**Katholieke Universiteit Leuven**

Departement Elektrotechniek ESAT-SISTA/TR 04-149

# Acoustic Echo Cancellation in the Presence of Continuous Double-Talk[1]

Toon van Waterschoot[2] [3], Geert Rombouts[2], Kris Struyve[4] and Marc Moonen[2]

December 2004
Published in *Proceedings of the first annual IEEE BENELUX/DSP Valley Signal Processing Symposium (SPS-DARTS 2005)*, Het Provinciehuis, Antwerp, Belgium, April 19-20, 2005, pp. 53-57

[1]This report is available by anonymous ftp from *ftp.esat.kuleuven.ac.be* in the directory *pub/sista/vanwaterschoot/reports/04-149.pdf*

[2]K.U.Leuven, Dept. of Electrical Engineering (ESAT), Research group SCD(SISTA), Kasteelpark Arenberg 10, 3001 Leuven, Belgium, Tel. +32 16 321927, Fax +32 16 321970, WWW: *http://www.esat.kuleuven.ac.be/sista-cosic-docarch*. E-mail: *toon.vanwaterschoot@esat.kuleuven.ac.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-MEFISTO-666 (Mathematical Engineering for Information and Communication Systems Technology) of the Flemish Government and IWT project 020476: 'SMS4PA: Sound Management System for Public Address Systems'. The scientific responsibility is assumed by its authors.

[4]Televic N.V., L. Bekaertlaan 1, B-8870 Izegem, Belgium

# ACOUSTIC ECHO CANCELLATION IN THE PRESENCE OF CONTINUOUS DOUBLE-TALK

[1,2]*Toon van Waterschoot,* [2]*Geert Rombouts,* [3]*Kris Struyve and* [2]*Marc Moonen*

[1]toon.vanwaterschoot@esat.kuleuven.ac.be
[2]Katholieke Universiteit Leuven, ESAT-SCD, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium
[3]Televic N.V., L. Bekaertlaan 1, B-8870 Izegem, Belgium

**ABSTRACT**

The use of a double-talk detector in acoustic echo cancellation cannot improve the adaptive algorithm's performance if near-end noise is continuously present. We propose a new way of dealing with such a continuous double-talk situation, which may occur e.g. in an automatic gain adjustment application. If the microphone and loudspeaker signals are prefiltered with the inverse near-end signal model, a minimum variance room impulse response estimate can be obtained. However, the near-end signal model is unknown and time-varying and has to be estimated concurrently with the room impulse response. We apply three different prediction error identification algorithms to this problem, that were originally developed for adaptive feedback cancellation. Simulation results indicate that only the prediction error method based adaptive filtering algorithm applying row operations (PEM-AFROW) outperforms standard RLS or NLMS adaptive algorithms.

## 1. 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 classical AEC scenario is shown in **Figure 1**. A speech signal $u(t)$ from the far-end side is broadcasted 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 modelled 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

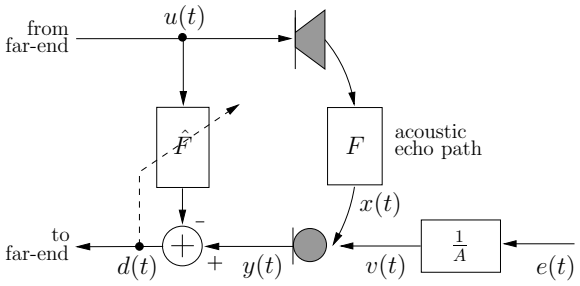



Figure 1: AEC scenario with AR modelling of the near-end signal.

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 $F(q,t)$.

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 a well-known problem in AEC that the convergence speed and hence the tracking capabilities of standard adaptive algorithms (like recursive least squares (RLS) or normalized least mean squares (NLMS)) may decrease severely when near-end noise is present ('double-talk' periods). A lot of effort has been spent on the design of efficient double-talk detectors (DTD), which are used to slow down or switch off the adaptation during double-talk periods [1]. Nevertheless in some scenarios near-end noise will be continuously present and the use of a DTD becomes futile. This may be the case for example when a loudspeaker message is broadcasted in a noisy environment and one wants to automatically adjust the broadcasting volume to the ambient noise level. For accurate estimation of the ambient noise level and the level of the loudspeaker signal arriving at the microphone, identification of the acoustic echo path during double-talk periods is required. An example application is a crowded train station in which the acoustic impulse response from a public address loudspeaker to a measurement microphone on the platform is to be identified. The sound levels of the separated components can then be used to calculate a gain factor for the public address system.

In this paper we aim at developing a recursive identification algorithm that allows for continuous adaptation of the RIR estimate, yet behaves in a robust way in double-talk situations. From linear estimation theory [2], we know that the best (i.e. minimum

[1]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-MEFISTO-666 (Mathematical Engineering for Information and Communication Systems Technology) of the Flemish Government and IWT project 020476: 'SMS4PA: Sound Management System for Public Address Systems'. The scientific responsibility is assumed by its authors.

variance) linear unbiased estimate (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 which 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. However the near-end signal characteristics are typically unknown and may be highly time-varying. Therefore they need to be estimated concurrently with the unknown RIR.

The problem of concurrently estimating the RIR and the near-end signal characteristics has been studied in the context of adaptive feedback cancellation (AFC) [3], [4], [5]. We will explore two algorithms developed for AFC in hearing aids [3], namely the two-channel adaptive filtering (2ch-AF) algorithm and the prediction error method based adaptive filtering (PEM-AF) algorithm. These 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 [4] 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. We will demonstrate by means of simulations that indeed the 2ch-AF and PEM-AF algorithms perform similar to or even worse than a standard RLS or NLMS adaptive algorithm when applied to AEC in the presence of continuous double-talk.

We have recently proposed a new prediction error method based AFC algorithm [5] that does not rely on the above assumption and is therefore better suited for room acoustics applications. In contrast to the hearing aid PEM-AF algorithm, this new algorithm performs only row operations on the loudspeaker signal data matrix, hence the name PEM-AFROW. When applied to the AEC problem with continuous double-talk, the PEM-AFROW identification algorithm shows an improved convergence compared to the RLS, 2ch-AF and PEM-AF algorithms. The improvement is even more distinct when a stochastic gradient implementation of the algorithms is considered.

## 2. PREDICTION ERROR CRITERION

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

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

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

$$F(q,t) = f_0(t) + f_1(t)q^{-1} + \ldots + f_{n_F}(t)q^{-n_F} \tag{2}$$

$$A(q,t) = 1 + a_1(t)q^{-1} + \ldots + a_{n_A}(t)q^{-n_A}. \tag{3}$$

This implies that the near-end signal $v(t)$ is modelled as an auto-regressive (AR) sequence, a model which is physically relevant for speech signals and also not uncommon for audio signals other than speech. Note that $A(q,t)$ is constrained to be monic, i.e. $a_0 = 1$. The signal $e(t)$ represents the excitation signal generating the near-end signal, and is assumed to be white noise with variance $\sigma_t^2$.

Following system identification theory [6] the best one-step-ahead predictor for $y(t)$ is given by

$$\hat{y}(t|\boldsymbol{\theta}(t)) = [1 - A(q,t)]y(t) + A(q,t)F(q,t)u(t). \tag{4}$$

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

$$\boldsymbol{\theta}(t) \triangleq \begin{bmatrix} \mathbf{f}(t) \\ \mathbf{a}(t) \end{bmatrix} \tag{5}$$

with

$$\mathbf{f}(t) \triangleq \begin{bmatrix} f_0(t) & f_1(t) & \ldots & f_{n_F}(t) \end{bmatrix}^T, \tag{6}$$

$$\mathbf{a}(t) \triangleq \begin{bmatrix} a_1(t) & a_2(t) & \ldots & a_{n_A}(t) \end{bmatrix}^T. \tag{7}$$

The prediction error, defined as

$$\varepsilon(t, \boldsymbol{\theta}(t)) \triangleq y(t) - \hat{y}(t|\boldsymbol{\theta}(t)), \tag{8}$$

can with (4) be calculated as

$$\varepsilon(t, \boldsymbol{\theta}(t)) = A(q,t)y(t) - A(q,t)F(q,t)u(t) \tag{9}$$

$$= A(q,t)[y(t) - F(q,t)u(t)]. \tag{10}$$

Hence the prediction error criterion for estimating $\boldsymbol{\theta}(t)$ may be defined as [6]

$$V_{PE}(t, \boldsymbol{\theta}(t)) = \frac{1}{2}\sum_{k=1}^{t} \frac{\lambda_{\boldsymbol{\theta}}^{t-k}}{\hat{\sigma}_k^2}\varepsilon^2(k, \boldsymbol{\theta}(t)) \tag{11}$$

$$= \frac{1}{2}\sum_{k=1}^{t} \frac{\lambda_{\boldsymbol{\theta}}^{t-k}}{\hat{\sigma}_k^2}\big(A(q,t)[y(k) - F(q,t)u(k)]\big)^2 \tag{12}$$

where the exponential weighting with forgetting factor $\lambda_{\boldsymbol{\theta}}$ is included for discounting old data and $\hat{\sigma}_k^2$ is an estimate of the prediction error variance at time $k$, calculated using the most recent estimate $\hat{\boldsymbol{\theta}}(k-1)$ of the unknown parameter vector and an appropriate exponential forgetting factor $\lambda_\sigma$:

$$\hat{\sigma}_k^2 = \lambda_\sigma \hat{\sigma}_{k-1}^2 + (1 - \lambda_\sigma)\varepsilon^2(k, \hat{\boldsymbol{\theta}}(k-1)). \tag{13}$$

If the near-end signal $v(t)$ can be modelled as an AR process of order $n_A$ and if the true AR polynomial $A(q,t)$ and the true near-end excitation signal variance $\sigma_t^2$ are known at each time instant, minimization of the prediction error criterion (12) w.r.t. $\mathbf{f}(t)$ only, on a data window in which $F(q,t)$ can be considered stationary, yields the best (minimum variance) linear unbiased estimate (BLUE) of $\mathbf{f}(t)$. Of course, in practice the AR polynomial $A(q,t)$ is unknown and the criterion in (12) is minimized w.r.t. both $\mathbf{f}(t)$ and $\mathbf{a}(t)$.

## 3. RECURSIVE IDENTIFICATION ALGORITHMS

We will now review three recursive algorithms that aim at minimizing the prediction error criterion (12) and were derived in an adaptive feedback cancellation (AFC) framework. We will show these algorithms are also applicable to the current AEC problem.

### 3.1. Two-channel adaptive filtering (2ch-AF) algorithm

In the two-channel adaptive filtering (2ch-AF) algorithm proposed in [3] it is noted that expression (9) for the prediction error can be written as

$$\varepsilon(t, \boldsymbol{\theta}_{\boldsymbol{2ch}}(t)) = A(q,t)y(t) + B(q,t)u(t) \tag{14}$$

with

$$B(q,t) \triangleq -A(q,t)F(q,t) \tag{15}$$

and

$$\boldsymbol{\theta_{2ch}}(t) \triangleq [b_0(t) \ \dots \ b_{n_B}(t) \quad a_1(t) \ \dots \ a_{n_A}(t)]^T. \tag{16}$$

This interpretation leads to a criterion that is quadratic in the parameter vector $\boldsymbol{\theta_{2ch}}(t)$:

$$\begin{aligned} V_{2ch}(t,\boldsymbol{\theta_{2ch}}(t)) &= \frac{1}{2}\sum_{k=1}^{t}\frac{\lambda_{\boldsymbol{\theta}}^{t-k}}{\hat{\sigma}_k^2}\big(A(q,t)y(k)+B(q,t)u(k)\big)^2 \\ &= \frac{1}{2}\sum_{k=1}^{t}\frac{\lambda_{\boldsymbol{\theta}}^{t-k}}{\hat{\sigma}_k^2}\big(y(k)+\boldsymbol{\phi_{2ch}}^T(k)\boldsymbol{\theta_{2ch}}(t)\big)^2 \end{aligned}$$

with

$$\boldsymbol{\phi_{2ch}}(k) \triangleq [u(k) \ \dots \ u(k-n_B) \quad y(k-1) \ \dots \ y(k-n_A)]^T,$$

so that an ordinary recursive least squares (RLS) algorithm can be applied for recursive estimation of $\boldsymbol{\theta_{2ch}}(t)$:

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{2ch}}(t) = \hat{\boldsymbol{\theta}}_{\boldsymbol{2ch}}(t-1) - \frac{1}{\hat{\sigma}_t^2}\mathbf{R}^{-1}(t)\boldsymbol{\phi_{2ch}}(t)\varepsilon(t,\hat{\boldsymbol{\theta}}_{\boldsymbol{2ch}}(t-1)),$$

$$\mathbf{R}(t) = \lambda_{\boldsymbol{\theta}}\mathbf{R}(t-1) + \frac{1}{\hat{\sigma}_t^2}\boldsymbol{\phi_{2ch}}(t)\boldsymbol{\phi_{2ch}}^T(t), \tag{17}$$

$$\varepsilon(t,\hat{\boldsymbol{\theta}}_{\boldsymbol{2ch}}(t-1)) = y(t) + \boldsymbol{\phi_{2ch}}^T(t)\hat{\boldsymbol{\theta}}_{\boldsymbol{2ch}}(t-1). \tag{18}$$

The main drawback of this algorithm is that the coefficients of $F(q,t)$ and $A(q,t)$, which are combined in $B(q,t)$, are estimated on the same data window. As pointed out in [4] for the AFC case, this makes the algorithm, which was originally derived for hearing aid applications, unsuitable when dealing with room acoustics. The reason is that a room impulse response is typically of very high order and thus a relatively large data window is required to obtain a reliable room impulse response estimate. On the other hand, audio and in particular speech signal characteristics may be highly time-varying and can only be estimated accurately on relatively short data windows. Therefore it is desirable to identify $F(q,t)$ and $A(q,t)$ on data windows of different sizes, as is the case in the PEM-AF and PEM-AFROW algorithms described below.

### 3.2. Prediction error method based adaptive filtering (PEM-AF) algorithm

Next to the 2ch-AF algorithm, a prediction error method based adaptive filtering (PEM-AF) algorithm for hearing aid AFC was derived in [3]. This algorithm features a decoupling of the data windows on which $F(q,t)$ and $A(q,t)$ are estimated. The algorithm is based on the recursive update equation [6]

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{1}{\hat{\sigma}_t^2}\mathbf{R}^{-1}(t)\boldsymbol{\psi}(t,\hat{\boldsymbol{\theta}}(t-1))\varepsilon(t,\hat{\boldsymbol{\theta}}(t-1)), \tag{19}$$

with $\boldsymbol{\theta}(t)$ defined in (5) and the $(n_F+n_A+1)\times 1$ gradient vector $\boldsymbol{\psi}(t,\boldsymbol{\theta}(t))$ defined as

$$\boldsymbol{\psi}(t,\boldsymbol{\theta}(t)) \triangleq -\frac{\partial}{\partial\boldsymbol{\theta}(t)}\varepsilon(t,\boldsymbol{\theta}(t)) \tag{20}$$

or, in partioned form,

$$\begin{bmatrix}\boldsymbol{\psi_f}(t,\mathbf{a}(t))\\ \boldsymbol{\psi_a}(t,\mathbf{f}(t))\end{bmatrix} \triangleq -\begin{bmatrix}\frac{\partial}{\partial\mathbf{f}(t)}\varepsilon(t,\boldsymbol{\theta}(t))\\ \frac{\partial}{\partial\mathbf{a}(t)}\varepsilon(t,\boldsymbol{\theta}(t))\end{bmatrix}. \tag{21}$$

The matrix $\mathbf{R}(t)$ can be any positive definite approximation of the Hessian $\frac{\partial^2}{\partial\boldsymbol{\theta}^2(t)}V_{PE}(t,\boldsymbol{\theta}(t))$ of the criterion in (12). If $\mathbf{R}(t)$ is updated according to

$$\mathbf{R}(t) = \lambda_{\boldsymbol{\theta}}\mathbf{R}(t-1) + \frac{1}{\hat{\sigma}_t^2}\boldsymbol{\psi}(t,\hat{\boldsymbol{\theta}}(t-1))\boldsymbol{\psi}^T(t,\hat{\boldsymbol{\theta}}(t-1)), \tag{22}$$

the stochastic Gauss-Newton method is obtained. Decoupling of the data windows on which $F(q,t)$ and $A(q,t)$ are estimated is then possible by block-diagonalizing $\mathbf{R}(t)$ as

$$\mathbf{R}(t) = \begin{bmatrix}\mathbf{R_f}(t) & \mathbf{0}_{(n_F+1)\times n_A}\\ \mathbf{0}_{n_A\times(n_F+1)} & \mathbf{R_a}(t)\end{bmatrix} \tag{23}$$

and by introducing two different forgetting factors $\lambda_{\mathbf{f}}$ and $\lambda_{\mathbf{a}}$, where typically $\lambda_{\mathbf{f}}$ is much closer to 1 than $\lambda_{\mathbf{a}}$:

$$\mathbf{R_f}(t) = \lambda_{\mathbf{f}}\mathbf{R_f}(t-1) + \frac{1}{\hat{\sigma}_t^2}\boldsymbol{\psi_f}(t,\hat{\mathbf{a}}(t-1))\boldsymbol{\psi_f}^T(t,\hat{\mathbf{a}}(t-1)), \tag{24}$$

$$\mathbf{R_a}(t) = \lambda_{\mathbf{a}}\mathbf{R_a}(t-1) + \frac{1}{\hat{\sigma}_t^2}\boldsymbol{\psi_a}(t,\hat{\mathbf{f}}(t-1))\boldsymbol{\psi_a}^T(t,\hat{\mathbf{f}}(t-1)). \tag{25}$$

As is common in recursive system identification [6], the prediction error in (10) and the gradient vector in (21) are in the PEM-AF algorithm approximated as:

$$\varepsilon(t) = \hat{A}(q,t-1)\xi(t), \tag{26}$$

$$\boldsymbol{\psi_f}(t) = [\hat{A}(q,t-1)u(t) \ \dots \ \hat{A}(q,t-n_F-1)u(t-n_F)]^T, \tag{27}$$

$$\boldsymbol{\psi_a}(t) = \begin{bmatrix}\xi(t-1) & \dots & \xi(t-n_A)\end{bmatrix}^T, \tag{28}$$

with

$$\xi(t) \triangleq y(t) - \hat{F}(q,t-1)u(t). \tag{29}$$

With this approximation it is assumed that $\hat{A}(q,t-1) = \hat{A}(q,t-1-i)$ for $i\in[0,n_F]$ and that $\hat{F}(q,t-1) = \hat{F}(q,t-1-j)$ for $j\in[1,n_A]$. As already mentioned before, in an AEC scenario $F(q,t)$ will typically be of very high order and $A(q,t)$ may be highly time-varying over a data window of length $n_F$. Therefore the first assumption, and hence the approximation (27) of the gradient vector $\boldsymbol{\psi_f}(t,\hat{\mathbf{a}}(t-1))$, will not hold for the current AEC application, in contrast to the hearing aid AFC application considered in [3].

### 3.3. Prediction error method based adaptive filtering algorithm applying row operations (PEM-AFROW)

We have recently proposed a new prediction error method based adaptive filtering algorithm in an AFC framework [5]. In contrast to the PEM-AF algorithm described above, this algorithm calculates the gradient vector $\boldsymbol{\psi_f}(t,\hat{\mathbf{a}}(t-1))$ by applying only row operations to the loudspeaker signal data matrix, hence the name PEM-AFROW.

The PEM-AFROW algorithm is frame-based and in [5] the frame length is set to $20ms$, which is the average period over which a speech signal can be considered locally stationary. Let the frame length be denoted by $M$ and the current frame index by $i$, and suppose that consecutive frames overlap with $n_A$ samples. The first step of the algorithm consists in calculating a residual vector $\mathbf{d_i}$ based on the loudspeaker and microphone data of the $i$-th frame and by using a previous estimate $\hat{\mathbf{f}}((i-1)M)$ of the room impulse response:

$$\mathbf{d_i} = \mathbf{y_i} - \mathbf{U_i}\hat{\mathbf{f}}((i-1)M) \tag{30}$$

with

$$\mathbf{d_i} \triangleq \begin{bmatrix} d(iM) & \dots & d((i-1)M+1-n_A) \end{bmatrix}^T, \quad (31)$$

$$\mathbf{y_i} \triangleq \begin{bmatrix} y(iM) & \dots & y((i-1)M+1-n_A) \end{bmatrix}^T, \quad (32)$$

$$\mathbf{U_i} \triangleq \begin{bmatrix} u(iM) & \dots & u(iM-n_F) \\ \vdots & \ddots & \vdots \\ u((i-1)M+1-n_A) & \dots & u((i-1)M+1-n_A-n_F) \end{bmatrix}.$$

A linear prediction using the Levinson-Durbin recursion is then performed on $\mathbf{d_i}$ to estimate the AR coefficients $\mathbf{a_i}$ and the prediction error variance $\sigma_i^2$ which describe the near-end signal characteristics in the $i$-th frame.

Then in the second step of the PEM-AFROW algorithm an exponentially windowed RLS algorithm is applied to estimate $F(q,t)$ with $t \in [(i-1)M+1, iM]$ from the criterion

$$V_{PE}(t, \mathbf{f}(t)) = \frac{1}{2} \sum_{j=1}^{i-1} \sum_{k=k_0}^{jM} \frac{\lambda^{t-k}}{\hat{\sigma}_j^2} \left( y_{\hat{\mathbf{a}}_\mathbf{j}}(k) - F(q,t) u_{\hat{\mathbf{a}}_\mathbf{j}}(k) \right)^2 + \frac{1}{2} \sum_{k=(i-1)M+1}^{t} \frac{\lambda^{t-k}}{\hat{\sigma}_i^2} \left( y_{\hat{\mathbf{a}}_\mathbf{i}}(k) - F(q,t) u_{\hat{\mathbf{a}}_\mathbf{i}}(k) \right)^2 \quad (33)$$

with $k_0 \triangleq (j-1)M+1$ and, for $j \in [1, i]$,

$$y_{\hat{\mathbf{a}}_\mathbf{j}}(k) \triangleq \begin{bmatrix} y(k) & \dots & y(k-n_A) \end{bmatrix} \begin{bmatrix} 1 \\ \hat{\mathbf{a}}_\mathbf{j} \end{bmatrix}, \quad (34)$$

$$u_{\hat{\mathbf{a}}_\mathbf{j}}(k) \triangleq \begin{bmatrix} u(k) & \dots & u(k-n_A) \end{bmatrix} \begin{bmatrix} 1 \\ \hat{\mathbf{a}}_\mathbf{j} \end{bmatrix}. \quad (35)$$

Note that in (33) the summation over the entire time interval $[1, t]$ is split up into a summation over the frame index $j$ and consequently over the sample index $k$ inside each frame.

The RLS updating equations for this second step of the PEM-AFROW algorithm on the $i$-th frame are then given by

$$\hat{\mathbf{f}}(t) = \hat{\mathbf{f}}(t-1) + \frac{1}{\hat{\sigma}_i^2} \mathbf{R}^{-1}(t) \boldsymbol{\psi_f}(t, \hat{\mathbf{a}}_\mathbf{i}) \varepsilon(t, \hat{\mathbf{f}}(t-1), \hat{\mathbf{a}}_\mathbf{i}) \quad (36)$$

$$\mathbf{R}(t) = \lambda_\mathbf{f} \mathbf{R}(t-1) + \frac{1}{\hat{\sigma}_i^2} \boldsymbol{\psi_f}(t, \hat{\mathbf{a}}_\mathbf{i}) \boldsymbol{\psi_f}^T(t, \hat{\mathbf{a}}_\mathbf{i}) \quad (37)$$

$$\varepsilon(t, \hat{\mathbf{f}}(t-1), \hat{\mathbf{a}}_\mathbf{i}) = y_{\hat{\mathbf{a}}_\mathbf{i}}(t) - \boldsymbol{\psi_f}^T(t, \hat{\mathbf{a}}_\mathbf{i}) \hat{\mathbf{f}}(t-1) \quad (38)$$

where the gradient vector is now defined as

$$\boldsymbol{\psi_f}(t, \hat{\mathbf{a}}_\mathbf{i}) \triangleq \begin{bmatrix} u_{\hat{\mathbf{a}}_\mathbf{i}}(t) & \dots & u_{\hat{\mathbf{a}}_\mathbf{i}}(t-n_F) \end{bmatrix}^T. \quad (39)$$

In this way obviously all elements of the loudspeaker signal data vector $[u(t) \ \dots \ u(t-n_F)]^T$ are prefiltered with the same AR coefficients $[1 \quad \hat{\mathbf{a}}_\mathbf{i}^T]$ to obtain the gradient vector $\boldsymbol{\psi_f}(t, \hat{\mathbf{a}}_\mathbf{i})$. In other words, only row operations are performed on the loudspeaker signal data matrix.

As pointed out in [5] for the AFC case, the PEM-AFROW algorithm outperforms the PEM-AF algorithm proposed in [3] for long acoustic paths. Simulations will reveal that this is also true in an AEC scenario, see **Section 4**.

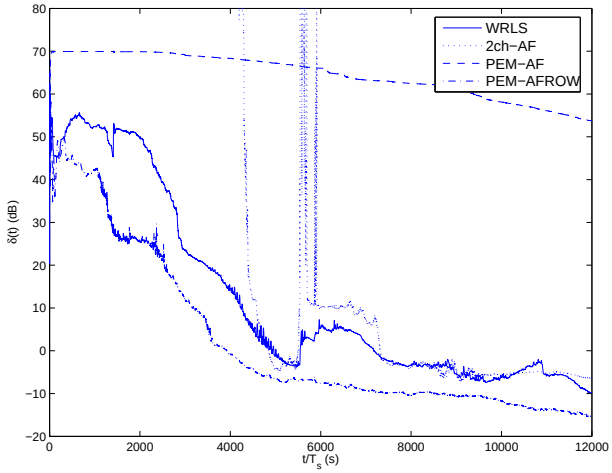



Figure 2: Convergence curves of recursive identification algorithms.

## 4. SIMULATION RESULTS

MATLAB simulations were performed to compare the convergence properties of the three recursive identification algorithms described in **Section 3**. Moreover, a weighted recursive least squares (WRLS) algorithm was implemented as a reference algorithm, with the inverse weights calculated as in (13). At a sampling rate $f_s = 8kHz$, $F(q,t)$ was a fixed, realistic room impulse response of length $n_F+1 = 1000$. The AR model order was set to $n_A = 55$, a value high enough to predict also the pitch of the near-end excitation signal during voiced speech segments. In the 2ch-AF algorithm we chose $n_B = n_F + n_A$. The far-end signal $u(t)$ was a $1.5s$ male speech fragment and the near-end signal $v(t)$ a $1.5s$ female speech fragment. The near-end signal $v(t)$ was scaled such that the average echo-to-background ratio (EBR) was equal to $10dB$:

$$EBR \triangleq \frac{\sum_{k=1}^{N} |x(k)|^2}{\sum_{k=1}^{N} |v(k)|^2} = 10dB. \quad (40)$$

$N = 12000$ denotes the number of data points used. Every exponential forgetting factor was set such that the data window length for estimating the respective parameter vector was roughly three times the parameter vector length, i.e. $\lambda_\mathbf{f} = \lambda_{\boldsymbol{\theta}} = 0.9997$ and $\lambda_\mathbf{a} = \lambda_\sigma = 0.9940$. The performance measure used for comparison was the logarithmic normalized distance between the true and the estimated RIR, defined as

$$\delta(t) = 20 \log_{10} \frac{\|\hat{\mathbf{f}}(t) - \mathbf{f}\|}{\|\mathbf{f}\|}. \quad (41)$$

The convergence curves obtained with these simulation parameters are shown in **Figure 2**. The 2ch-AF algorithm initially performs extremely bad but eventually settles down to a convergence rate comparable to the WRLS algorithm, with occasionally excessive performance dips (which may be due to the deconvolution which is necessary to extract the RIR coefficients from $B(q,t)$, see (15)). The PEM-AF algorithm performs even worse, as it does not converge to the desired solution at all. This is probably due to convergence to a local minimum of the prediction error criterion. The PEM-AFROW algorithm obviously

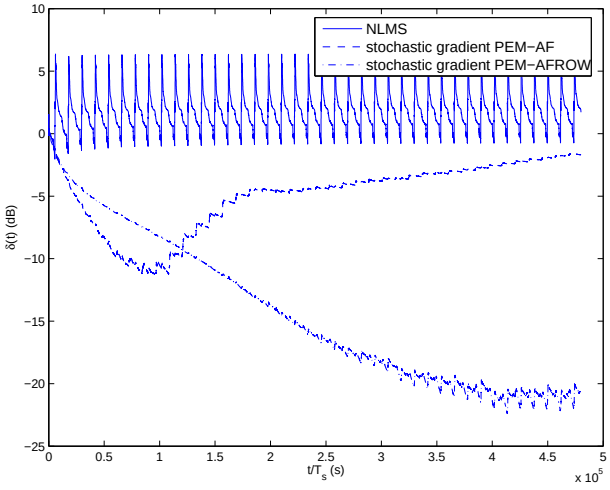



Figure 3: Convergence curves of stochastic gradient algorithms.

shows the best performance. The normalized distance is with the PEM-AFROW algorithm on average $10dB$ smaller than the WRLS normalized distance.

In a second experiment the performance of a stochastic gradient implementation of the algorithms described in **Section 3** was compared to a normalized least mean squares (NLMS) algorithm. Only the RIR identification was performed using the stochastic gradient method, whereas the identification of the AR coefficients was performed as in the previous experiment. **Figure 3** shows the convergence plots for the NLMS algorithm and the stochastic gradient PEM-AF and PEM-AFROW algorithm. The simulation setup was the same as for the first experiment. The step size was set to $\mu = 0.5$ for all algorithms. The speech fragments from the previous experiment were repeated until a sequence of $N = 480000$ data points was obtained. Both the NLMS and the stochastic gradient 2ch-AF algorithm do not converge to the desired solution and their convergence curves show a large variance. The stochastic gradient 2ch-AF algorithm exhibits an extremely large variance and for clarity its convergence curve is not drawn in **Figure 3**. The stochastic gradient PEM-AF algorithm shows the best initial convergence, but afterwards diverges away from the desired solution. The stochastic gradient PEM-AFROW algorithm obviously shows the best performance. In the stochastic gradient case the performance improvement obtained with the PEM-AFROW algorithm is even more distinct than in the first experiment.

## 5. CONCLUSION

We have proposed a new way of coping with a continuous double-talk situation in acoustic echo cancellation. Three algorithms were presented in a common framework, which all aim at minimizing a prediction error criterion for the identification of the acoustic echo path. The 2ch-AF and PEM-AF algorithms however rely on the assumption that the near-end signal characteristics do not change significantly on a time scale comparable to the room impulse response length. Therefore these algorithms do not perform better than an ordinary RLS or NLMS algorithm. The PEM-AFROW algorithm on the other hand appears to be well suited to the current problem, as it exhibits an improved convergence compared to the other algorithms.

## 6. REFERENCES

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

[2] S. M. Kay, *Fundamentals of statistical signal processing: estimation theory*, Prentice-Hall Inc., Upper Saddle River, New Jersey, USA, 1993.

[3] A. Spriet, I. Proudler, M. Moonen, and J. Wouters, "Adaptive feedback cancellation with linear prediction of the desired signal," Accepted for publication in IEEE Trans. Signal Processing. Available at ftp://ftp.esat.kuleuven.ac.be/pub/sista/spriet/reports/03-167.pdf, Aug. 2003.

[4] T. van Waterschoot, G. Rombouts, and M. Moonen, "On the performance of decorrelation by prefiltering for adaptive feedback cancellation in Public Address systems," in *Proceedings of the 2004 IEEE Benelux Signal Processing Symposium (SPS 2004)*, Hilvarenbeek, The Netherlands, Apr. 2004, pp. 167–170, Available at ftp://ftp.esat.kuleuven.ac.be/pub/sista/vanwaterschoot/reports/04-24.pdf.

[5] G. Rombouts, T. van Waterschoot, K. Struyve, and M. Moonen, "Acoustic feedback cancellation for long acoustic paths using a nonstationary source model," ESAT-SISTA Technical Report TR 04-151, Katholieke Universiteit Leuven, Belgium. Submitted to IEEE Trans. Signal Processing, Mar. 2005.

[6] L. Ljung and T. Söderström, *Theory and practice of recursive identification*, MIT Press, Cambridge, Massachusetts, USA, 3rd edition, 1986.