Katholieke Universiteit Leuven

Departement Elektrotechniek

ESAT-SISTA/TR 04-24

On the performance of decorrelation by prefiltering for adaptive feedback cancellation in Public Address systems[1]

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

January 2004
Published in *Proceedings of the 2004 IEEE Benelux Signal Processing Symposium (SPS 2004)*, Hilvarenbeek, the Netherlands, April 15-16, 2004, pp.167–170

[1]This report is available by anonymous ftp from *ftp.esat.kuleuven.ac.be* in the directory *pub/sista/vanwaterschoot/reports/04-24.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 State, Prime Minister's Office - Federal Office for Scientific, Technical and Cultural Affairs - Interuniversity Poles of Attraction Programme (2002-2007) - IUAP P5/22 ('Dynamical Systems and Control: Computation, Identification and Modelling') and P5/11 ('Mobile multimedia communication systems and networks'), 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.

ON THE PERFORMANCE OF DECORRELATION BY PREFILTERING FOR ADAPTIVE FEEDBACK CANCELLATION IN PUBLIC ADDRESS SYSTEMS

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

[1]toon.vanwaterschoot@esat.kuleuven.ac.be
[2]Katholieke Universiteit Leuven, ESAT-SCD, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium

ABSTRACT

The presence of an acoustic feedback path in Public Address (P.A.) systems limits the maximum allowable amplification if howling is to be avoided. Adaptive feedback cancellation (AFC) techniques aim at identifying the unknown feedback path, usually also by incorporating some form of decorrelation in the electroacoustic forward path. Recently, decorrelation by prefiltering was proposed for AFC in hearing aids, where the signals used for identification are filtered with the inverse source signal model. We apply this technique to closed-loop identification of room impulse responses. However, to obtain a useful estimate a much larger time window has to be used compared to the hearing aid case and the characteristics of the source signal may then vary significantly throughout the data set used. We propose a basis expansion technique to cope with this non-stationarity and we compare the proposed technique with an existing AFC method for P.A. systems.

1. INTRODUCTION

The acoustic feedback problem has been a subject of scientific research for four decades and more. Originally the problem was typically encountered in Public Address (P.A.) systems but with the advent of hearing aids, research has been focused on the latter application. The feedback problem can be illustrated by considering a 1-microphone/1-loudspeaker (SISO) setup, as depicted in **Figure 1**. Source signal $v(t)$ is picked up by the microphone, together with direct and reflected versions of the loudspeaker signal $u(t)$. We call $x(t) = F(q,t)u(t)$ the (unwanted) feedback signal, where $F(q,t)$ represents the unknown and possibly time-varying feedback path transfer function (with q the time shift operator: $q^{-1}u(t) = u(t-1)$). The composite microphone signal $y(t) = x(t) + v(t)$ is amplified and filtered in the forward path $G(q,t)$ before it is sent to the loudspeaker.

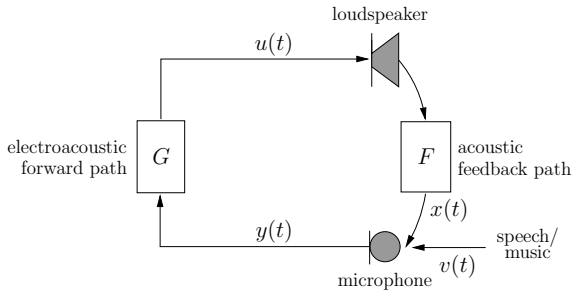


Figure 1: *The acoustic feedback problem in a SISO setup.*

Following Nyquist's stability criterion, the closed-loop system at hand will become unstable if a frequency ω exists for which

$$\begin{cases} |G(e^{i\omega})F(e^{i\omega})| & \geq & 1 \\ \angle G(e^{i\omega})F(e^{i\omega}) & = & n2\pi, n \in \mathbb{Z}. \end{cases} \quad (1)$$

Instability results in howling, and even a stable system operating close to instability (i.e. with a small gain margin) produces audible distortion due to ringing and excessive reverberation effects. Taking this into account, the maximum amplification that can be used without feedback cancellation, is limited to a level that is often insufficiently high.

Most feedback cancellation techniques aim at preventing one of the two conditions in (1) to be met for any frequency ω. In P.A. systems, equalization and notch filtering (either manually or automatically) are most commonly used to flatten the open-loop response $|G(e^{i\omega})F(e^{i\omega})|$. Another technique for smoothing $|G(e^{i\omega})F(e^{i\omega})|$ is to include some modulation device in the forward path $G(q,t)$, e.g. a frequency shifter [1],[2].

In hearing aids, a commonly used approach to avoid acoustic feedback is adaptive feedback cancellation (AFC), see e.g. [3]. As in acoustic echo cancellation (AEC), the loudspeaker and microphone signals are used to identify the unknown acoustic transfer function $F(q,t)$. Hence an estimate of the unwanted feedback signal $x(t)$ can be subtracted from the microphone signal $y(t)$. Unfortunately, for non-white source signals there is a strong correlation between the signals $u(t)$ and $v(t)$ which generally leads to a biased estimate of the feedback path, unless some form of decorrelation is used (see Section 2). In particular adaptive feedback cancellation in P.A. systems is a largely unexplored issue, except for the method proposed in [4],[5].

Recently, decorrelation by prefiltering $u(t)$ and $y(t)$ with the inverse source signal model was proposed for AFC in hearing aids [6],[7]. When $F(q,t)$ is a room transfer function (requiring a larger data set for identification), this technique is expected to

[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 State, Prime Minister's Office - Federal Office for Scientific, Technical and Cultural Affairs - Interuniversity Poles of Attraction Programme (2002-2007) - IUAP P5/22 ('Dynamical Systems and Control: Computation, Identification and Modelling') and P5/11 ('Mobile multimedia communication systems and networks'), 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.

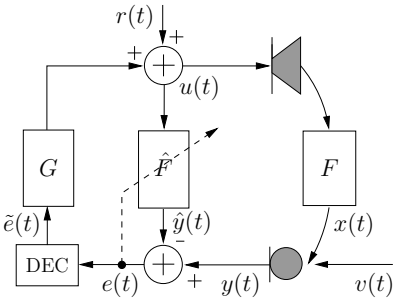


Figure 2: *Standard adaptive feedback cancellation (AFC) scheme with decorrelation in the forward path.*

perform worse due the non-stationarity typical to audio signals. We propose one way to cope with this, by expanding the time-varying model parameters onto a predefined set of basis functions (see Section 3). The proposed technique is compared to the method described in [5] based on computer simulations (see Section 4) and some observations are summarized in Section 5.

2. LEAST SQUARES CLOSED-LOOP IDENTIFICATION

Throughout this paper we will assume the acoustic feedback path $F(q)$ to be time-invariant with a finite impulse response $\mathbf{f} = \begin{bmatrix} f_0 & \dots & f_{n_F} \end{bmatrix}^T$ of known order n_F. This impulse response may be estimated at time t by solving the overdetermined system of equations:

$$\begin{bmatrix} u(t) & \dots & u(t-n_{\hat{F}}) \\ u(t-1) & \dots & u(t-n_{\hat{F}}-1) \\ \vdots & \ddots & \vdots \\ u(1) & \dots & u(-n_{\hat{F}}+1) \end{bmatrix} \cdot \begin{bmatrix} \hat{f}_0(t) \\ \vdots \\ \hat{f}_{n_{\hat{F}}}(t) \end{bmatrix} = \begin{bmatrix} y(t) \\ y(t-1) \\ \vdots \\ y(1) \end{bmatrix}$$

$$\Updownarrow$$

$$\mathbf{U}_{t\times(n_{\hat{F}}+1)}\hat{\mathbf{f}}(t)_{(n_{\hat{F}}+1)\times 1} = \mathbf{y}_{t\times 1}. \quad (2)$$

As model order selection will not be treated, we assume $n_{\hat{F}} = n_F$. The least squares (LS) solution to (2) is then given by

$$\hat{\mathbf{f}}_{\mathbf{LS}}(t) = (\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{y} \quad (3)$$

$$= \mathbf{f} + \underbrace{(\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{v}}_{\text{bias}} \quad (4)$$

with $\mathbf{v} \triangleq [v(t)\ v(t-1)\ \dots\ v(1)]^T$ representing the source signal vector. Since loudspeaker signal $u(t)$ is an amplified and filtered version of microphone signal $y(t)$, and $y(t)$ contains the source signal $v(t)$, the bias term $(\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{v}$ in (4) will generally not converge to zero as $t \to \infty$. Note that source signal $v(t)$ acts as a noise signal in the closed-loop identification.

2.1. Decorrelation in the forward path

A first solution to this consistency problem is to include some decorrelating signal operation in cascade with the electroacoustic forward path $G(q,t)$. In combination with an adaptive feedback canceller, i.e. an adaptive filter with input signal $u(t)$ and desired output signal $y(t)$, this yields the 'standard' AFC scheme as depicted in **Figure 2**. The decorrelation device ideally removes any correlation between loudspeaker signal $u(t)$ and source signal $v(t)$, such that the bias term in (4) disappears as $t \to \infty$. In practice this is attempted by decorrelating $u(t)$ from the adaptive filter error signal $e(t)$, which approximates $v(t)$ if $\hat{F}(q,t)$ converges to $F(q)$.
Several decorrelation techniques have been proposed in the AFC (and stereo AEC) literature, such as injecting a persistently exciting *probe signal* $r(t)$ at the loudspeaker, or adding an appropriate *delay* or *nonlinearity* (e.g. a frequency shift) to the forward path. Any one of these techniques (or any combination of them) allows a trade-off to be made between decorrelation on the one hand and audible signal distortion on the other hand, by tuning some parameter set. Signal distortion is unavoidable though, either because $\tilde{e}(t)$ is perceptually different from $e(t)$ (when strong decorrelation is applied) or because $e(t)$ is perceptually different from $v(t)$ due to a biased estimate $\hat{F}(q,t)$ (when weak decorrelation is applied). Hence an optimal parameter set for the decorrelating operation is hard to find.

2.2. Decorrelation in the identification algorithm

A second solution to this consistency problem is to use a model of the source signal in the identification procedure. Suppose we write the source signal as $v(t) = H(q,t)w(t)$ with $w(t)$ a stationary white noise signal and $H(q,t)$ a monic, stable, and inversely stable filter for all t. If we were able to consistently estimate $H(q,t)$, $\forall t$, an unbiased estimate of $F(q)$ could be obtained by *prefiltering* the loudspeaker and microphone signals $u(t)$ resp. $y(t)$ with a filter $L(q,t)$ equal to the inverse source signal model $H^{-1}(q,t)$ [6]. In this case an LS solution is sought to the modified system of equations:

$$\mathbf{L}_{t\times(t+n_L)}\mathbf{U}_{(t+n_L)\times(n_{\hat{F}}+1)}\hat{\mathbf{f}}(t) = \mathbf{L}_{t\times(t+n_L)}\mathbf{y}_{(t+n_L)\times 1} \quad (5)$$

where $\mathbf{U}$ and $\mathbf{y}$ are augmented by adding n_L rows at the bottom and $\mathbf{L}$ is a prefiltering matrix whose structure is defined by

$$\mathbf{L}_{t\times(t+n_L)} = \begin{bmatrix} l_0(t) & l_1(t) & \dots & 0 \\ 0 & l_0(t-1) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & l_{n_L}(1) \end{bmatrix}. \quad (6)$$

It is assumed here that the inverse source signal model $H^{-1}(q,t)$ has a finite impulse response, which is generally true for finite order autoregressive (AR) models of the source signal (see Section 3). The LS solution to the modified system (5) is given by

$$\hat{\mathbf{f}}_{\mathbf{LS},L}(t) = (\mathbf{U}^T\mathbf{L}^T\mathbf{L}\mathbf{U})^{-1}\mathbf{U}^T\mathbf{L}^T\mathbf{L}\mathbf{y} \quad (7)$$

$$= \mathbf{f} + \underbrace{(\mathbf{U}^T\mathbf{L}^T\mathbf{L}\mathbf{U})^{-1}\mathbf{U}^T\mathbf{L}^T\mathbf{L}\mathbf{v}}_{\text{bias}} \quad (8)$$

Now if $L(q,t)$ is a consistent estimate of $H^{-1}(q,t)$, $\forall t$, then $\mathbf{L}\mathbf{v} = \mathbf{w}$ where $\mathbf{w} \triangleq [w(t)\ w(t-1)\ \dots\ w(1)]^T$ is a white noise signal vector. The LS solution (8) will then be unbiased since every sample $w(t)$ of the stationary white noise sequence in $\mathbf{w}$ is uncorrelated with all samples of the prefiltered loudspeaker signal $L(q,t)u(t)$ up to and including time t (assuming that $G(q,t)$ contains at least one pure time delay).
The setup for performing the above closed-loop system identification procedure is shown in **Figure 3**. The coefficients of the identified impulse response $\hat{\mathbf{f}}_{\mathbf{LS},L}$ can be copied to the cancellation filter $F_0(q,t)$ at regular time intervals (not necessarily at

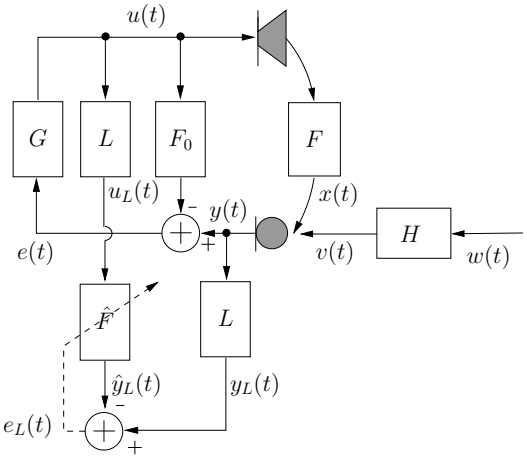


Figure 3: *Adaptive feedback cancellation scheme with decorrelation by prefiltering the loudspeaker and microphone signals.*

every time instant). The advantage of performing decorrelation by prefiltering is that one can obtain a consistent estimate of the feedback path $F(q)$ without distorting the signals in the closed-loop system. In other words it is possible to achieve the overall desired outcome that $u(t)$ is a consistent estimate of $G(q,t)v(t)$.

3. TIME-VARYING AUTOREGRESSIVE MODELLING

Audio signals are often modelled as autoregressive (AR) processes, hence our choice of imposing an AR model structure on the source signal model:

$$H(q,t) = \frac{1}{1 + a_1(t-1)q^{-1} + \ldots + a_p(t-p)q^{-p}} \qquad (9)$$

Note the explicit time-dependence of the AR coefficients in (9), which reflects the non-stationarity of a typical audio spectrum. In [6] this non-stationarity is ignored and $L(q)$ is implemented as a fixed high-pass prefilter, representing an average inverse speech spectrum. Although some decorrelation can be achieved, this method does not yield an unbiased estimate of $F(q)$.

Another approach would consist in estimating the inverse source signal model $H^{-1}(q,t)$ adaptively from the closed-loop data $\{u(k), y(k)\}_{k=1}^{t}$. Concurrent estimation of $F(q)$ and $H^{-1}(q,t)$ generally leads to the minimization of a non-linear cost function. In [7] this difficulty is avoided by reformulating the problem as a two-channel adaptive filtering problem, in which one adaptive filter aims at modelling $H^{-1}(q,t)$ and the second one at modelling $-H^{-1}(q,t)F(q)$. This allows for a time-varying source signal model, although it is assumed that $H^{-1}(q,t)$ does not vary significantly throughout the time window used for identification. This is a reasonable assumption for hearing aid applications since the length of the feedback path impulse response is then typically $10ms$ or less and a speech signal can be considered locally stationary for time windows with an average length of $20ms$. Given that a typical acoustic impulse response in a small office is at least $100ms$ long, it can be seen that the above method cannot be applied when dealing with room acoustics.

To obtain a time-invariant parameter set that can represent an audio spectrum over a sufficiently long time window, we propose to apply time-varying autoregressive (TVAR) modelling using a basis expansion method (BEM). Following [8], the idea is to expand each of the TVAR coefficients $a_i(t-i)$ (from (9)) onto a set of $m+1$ predefined basis functions $f_l(t-i)$:

$$a_i(t-i) = \sum_{l=0}^{m} a_{il} f_l(t-i) \quad i = 1, \ldots, p \qquad (10)$$

Depending on the kind and the number of basis functions used and depending on the frequency of the TVAR coefficients' evolution, the time-invariant parameter set

$$\theta_{H^{-1}} \triangleq \begin{bmatrix} a_{10} & \ldots & a_{1m} & \ldots & a_{p0} & \ldots & a_{pm} \end{bmatrix}^T \qquad (11)$$

can be used to describe the dynamics of the source signal over a time window of a certain length N_{win}. The problem of concurrently estimating $F(q)$ and $H^{-1}(q,t)$ from $\{u(k), y(k)\}$ with $k \in (t - N_{win}, t]$ is then reduced to concurrent estimation of $F(q)$ and $\theta_{H^{-1}}$ from the expanded data set $\{U(k), Y(k)\}$ containing the loudspeaker and microphone signals, expanded onto the given basis:

$$U(k) \triangleq \begin{bmatrix} f_0(k)u(k) \\ \vdots \\ f_m(k)u(k) \end{bmatrix}, \; Y(k) \triangleq \begin{bmatrix} f_0(k)y(k) \\ \vdots \\ f_m(k)y(k) \end{bmatrix} \qquad (12)$$

Deriving an algorithm for concurrent estimation of $F(q)$ and $\theta_{H^{-1}}$ is beyond the scope of this paper, but the performance of prefiltering with TVAR coefficients identified from the basis expanded source signal $V(k) \triangleq [f_0(k)v(k) \; \ldots \; f_m(k)v(k)]^T$ is evaluated in the next section.

4. SIMULATIONS

MATLAB simulations were conducted to illustrate the performance of different AFC techniques described above. At a sampling rate $f_s = 8kHz$, $F(q)$ was a fixed, artificially generated room impulse response of length $n_F + 1 = 1000$. $G(q)$ consisted of a unit time delay, cascaded with a constant, linear amplification factor. This factor was determined for each experiment such that the average signal-to-noise ratio (SNR) at the microphone

$$SNR_{mic} = \frac{\sum_{k=1}^{N} |x(k)|^2}{\sum_{k=1}^{N} |v(k)|^2} \qquad (13)$$

was constant throughout different experiments (with $N = 24000$ the number of data points used). Typical microphone sensitivity and loudspeaker saturation functions were included in the simulated closed-loop system. The experiments performed closed-loop identification *without* subtracting the estimated feedback signal from the microphone signal. To avoid system instability, a gain margin of $3dB$ or more was maintained throughout all experiments. Recursive identification was performed using an exponentially windowed RLS algorithm with forgetting factor $\lambda = 0.9997$ (which corresponds to a data memory of roughly $3(n_F + 1) = 3000$ samples). When prefiltering was used, the source signal $v(t)$ *was assumed to be known* and the AR model order was set to $p = 12$. Although in practice $v(t)$ is unknown, a consistent estimate $e(t)$ of the source signal may be available when a cancellation filter $F_0(q,t) = \hat{F}(q,t)$ is included and provided that $\hat{F}(q,t)$ converges to $F(q)$. The performance measure used for comparison was the normalized bias, defined as

$$\delta(t) = 20 \log_{10} \frac{\|\hat{\mathbf{f}}_{\mathbf{LS}}(t) - \mathbf{f}\|}{\|\mathbf{f}\|} \qquad (14)$$

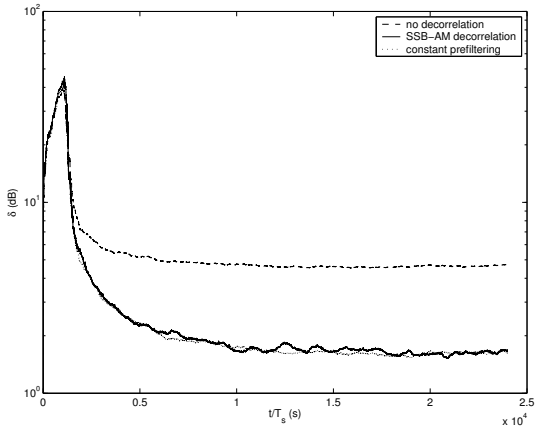


Figure 4: *Convergence curves for a stationary AR source signal.*

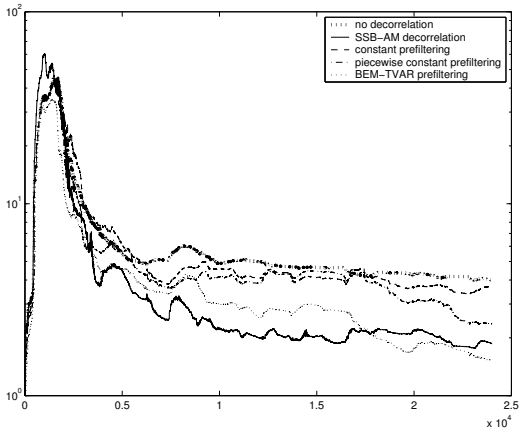


Figure 5: *Convergence curves for a true speech source signal.*

The first source signal we applied was a p-th order stationary AR sequence exhibiting a typical long-term speech spectrum. **Figure 4** shows the convergence plots for three different identification methods. Identification without decorrelation obviously performs poorly. Decorrelation by frequency shifting (which is implemented as an SSB-AM modulation as described in [5]) shows a performance comparable to the performance of decorrelation by constant prefiltering. The prefilter used was equal to the exact inverse source signal model. Note that the normalized bias is excessively large even after convergence. This is due to the low SNR at the microphone (typically $SNR_{mic} \approx -20dB$), which for stability reasons can only be increased if a cancellation filter $F_0(q,t)$ is included.

The second source signal we applied was a true speech signal. Five different identification methods were used, as shown in **Figure 5**. Again, identification without decorrelation shows poor performance, as does prefiltering with a constant prefilter equal to the inverse long-term spectrum of the speech signal. To have an idea on how the two-channel adaptive filtering scheme of [7] would perform in this situation, the speech signal was divided in frames of $N_{win} = 3(n_F + 1) = 3000$ samples. An AR model was fitted to each frame, and identification with prefiltering was performed keeping the prefilter coefficients constant during 3000 samples and equal to the identified AR coefficients. This piecewise constant prefiltering technique performs only slightly better than the previous one. Decorrelation by SSB-AM frequency shifting shows a performance comparable to the performance of prefiltering with the BEM-TVAR method proposed in Section 3. A Fourier basis with 51 basis functions was used to estimate the time-invariant BEM-TVAR coefficients on non-overlapping frames of $N_{win} = 3000$ samples. We note that in a similar experiment with $F(q)$ being a 100-tap hearing aid impulse response, the piecewise constant prefiltering technique performed similar to the SSB-AM and BEM-TVAR prefiltering techniques, which illustrates our comment in Section 3 on the influence of the feedback path length when performing identification with speech signals.

5. CONCLUSION

From a system identification point of view, prefiltering closed-loop data with the inverse source signal model is probably the most straightforward way of obtaining a consistent estimate of the feedback path transfer function in an acoustic feedback situation. In practice, this method is hard to apply in a correct way because the source signal model is unknown and time-varying, and thus should be estimated concurrently with the feedback path transfer function. Nevertheless, initial simulations indicate that —if the inverse source signal model can be estimated consistently— the method performs as well as decorrelation techniques applied in the forward path, with the additional advantage of not causing any signal distortion.

6. REFERENCES

[1] M. R. Schroeder, "Improvement of acoustic-feedback stability by frequency shifting," *J. Acoust. Soc. Amer.*, vol. 36, no. 9, pp. 1718–1724, Sept. 1964.

[2] J. L. Nielsen and U. P. Svensson, "Performance of some linear time-varying systems in control of acoustic feedback," *J. Acoust. Soc. Amer.*, vol. 106, no. 1, pp. 240–254, July 1999.

[3] J. A. Maxwell and P. M. Zurek, "Reducing acoustic feedback in hearing aids," *IEEE Trans. Speech Audio Processing*, vol. 3, no. 4, pp. 304–313, July 1995.

[4] S. Kamerling, K. Janse, and F. van der Meulen, "A new way of acoustic feedback suppression," in *Preprints 104th AES Convention*, Amsterdam, The Netherlands, May 16–19 1998.

[5] C. P. Janse and P. A. A. Timmermans, "Signal amplifier system with improved echo cancellation," U.S. Patent 5,748,751, May 1998.

[6] J. Hellgren, "Analysis of feedback cancellation in hearing aids with Filtered-X LMS and the direct method of closed-loop identification," *IEEE Trans. Speech Audio Processing*, vol. 10, no. 2, pp. 119–131, Feb. 2002.

[7] A. Spriet, M. Moonen, and I. Proudler, "Feedback cancellation in hearing aids: an unbiased modelling approach," in *Proc. European Signal Processing Conf.*, Toulouse, France, Sept. 2002, EUSIPCO-02, pp. 531–534.

[8] Y. Grenier, "Time-dependent ARMA modeling of nonstationary signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 31, pp. 899–911, Aug. 1983.