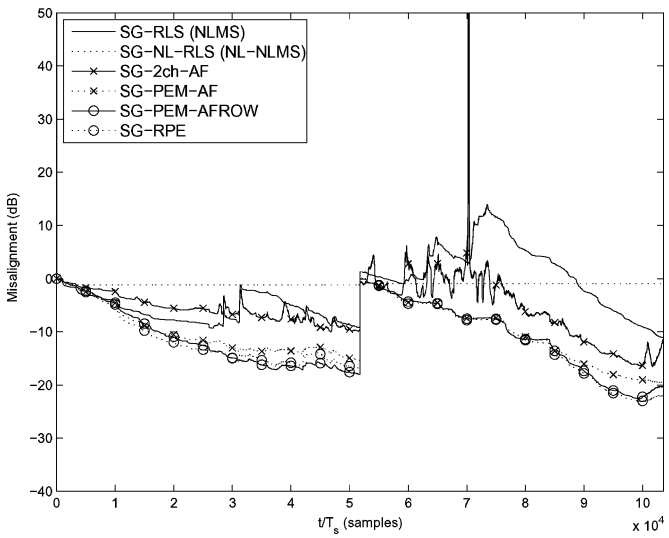


Fig. 10. Echo path change in double-talk: stochastic gradient simulations. Convergence curves for $n_A = 1$.

PEM-AF algorithms clearly outperform all other stochastic gradient algorithms, and they do not even diverge at the onset of double-talk (around $t/T_{\rm s} = 27\,000$ samples).

In a third and fourth scenario, we investigate the behavior of the algorithms when an echo path change occurs. In the middle of the experiment, the echo path gain is decreased by 6 dB, i.e., all RIR coefficients are halved. The far-end signals are the same as before. In the third scenario, the echo path change occurs during single-talk. The convergence curves with $n_A = 1$ are shown in Figs. 7 and 8. The Gauss–Newton algorithms are obviously non-robust to echo path changes. Memory resetting after an echo path change is necessary, and hence these algorithms should be equipped with an echo path change detector (as proposed, e.g., in [5] and [41]–[43]). On the other hand, the stochastic gradient algorithms converge after an echo path change as fast as in the initial convergence phase. In the fourth scenario, the echo path change occurs during double-talk (a worst-case scenario). The near-end signals are the same as in the bursting double-talk scenario. The convergence curves with $n_A = 1$ are drawn in Figs. 9 and 10. The behavior of the Gauss–Newton algorithms after the echo path change is comparable to the third scenario. However, the stochastic gradient RPE, PEM-AF, and PEM-AFROW algorithms converge remarkably faster after the echo path change than the other stochastic gradient algorithms.

VI. Conclusion

We have proposed a new approach to improve double-talk-robustness in adaptive algorithms for acoustic echo cancellation. Minimization of a PE criterion that includes a model of the near-end signal leads to a minimum-variance type, fast-converging recursive echo path estimate. We have applied a new RPE identification algorithm and three PE identification algorithms from adaptive feedback cancellation (PEM-AF, PEM-AFROW, and 2ch-AF) to different double-talk scenarios that

may occur in echo cancellation. Especially in the stochastic gradient case, the PE identification algorithms outperform the standard NLMS algorithm, which is highly sensitive to double-talk. In most double-talk scenarios, the RPE and PEM-AFROW algorithms tend to show the fastest convergence behavior with only a few additional computations and without sacrificing convergence speed during single-talk.

REFERENCES

[1] D. L. Duttweiler, "A twelve-channel digital echo canceler," *IEEE Trans. Commun.*, vol. 26, no. 5, pp. 647–653, May 1978.

[2] J. H. Cho, D. R. Morgan, and J. Benesty, "An objective technique for evaluating doubletalk detectors in acoustic echo cancelers," *IEEE Trans. Speech Audio Process.*, vol. 7, no. 6, pp. 718–724, Nov. 1999.

[3] J. Benesty, D. R. Morgan, and J. H. Cho, "A new class of doubletalk detectors based on cross-correlation," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 2, pp. 168–172, Mar. 2000.

[4] K. Ghose and V. U. Reddy, "A double-talk detector for acoustic echo cancellation applications," *Signal Process.*, vol. 80, no. 8, pp. 1459–1467, Aug. 2000.

[5] J. C. Jenq and S. F. Hsieh, "Decision of double-talk and time-variant echo path for acoustic echo cancellation," *IEEE Signal Process. Lett.*, vol. 10, no. 11, pp. 317–319, Nov. 2003.

[6] T. Jia, Y. Jia, J. Li, and Y. Hu, "Subband doubletalk detector for acoustic echo cancellation systems," in *Proc. 2003 IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Apr. 2003, vol. 5, pp. 604–607.

[7] M. Kallinger, A. Mertins, and K.-D. Kammeyer, "Enhanced double-talk detection based on pseudo-coherence in stereo," in *Proc. 2005 Int. Workshop Acoustic Echo Noise Control (IWAENC)*, Eindhoven, The Netherlands, Sep. 2005, pp. 177–180.

[8] T. Gänsler, J. Benesty, S. L. Gay, and M. M. Sondhi, "A robust proportionate affine projection algorithm for network echo cancellation," in *Proc. 2000 IEEE Int. Conf. Acoustics., Speech, Signal Processing. (ICASSP)*, Istanbul, Turkey, Jun. 2000, vol. 2, pp. 793–796.

[9] P. J. Huber, *Robust Statistics*. New York: Wiley, 1981.

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

[11] H. Buchner, J. Benesty, T. Gänsler, and W. Kellermann, "An outlier-robust extended multidelay filter with application to acoustic echo cancellation," in *Proc. 2003 Int. Workshop Acoustic Echo Noise Control (IWAENC)*, Kyoto, Japan, Sep. 2003, pp. 19–22.

[12] W. Herbordt, H. Buchner, S. Nakamura, and W. Kellermann, "Application of a double-talk resilient DFT-domain adaptive filter for bin-wise stepsize controls to adaptive beamforming," in *Proc. 2005 IEEE EURASIP Workshop Nonlinear Signal Image Process.*, Sapporo, Japan, May 2005.

[13] M. M. Sondhi and D. A. Berkley, "Silencing echoes on the telephone network," *Proc. IEEE*, vol. 68, pp. 948–963, Aug. 1980.

[14] 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.

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

[16] G. Enzner and P. Vary, "Robust and elegant, purely statistical adaptation of acoustic echo canceler and postfilter," in *Proc. 2003 Int. Workshop Acoustic Echo Noise Control (IWAENC)*, Kyoto, Japan, Sep. 2003, pp. 43–46.

[17] M. Kallinger, K.-D. Kammeyer, and J. Bitzer, "Multi-microphone residual echo estimation," in *Proc. 2003 IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Apr. 2003, vol. 5, pp. 584–587.

[18] S. Y. Low and S. Nordholm, "A blind approach to joint noise and acoustic echo cancellation," in *Proc. 2005 IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Philadelphia, PA, Mar. 2005, vol. 3, pp. 69–72.

[19] J. F. Doherty and R. Porayath, "A robust echo canceler for acoustic environments," *IEEE Trans. Circuits Syst. II*, vol. 44, no. 5, pp. 389–396, May 1997.

[20] T. Okuno, M. Fukushima, and M. Tohyama, "Adaptive cross-spectral technique for acoustic echo cancellation," *IEICE Trans. Fundam.*, vol. E82-A, no. 4, pp. 634–639, Apr. 1999.

[21] X. Lu and B. Champagne, "Acoustic echo cancellation over a nonlinear channel," in *Proc. 2001 Int. Workshop Acoustic Echo Noise Control (IWAENC)*, Darmstadt, Germany, Sep. 2001, pp. 139–142.

[22] ——, "Pitch analysis-based acoustic echo cancellation over a nonlinear channel," in *Proc. 11th Eur. Signal Process. Conf. (EUSIPCO)*, Toulouse, France, Sep. 2002, pp. 159–162.

[23] X. Hu, A.-Q. Hu, Y. Chen, and X.-H. Zeng, "An adaptive acoustic echo cancellation without double-talk detection," in *Proc. 2003 Int. Conf. Machine Learning Cybernetics (ICMLC)*, Xi'an, China, Nov. 2003, pp. 972–975.

[24] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ: Prentice-Hall, 1993.

[25] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*. Cambridge, MA: MIT Press, 1986.

[26] T. van Waterschoot, G. Rombouts, and M. Moonen, "On the performance of decorrelation by prefiltering for adaptive feedback cancellation in public address systems," in *Proc. 4th IEEE Benelux Signal Process. Symp. (SPS)*, Hilvarenbeek, The Netherlands, Apr. 2004, pp. 167–170 [Online]. Available: ftp.esat.kuleuven.be/pub/sista/vanwaterschoot/abstracts/04-24.html

[27] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[28] T. van Waterschoot and M. Moonen, "Double-talk robust acoustic echo cancellation with continuous near-end activity," in *Proc. 13th Eur. Signal Process. Conf. (EUSIPCO)*, Antalya, Turkey, Sep. 2005.

[29] R. Hastings-James and M. W. Sage, "Recursive generalised-least-squares procedure for online identification of process parameters," *Proc. Inst. Elect. Eng.*, vol. 116, no. 12, pp. 2057–2062, Dec. 1969.

[30] T. Söderström, L. Ljung, and I. Gustavsson, "A theoretical analysis of recursive identification methods," *Automatica*, vol. 14, no. 3, pp. 231–244, May 1978.

[31] D. W. Clarke, "Generalized least squares estimation of the parameters of a dynamic model," in *Proc. 1st IFAC Symp. Identification Automatic Control Systems*, Prague, Czechoslovakia, 1967.

[32] J. Gertler and C. Bányász, "A recursive (on-line) maximum likelihood identification method," *IEEE Trans. Autom. Control*, vol. AC-19, no. 6, pp. 816–820, Dec. 1974.

[33] T. Söderström, "Convergence properties of the generalized least squares identification method," *Automatica*, vol. 10, no. 6, pp. 617–626, Dec. 1974.

[34] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[35] J. E. Greenberg, "Modified LMS algorithms for speech processing with an adaptive noise canceller," *IEEE Trans. Speech Audio Process.*, vol. 6, no. 4, pp. 338–351, Jul. 2000.

[36] X. Cheng, J. Xu, X. Li, and J. Tian, "Double-talk protection of acoustic echo cancellation based on weighted-sum NLMS algorithms," in *Proc. 6th Int. Conf. Signal Process. (ICSP)*, Beijing, China, Aug. 2002, vol. 1, pp. 297–300.

[37] R. P. Ramachandran and P. Kabal, "Pitch prediction filters in speech coding," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-37, no. 4, pp. 467–478, Apr. 1989.

[38] R. Leber and A. Schaub, "Circuit and method for the adaptive suppression of an acoustic feedback," U.S. Patent 6 611 600, Aug. 26, 2003.

[39] T. van Waterschoot, G. Rombouts, and M. Moonen, "Software for double-talk robust acoustic echo cancellation," Katholieke Universiteit Leuven, Belgium, Tech. Rep. ESAT-SISTA TR 05-162, Oct. 2005 [Online]. Available: ftp.esat.kuleuven.be/pub/sista/vanwaterschoot/abstracts/05-162.html

[40] T. van Waterschoot, G. Rombouts, K. Struyve, and M. Moonen, "Acoustic echo cancellation in the presence of continuous double-talk," in *Proc. 1st Annu. IEEE Benelux/DSP Valley Signal Process. Symp. (SPS-DARTS)*, Antwerp, Belgium, Apr. 2005, pp. 53–57 [Online]. Available: ftp.esat.kuleuven.be/pub/sista/vanwaterschoot/abstracts/04-149.html

[41] J. Huo, S. Nordholm, and Z. Zang, "A method for detecting echo path variation," in *Proc. 2003 Int. Workshop Acoustic Echo Noise Control (IWAENC)*, Kyoto, Japan, Sep. 2003, pp. 71–74.

[42] J. Liu, "Efficient and robust cancellation of echoes with long echo path delay," *IEEE Trans. Commun.*, vol. 52, no. 8, pp. 1288–1291, Aug. 2004.

[43] H. K. Jung, N. S. Kim, and T. Kim, "A new double-talk detector using echo path estimation," *Speech Commun.*, vol. 45, no. 1, pp. 41–48, Jan. 2005.



Toon van Waterschoot (S'04) was born in Lier, Belgium, on June 11, 1979. He received the Master's degree from Katholieke Universiteit Leuven (K.U. Leuven), Belgium, in 2001, where he is currently working towards the Ph.D. degree.

In 2002, he spent a year as a Teaching Assistant with the Hogere Zeevaartschool Antwerpen, Belgium, after which he returned to K.U. Leuven to start a Ph.D. research project on signal processing applied to audio and acoustics. His current research is focused on the development of adaptive filtering algorithms for acoustic echo and feedback cancellation.



Geert Rombouts was born in Turnhout, Belgium, in 1973. He received the Master's degree in electrical engineering and the Ph.D. degree in engineering (Toegepaste Wetenschappen) from the Katholieke Universiteit Leuven, Belgium, in 1997 and 2003, respectively.

Currently, he is a Postdoctoral Researcher at the ESAT-SCD laboratories of the Katholieke Universiteit Leuven.



Piet Verhoeve (S'88–M'92–SM'99) received the M.S. degree in electrical mechanical engineering from the Katholieke Universiteit Leuven, Belgium, in June 1991 and the Ph.D. degree in electronic engineering from Ghent University, Belgium, in 1998.

From August 1992 to September 1998, he was a Research Assistant at IMEC (Ghent). From October 1998 to January 2001, he worked as a Software Architect at Lernout & Hauspie, a speech software company. Since February 2001, he is working as a Senior Research Associate at Televic, Izegem, Belgium, where he is responsible for innovation and research in collaboration with institutes, universities, and engineering schools. His main research interests include low-latency digital audio and video, wireless sensor networks, and location aware computing. He is author and coauthor of 20 publications and three patent applications.



Marc Moonen (M'94–SM'06–F'07) received the electrical engineering degree and the Ph.D. degree in applied sciences from Katholieke Universiteit Leuven (K.U. Leuven), Belgium, in 1986 and 1990, respectively.

Since 2004, he has been a Full Professor at the Electrical Engineering Department of Katholieke Universiteit Leuven, where he is currently heading a research team of 16 Ph.D. candidates and postdocs, working in the area of numerical algorithms and signal processing for digital communications, wireless communications, DSL, and audio signal processing.

Dr. Moonen received the 1994 K.U. Leuven Research Council Award, the 1997 Alcatel Bell (Belgium) Award (with Piet Vandaele), and the 2004 Alcatel Bell (Belgium) Award (with Raphael Cendrillon) and was a 1997 "Laureate of the Belgium Royal Academy of Science." He received a journal Best Paper Award (with Geert Leus) from the IEEE TRANSACTIONS ON SIGNAL PROCESSING and (with Simon Doclo) from Elsevier Signal Processing. He was Chairman of the IEEE Benelux Signal Processing Chapter from 1998 to 2002 and has been a European Association for Signal, Speech and Image Processing (EURASIP) AdCom Member since 2000 and a member of the IEEE Signal Processing Society Technical Committee on Signal Processing for Communications. He has served as Editor-in-Chief for the *EURASIP Journal on Applied Signal Processing* from 2003 to 2005 and has been a member of the editorial board of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I from 2002 to 2003 and IEEE SIGNAL PROCESSING MAGAZINE from 2003 to 2005. He is currently a member of the editorial board of *Integration, the VLSI Journal*, *EURASIP Journal on Applied Signal Processing*, *EURASIP Journal on Wireless Communications and Networking*, and *Signal Processing*.