Available online at www.sciencedirect.com
ScienceDirect
Signal Processing 88 (2008) 571–581
SIGNAL PROCESSING
www.elsevier.com/locate/sigpro

Generalized sidelobe canceller based combined acoustic feedback- and noise cancellation

Geert Rombouts*, Ann Spriet, Marc Moonen

Katholieke Universiteit Leuven/ESAT-SCD, Kasteelpark Arenberg 10, 3001 Heverlee, Belgium

Received 16 March 2007; received in revised form 14 August 2007; accepted 31 August 2007
Available online 11 September 2007

Abstract

We propose a combination of the well-known generalized sidelobe canceller (GSC) or Griffiths–Jim beamformer, and the so-called PEM-AFROW algorithm for joint estimation under closed loop conditions of a room impulse response and a desired speech signal model, resulting in a system for multimicrophone combined acoustic feedback and noise cancellation. For specific applications (e.g. public address systems), the computational complexity may be reduced dramatically compared to state-of-the-art proactive acoustic feedback cancellers, while feedback cancellation performance is only marginally degraded.
© 2007 Elsevier B.V. All rights reserved.

Keywords: Generalized sidelobe canceller; Acoustic feedback cancellation; Whitening

1. Introduction

Acoustic feedback is a well-known phenomenon that appears in systems that have an electro-acoustic closed loop. Such systems (Fig. 1) consist of a microphone, a forward path gain g and delay q^{-D} (remark: the delay operator q^{-1} is used to denote a unit delay in the time domain, as is often done in system identification literature, while z^{-1} is a unit delay in the z-domain), a loudspeaker and a room impulse response (RIR) $\mathbf{f}_r$. Hearing aids (HA) and public address (PA) systems are typical examples. If the loop gain exceeds unity for frequencies ω_i where the loop phase is $2n_i\pi$ radians (with n_i integer), system instability becomes audible as a loud 'howling' sound (the Larsen-effect). In such applications, background *noise* is also picked up by the microphones, and degrades the signal quality. Examples are babble noise in the vicinity of the speaker, or cabin noise in a train.

Many approaches can be found in literature for acoustic feedback cancellation as well as for noise cancellation.

Most of the *acoustic feedback cancellation* techniques that have been derived up till now are single channel techniques [1–5] (although some multichannel examples exist [6,7]). Traditional approaches are mostly 'reactive', as they allow the system to become unstable in order to then identify the frequencies where acoustic feedback occurs, and introduce notch filters for these frequencies into the signal path. More recent approaches [8–10] are 'proactive' and do not introduce signal distortion, as they are based on an adaptive filter that models

*Corresponding author.
E-mail address: geert.rombouts@esat.kuleuven.be (G. Rombouts).

0165-1684/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2007.08.018

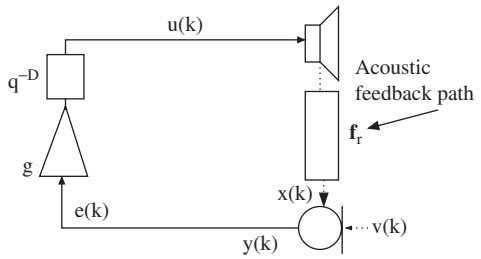

Fig. 1. An electro-acoustic loop.

the RIR, and insert a so-called controller into the scheme, which effectively removes the component from the microphone signal that stems from the loudspeaker. Such approaches are based on acoustic echo cancellation procedures, where an additional signal modelling is required to avoid a biased RIR estimate due to the correlation between the desired speech signal (''near-end signal'') and the (''far-end'') loudspeaker signal (indeed meant to be a processed (amplified) version of the speech signal). The PEM-AFROW algorithm of [8], applies signal model based prewhitening and effectively achieves an unbiased estimate. It is a prediction error method, where the whitening process performs row-operations on a least-squares system of equations, hence the name. Modelling the RIR may, however, be computationally expensive, especially in the PA context, and therefore we will concentrate on a simpler approach.

Good *noise cancellation methods* typically rely on multiple microphones (microphone arrays) providing spatial information on the location of the desired speech source and the noise source. Single channel approaches can only use spectral information to discriminate between noise and speech, and their performance is inferior to multichannel methods (e.g. they are often prone to a phenomenon called 'musical noise', although remedies exist for this [11]).

In several applications (e.g. HA and PA again), it is desirable to combine feedback- and noise cancellation. A simple cascading of both systems, even for the single channel case, does not lead to an optimal solution (see Section 4.1). It is also expected that PEM-AFROW based feedback cancellation schemes could benefit from additional multichannel processing. In this paper, we first derive a PEM-AFROW based proactive multichannel feedback cancellation algorithm which does not introduce signal distortion, and which especially for PA applications exhibits a dramatically reduced computational complexity compared to existing adaptive filtering based proactive feedback cancellation schemes. We show that this scheme is robust to noise, and in certain circumstances even provides some noise reduction. Then we extend this setup by adding multiple noise references, such that it exhibits both feedback- and noise cancellation behaviour in low-reverberant environments.

In Section 2, the Griffiths–Jim (or GSC) based multichannel noise reduction scheme is reviewed. In Section 3, it is shown that for white input signals, a GSC scheme can already be used for feedback cancellation. When signal model based whitening filters are added, it can also provide feedback cancellation for (non-white) speech input signals. In Section 4, we first discuss the problems inherent to a simple cascading of a noise-reduction and a feedback-reduction scheme, and then proceed by stating and analyzing a minimization problem corresponding to the GSC-based combined noise- and feedback cancellation scheme, and describing an implementation (the PEM-AFROW algorithm). In Section 5, a number of simulations are shown for different scenarios, which prove the effectiveness of the new scheme. Conclusions are given in Section 6.

2. Noise reduction

A traditional approach to multichannel noise reduction is the so-called Griffiths–Jim beamformer [12] or generalized sidelobe canceller (GSC), see Fig. 2, which we briefly review in this section. For the time being, we assume a simple scenario with one desired speech source $v(k)$ and one noise source $x(k)$. The input vector of the M-microphone array can be written as

$$\mathbf{y}(k) = \mathbf{v}(k) + \mathbf{x}(k),$$

$$\mathbf{y}(k) = \begin{pmatrix} \mathbf{y}^{(1)}(k) \\ \mathbf{y}^{(2)}(k) \\ \vdots \\ \mathbf{y}^{(M)}(k) \end{pmatrix},$$

i.e. as a stacked vector of the input vectors for microphones $1 \ldots M$,

$$\mathbf{y}^{(i)}(k) = \begin{pmatrix} y^{(i)}(k) \\ y^{(i)}(k-1) \\ \vdots \\ y^{(i)}(k-N_m+1) \end{pmatrix}.$$

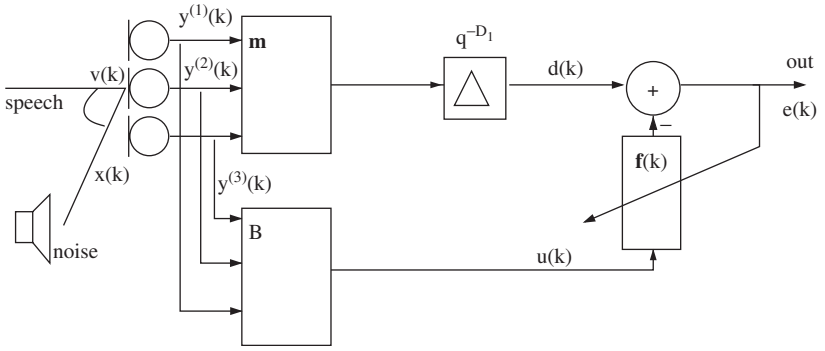

Fig. 2. Griffiths–Jim beamformer. Signal $u(k)$ does not contain a component correlated with the broadside speech signal $v(k)$, and hence forms a noise reference. Correlated noise is removed from $d(k)$ by means of an adaptive filter.

Vectors $\mathbf{v}(k)$ and $\mathbf{x}(k)$ have a similar structure as $\mathbf{y}(k)$. Vector $\mathbf{x}(k)$ is the component in the microphone signal which stems from the noise source $x(k)$ and vector $\mathbf{v}(k)$ stems from the desired speech source $v(k)$.

A fixed beamformer $\mathbf{m}$ with filter length N_m per channel, produces a speech reference signal $d(k)$ by "zooming in" on the desired speech signal source $v(k)$. It already suppresses some noise coming from directions other than the direction of the speech signal, as well as diffuse noise. This means $d(k)$ is a signal which contains both a speech- and a residual noise component. A blocking matrix B is chosen such that it suppresses the desired speech signal and hence creates a noise reference $u(k)$. Note that for the time being, the blocking matrix is a row *vector* creating a single noise reference, while further on we will use a blocking *matrix* creating multiple references. An adaptive filter $\mathbf{f}(k)$ then removes the component in $d(k)$ which is correlated with $u(k)$. This can be shown to be equivalent with implementing a beam-pattern on the sensor array which has a zero in the direction of the noise source $x(k)$.

The signal $d(k)$ can be written as

$$d(k) = \mathbf{m}^{\mathrm{T}}\mathbf{v}(k - D_1) + \mathbf{m}^{\mathrm{T}}\mathbf{x}(k - D_1),$$

and assuming a perfect blocking matrix, $u(k)$ can be written as

$$u(k) = B\mathbf{x}_b(k), \tag{1}$$

where $\mathbf{x}_b(k)$ is an input vector with a length which may differ from the length of $\mathbf{x}(k)$. The adaptive filter $\mathbf{f}(k)$ with length N_f should converge to the solution of $\min_{\mathbf{f}} \varepsilon\{\left\|d(k) - \mathbf{f}^{\mathrm{T}}\mathbf{u}(k)\right\|^2\}$, with

$$\mathbf{u}(k) = \begin{pmatrix} u(k) \\ u(k-1) \\ \vdots \\ u(k - N_f + 1) \end{pmatrix}.$$

Hence, upon convergence,

$$\mathbf{f}(k) = E\{\mathbf{u}(k)\mathbf{u}^{\mathrm{T}}(k)\}^{-1} E\{\mathbf{m}^{\mathrm{T}}\mathbf{x}(k)\mathbf{u}(k)\}.$$

The output in Fig. 2 is then

$$e(k) = \mathbf{m}^{\mathrm{T}}\mathbf{v}(k - D_1) + \mathbf{m}^{\mathrm{T}}\mathbf{x}(k - D_1) - \mathbf{f}^{\mathrm{T}}(k)\mathbf{u}(k). \tag{2}$$

If (as was assumed in (1)), no signal leakage occurs through B, that is if the signal $u(k)$ does not contain a component correlated with $v(k)$, then $e(k)$ contains an undistorted version of $\mathbf{m}^{\mathrm{T}}\mathbf{v}(k)$. This means the GSC approach works better in less reverberant environments, such that a good B can be found. With a sufficiently long $\mathbf{f}(k)$, $e(k) = \mathbf{m}^{\mathrm{T}}\mathbf{v}(k)$ is then obtained.

3. GSC-based feedback cancellation

We propose to use a GSC-like structure for acoustic feedback cancellation. In the electro-acoustic loop, the microphone signal is amplified, and emitted by a loudspeaker in the same room. Let the signal $x(k)$ of the previous section now represent the loudspeaker signal, instead of the usual noise signal. A GSC-like structure should then steer a zero in the direction of the loudspeaker in order to provide acoustic feedback cancellation. It is obvious that in a traditional GSC scheme as in Fig. 2, the

reference signal $u(k)$ would then be correlated with the desired signal component $v(k)$ and this would result in signal leakage, and hence in signal distortion or even cancellation.

Still referring to Fig. 2, let us first assume that $v(k)$ is a white noise signal instead of a speech signal (further on we will drop this assumption), that $\mathbf{m}$ contains FIR filters of length N_m, and finally that B is a perfect blocking matrix, which means that $v(k)$ does not leak directly into the signal $u(k)$. If a delay $D_2 > N_m$ is inserted into the loop, then signal $x(k)$ is not correlated with $\mathbf{m}^{\mathrm{T}}\mathbf{v}(k - D_1)$. The only component in $d(k)$ which is correlated with $u(k)$ (and which will hence be removed by the adaptive filter) is $\mathbf{m}^{\mathrm{T}}\mathbf{x}(k)$, hence $e(k) = \mathbf{m}^{\mathrm{T}}\mathbf{v}(k)$ can again be obtained with a sufficiently long $\mathbf{f}(k)$. This shows that for a white noise input signal, in an environment with low reverberation (which means that a good B can be found), feedback cancellation can be obtained by using a GSC.

This simple observation can be of use in e.g. a PA application where the geometry of the loudspeaker/microphone array setup is fixed. The adaptive filter can then be trained with a white noise sequence, and kept fixed from then on. On the other hand, it may be of interest to be able to adapt the system to changing acoustical environments. Hence we derive an algorithm which uses the (speech) signal to perform continuous training of the adaptive filter.

For a (non-white) speech signal $v(k)$, we propose the use of prewhitening filters, as shown in Fig. 3, and adopt a procedure based on the PEM-AFROW algorithm [8]. Here, the updating of the adaptive filter $\mathbf{f}(k)$ is performed with prewhitened versions of the signals $d(k)$ and $u(k)$. The prewhitening filter is a linear prediction error filter $\mathbf{a}(k)$, and the coefficients of $\mathbf{a}(k)$ are estimated together with $\mathbf{f}(k)$. The resulting adaptive filter coefficients $\mathbf{f}(k)$ are copied to the so-called 'controller' $\mathbf{f}_0(k)$ (dashed arrow).

With the assumption that a speech signal can be modelled as a time variant autoregressive process (TVAR), it can be shown that the scheme including prewhitening filters corresponds to a scheme without prewhitening filters, but then driven by a white noise input. This means that this scheme can be used for feedback suppression when the prewhitening filters $\mathbf{a}(k)$ can be estimated jointly with the adaptive filter $\mathbf{f}(k)$. This is accomplished by the PEM-AFROW algorithm and further analyzed in Section 4.2. For further details on the PEM-AFROW algorithm, we refer to [8].

4. Combined feedback- and noise cancellation

4.1. Cascaded schemes

It is desirable to use proactive techniques in the acoustic feedback cancellation context. These techniques typically involve an adaptive filter structure which models the acoustic feedback path even under conditions in which the loop gain is too low for the howling phenomenon to occur. Inspired by acoustic

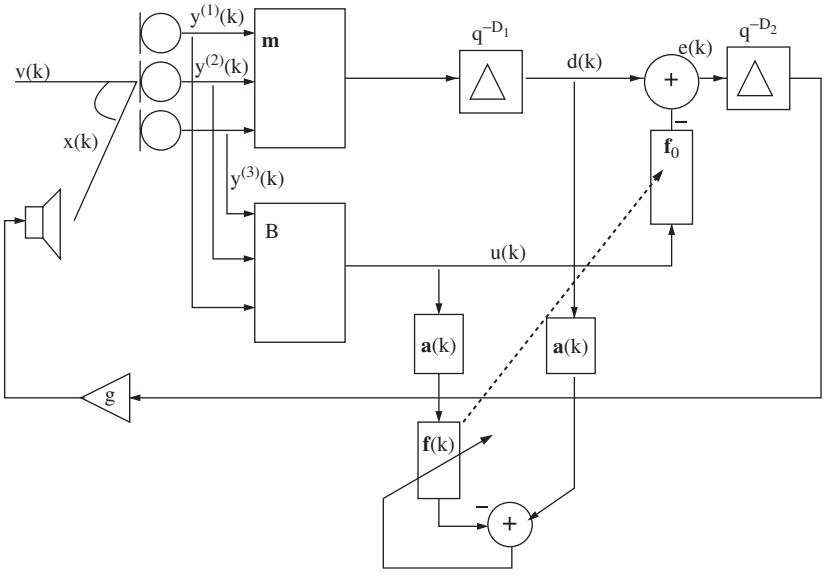

Fig. 3. New GSC-based scheme for acoustic feedback suppression.

echo cancellation, a so-called controller $\hat{\mathbf{f}}_0$ is then used (see Figs. 4 and 5) which models the RIR $\mathbf{f}_r$. This effectively removes the feedback loop, and oscillations cannot occur. For these schemes, simply cascading the feedback cancellation with a noise reduction scheme often does not work.

In a single microphone system with a single channel PEM-AFROW- or PEM-AFC [10] based acoustic feedback cancellation, the combination can be performed in two ways: either the noise reduction is performed first on the microphone signal, see Fig. 4 or the feedback cancellation is performed first, as shown in Fig. 5 When noise reduction is performed first, the noise reduction filter is included in the feedback path which has then to be modelled by the feedback canceller. Since in a single channel noise reduction scheme only highly time varying spectral differences between speech and noise are used to remove noise, the noise reduction filter will also be highly time varying, and it will be impossible for the feedback canceller to track these changes.

If on the other hand feedback reduction is performed first, another problem arises. Most single channel noise reduction techniques implement a frequency-dependent gain function, which is small on frequency intervals with a low signal-to-noise ratio. It is important to note that the (frequency dependent) loop gain of the closed loop system is independent of the signal-to-noise ratio: the latter is a feature of the signal, while the former is a feature of the system. The adaptive filter models the RIR, and hence the near-end speech+noise–signal may be considered as measurement noise for this identification process. In frequency intervals where the signal-to-noise ratio is considered low by the noise reduction algorithm, the gain function will be small, and hence the ratio of the measurement signal (the loudspeaker signal which arrives at the microphone after being filtered by the RIR) and the measurement noise (the near-end signal) becomes small. This will lead to a large variance on the estimate of the RIR, and thus to system instability. Briefly, the noise reduction is found to create adverse conditions for the feedback cancellation.

In multimicrophone cascading schemes, similar problems arise. If multichannel noise cancellation is performed first on several microphone signals, the feedback canceller again has to model the noise reduction filter, which complicates the feedback path estimation. The noise reduction filter may be less time varying than in the single-channel case, because it can now also take (more time invariant) spatial information into account.

If on the other hand feedback cancellation is first applied on each of the microphone signals, i.e. before the noise reduction algorithm, the measurement signal required for the feedback canceller will become too small compared to the measurement noise in low signal-to-noise frequency intervals.

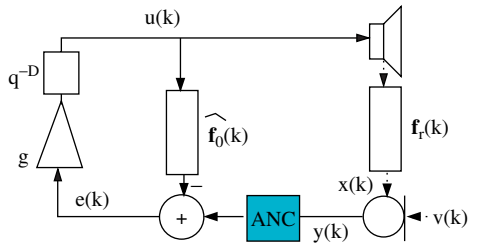

Fig. 4. First noise—then feedback cancellation.

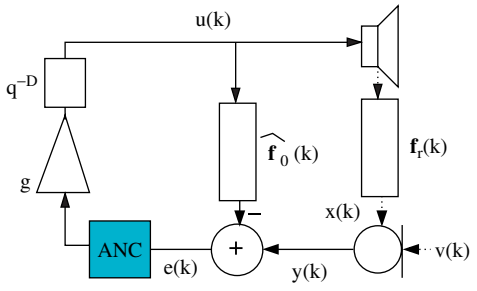

Fig. 5. First feedback—then noise cancellation.

4.2. GSC-based combined scheme

The scheme we propose in this paper is shown in Fig. 6. The main idea is that in a low-reverberant environment, noise reduction can be performed by the steering vector $\mathbf{m}$ for non-directional sources, and additionally by steering zeros in the direction of directional noise sources, while feedback cancellation can be performed by adaptively steering a zero in the direction of the loudspeaker. In a traditional GSC-based noise canceller, multiple adaptive zeros can be steered if multiple independent reference channels (multiple outputs of blocking matrix B) are available. For M microphones, $M-1$ independent noise references can be created. Note that if multiple loudspeakers are present, all producing the same signal, the scheme will be able to cancel all of them based on one feedback reference signal.

In the sequel we will give a design criterion for the filters $\mathbf{a}(k)$ and $\mathbf{f}(k)$ in Fig. 6 and we will demonstrate that this criterion leads to the desired

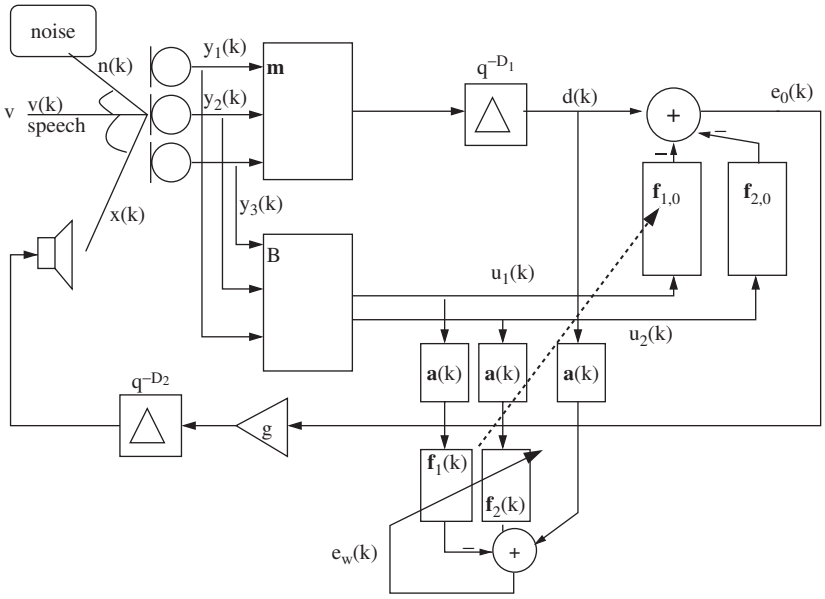

Fig. 6. GSC-based scheme for combined acoustic feedback and noise reduction.

feedback and noise cancellation solution in scenarios with M microphones and $M-2$ noise sources. We will assume stationarity of the input signal, which results in $\mathbf{a}(k)=\mathbf{a}$ being time invariant so that we can use z-transform notation for the derivation. Simulations show though that also for non-stationary input signals the scheme provides proper feedback- and noise cancellation. In order to avoid trivial solutions, we use a monic speech model, which means

$$\mathbf{a}=\begin{pmatrix}1\\ a_1\\ \vdots\\ a_P\end{pmatrix},$$

where P is the model order. Define

$$\mathbf{H}(z)=\begin{pmatrix}H^{(1)}(z)\\ H^{(2)}(z)\\ \vdots\\ H^{(M)}(z)\end{pmatrix},$$

where $H^{(i)}(z)$ is the transfer function of the loudspeaker to the ith microphone. Similarly $\mathbf{H}_v(z)$ is the transfer vector from the signal source to the microphone array and $\mathbf{H}_n^{(j)}(z)$ is defined as a transfer vector from the jth noise source to the microphone array. The noise transfer matrix

$$\mathbf{H}_n(z)=(\mathbf{H}_n^{(1)}\ \mathbf{H}_n^{(2)}\ \ldots\ \mathbf{H}_n^{(M-2)})$$

contains transfer vectors from $M-2$ noise sources to M microphones. The $M-2$ independent noise sources can be written as a vector

$$\mathbf{N}(z)=\begin{pmatrix}N_1(z)\\ N_2(z)\\ \vdots\\ N_{M-2}(z)\end{pmatrix}.$$

Beamformer $\mathbf{M}(z)$ is a $1\times M$-element vector, and blocking matrix $\mathbf{B}(z)$ has M columns, and $M-1$ rows, one for each reference signal. We assume that the response of $\mathbf{M}(z)$ in the direction of the signal of interest is designed to be unity,

$$\mathbf{M}^{\mathrm{T}}(z)\mathbf{H}_v(z)=1, \tag{3}$$

and that the blocking matrix is perfect (no signal leakage),

$$\mathbf{B}(z)\mathbf{H}_v(z)=\begin{pmatrix}0\\ \vdots\\ 0\end{pmatrix}. \tag{4}$$

The vector

$$\mathbf{Y}(z) = \begin{pmatrix} Y^{(1)}(z) \\ Y^{(2)}(z) \\ \vdots \\ Y^{(M)}(z) \end{pmatrix}$$

is the z-transform of the microphone array input signal. The beamformer and blocking matrix outputs are

$$D(z) = \mathbf{M}^{\mathrm{T}}(z)\mathbf{Y}(z)z^{-D_1},$$

$$\mathbf{U}(z) = \begin{pmatrix} U_1(z) \\ U_2(z) \\ \vdots \\ U_{M-1}(z) \end{pmatrix} = \mathbf{B}(z)\mathbf{Y}(z).$$

Define

$$\mathbf{F}_0(z) = \begin{pmatrix} F_{1,0}(z) \\ F_{2,0}(z) \\ \vdots \\ F_{M-1,0}(z) \end{pmatrix},$$

$$\mathbf{F}(z) = \begin{pmatrix} F_1(z) \\ F_2(z) \\ \vdots \\ F_{M-1}(z) \end{pmatrix}.$$

Then we have

$$E_0(z) = D(z) - \mathbf{U}^{\mathrm{T}}(z)\mathbf{F}_0(z), \tag{5}$$

$$E(z) = D(z) - \mathbf{U}^{\mathrm{T}}(z)\mathbf{F}(z), \tag{6}$$

and

$$\mathbf{Y}(z) = \mathbf{H}(z)E_0(z)z^{-D_2} + \mathbf{H}_v(z)V(z) + \mathbf{H}_n(z)\mathbf{N}(z).$$

The whitened residual signal is

$$E_w(z) = A(z)E(z).$$

The minimization problem that will be solved is now

$$\min_{A,\mathbf{F}} \varepsilon\{\|E_w(z)\|^2\}, \tag{7}$$

with $\varepsilon\{\}$ the expectation operator, and where the norm is defined as

$$\varepsilon\{\|G(z)\|^2\} = \varepsilon\{|g(k)|^2\}.$$

We have

$$E_w(z) = A(z)\mathbf{Q}(z)(I - gz^{-D_2}\mathbf{H}(z)\mathbf{Q}_0(z))^{-1} \times(\mathbf{H}_v(z)V(z) + \mathbf{H}_n(z)\mathbf{N}(z)), \tag{8}$$

$$\mathbf{Y}(z) = (I - gz^{-D_2}\mathbf{H}(z)\mathbf{Q}_0(z))^{-1}(\mathbf{H}_v(z)V(z) + \mathbf{H}_n(z)\mathbf{N}(z)), \tag{9}$$

where

$$\mathbf{Q}(z) = z^{-D_1}\mathbf{M}^{\mathrm{T}}(z) - \mathbf{F}^{\mathrm{T}}(z)\mathbf{B}(z),$$

$$\mathbf{Q}_0(z) = z^{-D_1}\mathbf{M}^{\mathrm{T}}(z) - \mathbf{F}_0^{\mathrm{T}}(z)\mathbf{B}(z).$$

$\mathbf{Q}_0(z)$ can be interpreted as the open loop transfer vector from the microphones to the error signal, which is equal to $\mathbf{Q}(z)$ iff $\mathbf{F}_0(z) = \mathbf{F}(z)$.

From the matrix inversion lemma, we have

$$\begin{aligned}(I - gz^{-D_2}\mathbf{H}(z)\mathbf{Q}_0(z))^{-1} &= I + gz^{-D_2}\mathbf{H}(z)(1 - \mathbf{Q}_0(z)gz^{-D_2}\mathbf{H}(z))^{-1}\mathbf{Q}_0(z) \\ &= I + \frac{1}{(1 - \mathbf{Q}_0(z)gz^{-D_2}H)}gz^{-D_2}\mathbf{H}(z)\mathbf{Q}_0(z).\end{aligned}$$

Substituting this into (8) results in

$$E_w(z) = A(z)\mathbf{Q}(z)\left(I + \frac{1}{(1 - \mathbf{Q}_0(z)gz^{-D_2}\mathbf{H}(z))} \times gz^{-D_2}\mathbf{H}(z)\mathbf{Q}_0(z)\right)(\mathbf{H}_v(z)V(z) + \mathbf{H}_n(z)\mathbf{N}(z)).$$

We assume the desired (speech) signal satisfies a time invariant model $V(z) = (1/\tilde{A}(z))W(z)$ where $W(z)$ is a white Gaussian noise process and $\tilde{A}(z)$ is a Pth order, monic polynomial (cfr. $A(z)$). Substituting the signal model gives

$$\begin{aligned}E_w(z) = \frac{A(z)}{\tilde{A}(z)}\mathbf{Q}(z)&\left(I + \frac{gz^{-D_2}\mathbf{H}(z)\mathbf{Q}_0(z)}{(1 - \mathbf{Q}_0(z)gz^{-D_2}\mathbf{H}(z))}\right) \times\mathbf{H}_v(z)W(z) \\ &+ A(z)\mathbf{Q}(z)\left(I + \frac{gz^{-D_2}H(z)\mathbf{Q}_0(z)}{(1 - \mathbf{Q}_0(z)gz^{-D_2}\mathbf{H}(z))}\right) \times\mathbf{H}_n(z)\mathbf{N}(z).\end{aligned}$$

Assuming that the noise sources are uncorrelated, this results in

$$\begin{aligned}&\varepsilon\{\|E_w(z)\|^2\}\\&=\varepsilon\left\{\left\|\frac{A(z)}{\tilde{A}(z)}\mathbf{Q}(z)\left(I+\frac{gz^{-D_2}H(z)\mathbf{Q}_0(z)}{(1-\mathbf{Q}_0(z)gz^{-D_2}\mathbf{H}(z))}\right)\mathbf{H}_v(z)W(z)\right\|^2\right\}\\&\quad+\varepsilon\left\{\left\|A(z)\mathbf{Q}(z)\left(I+\frac{gz^{-D_2}H(z)\mathbf{Q}_0(z)}{(1-\mathbf{Q}_0(z)gz^{-D_2}\mathbf{H}(z))}\right)\mathbf{H}_n(z)\mathbf{N}(z)\right\|^2\right\}.\end{aligned}$$

This expression is lower bounded by $\varepsilon\{\|W(z)\|^2\}=\varepsilon\{\|z^{-D_1}W(z)\|^2\}$ for monic $A(z),\tilde{A}(z)$. Furthermore when D_2 is larger than the order of $A(z),\tilde{A}(z)$, this bound is reached at a unique minimum [10], namely when $A(z)=\tilde{A}(z)$ and when $\mathbf{Q}(z)\mathbf{H}_n(z)\mathbf{N}(z)=0$ and $\mathbf{Q}(z)\mathbf{H}(z)=0$. Using also (3) and (4), this leads to

$$\begin{aligned}E_w(z)&=\frac{A(z)}{\tilde{A}(z)}\mathbf{Q}(z)\mathbf{H}_v(z)W(z)\\&=z^{-D_1}W(z),\end{aligned}$$

and hence

$$\varepsilon\{\|E_w(z)\|^2\}=\varepsilon\{\|z^{-D_1}W(z)\|^2\}.$$

The minimum is reached when $A(z)=\tilde{A}(z)$ and

$$\begin{pmatrix}\mathbf{H}^{\mathrm{T}}(z)\\ \mathbf{N}^{\mathrm{T}}(z)\mathbf{H}_n^{\mathrm{T}}(z)\end{pmatrix}\mathbf{Q}^{\mathrm{T}}(z)=0,$$

$$\begin{pmatrix}\mathbf{H}^{\mathrm{T}}(z)\\ \mathbf{N}^{\mathrm{T}}(z)\mathbf{H}_n^{\mathrm{T}}(z)\end{pmatrix}(z^{-D_1}\mathbf{M}(z)-\mathbf{B}^{\mathrm{T}}(z)\mathbf{F}(z))=0,$$

$$\begin{pmatrix}\mathbf{H}^{\mathrm{T}}(z)\\ \mathbf{N}^{\mathrm{T}}(z)\mathbf{H}_n^{\mathrm{T}}(z)\end{pmatrix}\mathbf{B}^{\mathrm{T}}(z)\mathbf{F}(z)=\begin{pmatrix}\mathbf{H}^{\mathrm{T}}(z)\\ \mathbf{N}^{\mathrm{T}}(z)\mathbf{H}_n^{\mathrm{T}}(z)\end{pmatrix}z^{-D_1}\mathbf{M}(z),$$

where $\mathbf{F}(z)$ contains $M-1$ unknowns. Recall that the number of rows in $\mathbf{H}_n^{\mathrm{T}}(z)$ equals the number of noise sources. Hence a solution exists (for infinitely long filters) if the number of noise sources is less than or equal to $M-2$. Note that for $\mathbf{Q}(z)\mathbf{H}_n(z)\mathbf{N}(z)=0$, recalling that $\mathbf{Q}(z)$ is the vector transfer function from the microphone array to the error (loudspeaker) signal, perfect noise reduction is achieved, and similarly with $\mathbf{Q}(z)\mathbf{H}(z)=0$, perfect feedback cancellation is achieved. Also note that the filter $A(z)$ is needed for this. If for instance $A(z)$ is set to $A(z)=1$ in (7), then in general a solution is obtained with $\mathbf{Q}(z)\mathbf{H}(z)\neq 0$, and so perfect feedback cancellation is not achieved. In conclusion: minimization of (7) has a solution (or multiple solutions if the number of noise sources is smaller than $M-2$), and for these solutions, perfect noise and feedback cancellation is achieved.

4.3. Implementation

Due to the joint estimation of $\mathbf{F}(z)$ and $A(z)$, the minimization problem (7) is nonlinear. It is similar to the minimization in [8,13], and so the same strategy (PEM-AFROW) can be applied. This technique consists of separating the nonlinear minimisation problem in two linear problems. First, $\mathbf{F}(z)$ is assumed known, and for a frame of data $e(k-L+1)\ldots e(k)$, a predictor $A(z)$ is computed. Note that if $\mathbf{F}(z)$ is correct, the estimated $A(z)$ corresponds to the z-transform $\tilde{A}(z)$ of the AR-coefficients $\tilde{\mathbf{a}}$ of $v(k)$. Then, in a second step, for the same time frame, the data $d(k-L+1)\ldots d(k)$ and $u_i(k-L+1)\ldots u_i(k)$, $i=1\ldots M-1$ is filtered with the prediction error filter corresponding to $A(z)$, as shown in Fig. 6, and the filter outputs are used to update the adaptive filters $\mathbf{F}(z)$. As we are dealing with a nonlinear cost function, it is possible that the procedure converges to a local minimum. However, due to the non-stationarity of the speech signal, and the fact that $\mathbf{F}(z)$ may be considered to be stationary over longer time frames, the cost function changes due to changes in one set of parameters $A(z)$, while the other set of parameters $\mathbf{F}(z)$ remains constant. Intuitively it is clear that these variations allow the procedure to escape from the local minima. This is confirmed in our simulations.

The actual PEM-AFROW algorithm of [8,13] as such implements a different approach to feedback cancellation, where the impulse response from the loudspeaker to the microphone is estimated as in an echo-canceller scenario. It also involves a signal whitening, to avoid a signal correlation that otherwise leads to a biased solution. The major advantage of the GSC-structure presented here compared to the direct PEM-AFROW approach in [8,13] is that the order of the adaptive filter in the GSC-structure can be significantly lower than in the PEM-AFROW room impulse response modelling setup. In other words, significantly less degrees of freedom are required to steer a zero in the direction of the loudspeaker than to model the acoustic path from the loudspeaker to the microphones. In addition, if the filter order is taken small enough, namely smaller than the minimal pitch period in the speech input signal, the use of both a short-term *and*

a long-term predictor as in [8,13] is not necessary anymore. The bias on the adaptive filter coefficients due to the pitch period (which is removed by the long-term predictor) does not occur if the stimulus signal for voiced sounds is modelled as an ideal impulse train. In practice, this is a good approximation, and so the long-term predictor can indeed be left out.

5. Simulations

In this section, we show (for a simulated environment) that even with a small adaptive filter length, feedback cancellation can be obtained in a noise robust way, by means of the algorithm described above.

5.1. GSC with one noise reference signal

First we evaluate a scheme with only one noise reference ($F_2(z) = 0$) in the noise-free case. The feedback cancellation performance is measured based on the 'added gain', which is the difference between the loop gain applied for the controlled system ($\mathbf{f}_{1,0} = \mathbf{f}_1$ after convergence) and the maximum stable loop gain for the uncontrolled system ($\mathbf{f}_{1,0} = 0$). The 'maximum added gain' is the value for the added gain for which the controlled system is still marginally stable (larger gains would render it unstable).

As near-end speech signal, a 3 s sentence is used sampled at 8 kHz and pronounced by a male speaker. The near-end speaker is assumed to be sufficiently close to the array so that no reverberation occurs on the near-end signal. Note that the PEM-AFROW algorithm uses order $P = 20$ TVAR-modelling of the near-end component in the microphone signals. This order is a typical value for clean speech modelling at 8 kHz (as is very well known from speech coding literature), but a strong reverberation in this component would lead to poor modelling (and poor whitening). Real time and real environment experiments in [8,13] have shown that for typical situations, this is no concern in practice, though.

A 500 taps simulated RIR (mirror image method) of a $5 \times 4 \times 3$ m room is used, once with wall reflection coefficients $r = 0.5$, and once with wall reflection coefficients $r = 0.1$. These settings correspond to the RIRs for the feedback path from the loudspeaker to the first of the microphones shown in Fig. 7. With the Eyring [14] approximation formula for reverberation time this corresponds to $T_{60} = 52$ ms and $T_{60} = 20$ ms, respectively. Note that these reverberation times are small for typical office rooms, but they are quite realistic in e.g. a car environment (example application: passenger communication system). Three microphones are spaced 5 cm apart on a linear array centered at room coordinates [1,2,1] (in meters). The loudspeaker is at [4.5, 3.5, 1]. An RLS-based adaptive filter with 40 taps is used. Values for the steering vector and the blocking matrix are $M = (\frac{1}{3}\ \ \frac{1}{3}\ \ \frac{1}{3})$, and $B = (1\ \ 1\ \ -2)$. The simulation shows that in this case a

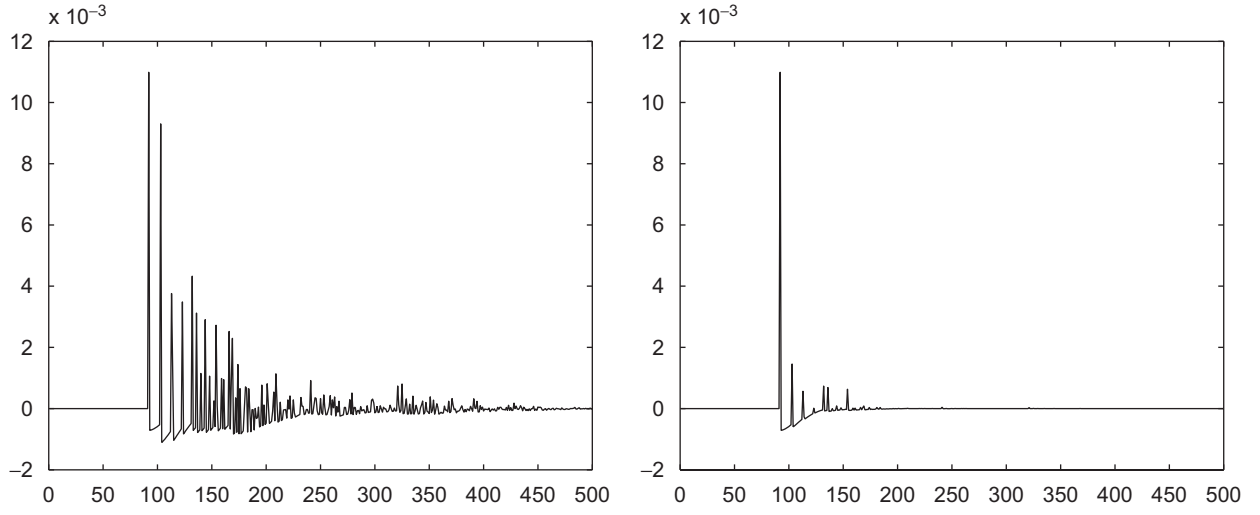

Fig. 7. Room impulse responses for the simulated reverberant environment (left) and the non-reverberant environment (right). Horizontal axes are filter taps.

maximum added gain of 7.2 dB can be reached when $r = 0.5$. Note that this value is lower than typical maximum added gains with the PEM-AFROW approach in [8,13]. On the other hand, if the loudspeaker to microphone array angle is fixed, a more robust performance can be expected because the parameters mostly depend on this angle, and not on small changes in the RIR (moving objects or persons).

If a white noise point source is added at position [1.29, 1.5, 1], with an SNR of the near-end signal versus the noise measured at the position of the middle microphone of 25 dB, the maximum added gain for $r = 0.5$ decreases to 6.4 dB. The GSC will now attempt to also perform noise reduction. We measure the noise reduction performance by letting the adaptive filter $\mathbf{f}_1$ in the GSC-system converge. Adaptation is then switched off, and the feedback loop is removed. Now only the noise input signal is applied (not the speech signal). The output energy σ^2_{proc} is measured. Then the filter coefficients in $\mathbf{f}_1$ are set to zero, and the simulation is repeated. This results in a reference output energy σ^2_{noproc}. The noise reduction measure is then $10\log\sigma^2_{\text{noproc}}/\sigma^2_{\text{proc}}$, where a negative value corresponds to noise amplification. This measure shows the noise reduction performance due to the adaptive filter, it does not contain the extra noise reduction provided by the steering matrix. Note that when a feedback loop is present, the noise is also reproduced in the feedback loudspeaker, and hence in a practical situation, the noise reduction of the scheme will be even larger than what is measured here.

The following table shows the results in different scenarios.

SNR	Added gain	Noise reduction	Added gain	Noise reduction
	High reverberation ($r = 0.5$)		Low reverberation ($r = 0.1$)	
∞	7.2 dB (max)	–	14 dB (max)	–
25 dB	6.4 dB (max)	−6 dB		
10 dB	3.5 dB (max)	−0.9 dB	14 dB (max)	−10 dB
			6 dB	−2 dB
			3 dB	0.5 dB
	0 dB	3.1 dB	0 dB	9.3 dB
5 dB	1.5 dB (max)	2.62 dB	14 dB (max)	−10 dB
0 dB	0 dB (max)	7.37 dB	14 dB (max)	−10 dB

Negative noise reduction values mean noise amplification. In the low-reverberation case, the maximum added gain (14 dB) is constant as a function of the SNR, and the system will make a trade-off between feedback suppression and noise reduction: if feedback is the most important part in the minimization criterion (high gains), then noise may be amplified (e.g. SNR = 10 dB, added gain = 14 dB), but the system will remain stable. On the other hand, for lower gains (e.g. SNR = 10 dB, added gain is 0/3/6 dB) more noise reduction is performed (up to 9.3 dB for 0 dB added gain).

In the case of high reverberation ($r = 0.5$), the maximum stable added gain decreases when the SNR decreases.

For the high-reverberation case, simulations were also performed with a 200 taps adaptive RLS filter. The following table shows the results. Again, if the gain is set to the maximum added gain, noise amplification occurs, but the system remains stable, and if a lower gain is used, noise reduction occurs.

SNR	Added gain	Noise reduction
	High reverberation ($r = 0.5$)	
10 dB	6 dB (max)	−4 dB
	3.5 dB	2.9 dB

5.2. GSC with two noise reference signals

Although the one noise reference signal scheme is robust to noise (and can also perform noise reduction), a better combined noise- and feedback cancellation performance can be obtained when two reference signals are available: one zero can then be steered towards the feedback loudspeaker, and one zero towards the noise source. The blocking matrix is now chosen as $B = \begin{pmatrix} 1 & 1 & -2 \\ 1 & -2 & 1 \end{pmatrix}$.

For more reverberant environments ($r = 0.5$), the performance of the system with the short adaptive filter (40 taps) is worse than in the single reference case. But for $r = 0.1$, with the same setup as above,

the following results are obtained:

SNR	Added gain	Noise reduction (dB)
10	17 dB (max)	−0.4
	12 dB	3
	6 dB	9
5	17 dB (max)	−1.3
	12 dB	9.8
	6 dB	12.8
0	17 dB (max)	−2.3
	12 dB	12.5
	6 dB	15.3

Whereas in the one noise reference signal case, noise amplification occurs and the maximum added gain is 14 dB, here the maximum added gain is 17 dB (with only minor noise amplification), and for lower gains, significant noise reduction is performed.

For the high-reverberation case, simulations were also performed for 10 dB SNR with a 200 taps adaptive RLS filter per reference channel. The following table shows the results:

SNR	Added gain	Noise reduction
	High reverberation ($r = 0.5$)	
10 dB	15.5 dB (max)	−3 dB
	6 dB	2.3 dB
	3.5 dB	4.6 dB

6. Conclusions

We have derived a GSC-based scheme for feedback cancellation which is robust to additive noise, and which can be extended to a combined noise- and feedback cancellation scheme. The practical implementation can be done by means of the PEM-AFROW algorithm. The complexity of this setup is much smaller compared to feedback-cancellation based on PEM-AFROW only. In applications with long acoustic paths, like PA-applications, this complexity reduction is very significant while performance can still be kept at the same level.

Acknowledgments

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') and P5/11 ('Mobile multimedia communication systems and networks'), the Concerted Research Action GOA-AMBioRICS, IWT project 040803: 'SMS4PA-II: Sound Management System for Public Address Systems'. The scientific responsibility is assumed by its authors.

References

[1] R. Porayath, D.J. Mapes-Riordan, Acoustic feedback elimination using adaptive notch filter algorithm. U.S. Patent 5,999,631, December 1999.
[2] M.G. Siqueira, A. Alwan, Steady-state analysis of continuous adaptation in acoustic feedback reduction systems for hearing-aids, IEEE Trans. Speech Audio Process. 8 (July 2000) 443–453.
[3] J. Hellgren, U. Forssell, Bias of feedback cancellation algorithms in hearing aids based on direct closed loop identification, IEEE Trans. Speech Audio Process. 9 (2001) 906–913.
[4] S. Kamerling, K. Janse, F. van der Meulen, A new way of acoustic feedback suppression, in: AES, AES, 1998.
[5] M.P. Lewis, T.J. Tucker, D.M. Oster, Method and apparatus for adaptive audio resonant frequency filtering. U.S. Patent 5,245,665, September 1993.
[6] C.P. Janse, H.J.W. Belt, Sound reinforcement system having an echo suppressor and loudspeaker beamformer. Patent WIPO WO 03/010996 A3, 24 June 2002.
[7] G.W. Elko, M.M. Goodwin, Beam dithering: acoustic feedback control using a modulated-directivity loudspeaker array, in: Proceedings of the IEEE International Conference Acoustics, Speech, Signal Processing, vol. 1, ICASSP-93, 1993, pp. 173–176.
[8] G. Rombouts, T. van Waterschoot, K. Struyve, M. Moonen, Acoustic feedback suppression for long acoustic paths using a nonstationary source model, IEEE Trans. Signal Process. 54 (September 2006) 3426–3434.
[9] A. Spriet, G. Rombouts, M. Moonen, J. Wouters, Adaptive feedback cancellation in hearing aids. KULEUVEN/ESAT-SCD Internal Report 05-61.
[10] A. Spriet, I. Proudler, M. Moonen, J. Wouters, Adaptive feedback cancellation in hearing aids with linear prediction of the desired signal, IEEE Trans. Signal Process. 53 (October 2005) 3749–3763.
[11] O. Cappe, Elimination of the musical noise phenomenon with the ephraim and malah noise suppressor, IEEE Trans. Speech and Audio Process. 2 (1994) 345–349.
[12] L.J. Griffiths, C.W. Jim, An alternative approach to linearly constrained adaptive beamforming, IEEE Trans. Antennas Propagation 30 (1982) 27–34.
[13] G. Rombouts, T. van Waterschoot, K. Struyve, M. Moonen, Acoustic feedback suppression for long acoustic paths using a nonstationary source model, in: Proceedings of Eusipco 2005, Antalya, Turkey, September 2005.
[14] C. Eyring, Reverberation time in 'dead' rooms, J. Acoust. Soc. Amer. 1 (1930) 217–241.