We propose a combination of the well known generalized sidelobe canceller based prewhitening and effectively achieves an unbiased speech signal). The PEM–AFROW algorithm of [8] applies signal (indeed meant to be a processed (amplified) version of the speech signal (“near–end signal”) and the (“far–end”) loudspeaker additional signal modelling is required to avoid a biased room impulse response, and insert a so–called controller \( \hat{c} \) or Griffiths–Jim beamformer, and the so–called approaches are based on acoustic echo cancellation procedures, where acoustic feedback occurs, and introduce a notch filter for this frequency into the signal path. More recent approaches [8, 9, 10] are proactive and do not introduce signal distortion, as they are based on an adaptive filter that models the loudspeaker–room–microphone impulse response, and insert a so–called controller \( \hat{f}_0 \), cfr. Figure 1. If the loop gain exceeds unity for a certain frequencies \( \omega_q \) where the loop phase is \( 2n_j \pi \) radians (with \( n_j \) integer), system instability becomes audible as a loud ‘howling’ sound.

Most of the acoustic feedback cancellation techniques that have been derived up till now are single channel techniques [1, 2, 3, 4, 5] (although some multi–channel examples exist [6, 7]). Traditional approaches are mostly ‘reactive’, because they allow the system to become unstable first, in order to then identify the frequency where acoustic feedback occurs, and introduce a notch filter for this frequency into the signal path. More recent approaches [8, 9, 10] are proactive and do not introduce signal distortion, as they are based on an adaptive filter that models the loudspeaker–room–microphone impulse response, and insert a so–called controller \( \hat{f}_0 \), cfr. Figure 1. In its usual setting, the GSC provides noise cancellation, by minimizing the noise energy in the GSC output, while preserving the desired speech signal energy. In the present feedback cancellation setting, the aim of the GSC is to cancel the ‘noise’ signal produced by the loudspeaker (plus other noises perhaps), while again preserving the desired speech signal.

However, a major impediment is that the noise signal (loudspeaker signal) is now correlated with the desired signal, and so cannot be cancelled by the standard GSC. An additional signal modelling and whitening step will be included, which will be adopted from the PEM–AFROW approach.

In section 2, a Griffiths–Jim (or GSC) based multichannel noise reduction scheme is reviewed. In section 3.1, we propose the new algorithm that combines the GSC with a signal whitening procedure. In section 3.2, we show how the optimisation problem introduced in section 3.1 can be solved with the PEM–AFROW algorithm. In section 4, a number of simulations are shown for different scenario’s, which prove the effectiveness of the new scheme, both in noiseless and noisy environments. Conclusions are given in section 5.

2. GSC BASED NOISE REDUCTION

A traditional approach to multichannel noise reduction is the so–called Griffiths–Jim beamformer [11], or GSC, shown in Figure 2. A fixed beamformer \( \mathbf{w} \) produces a speech reference signal \( d(k) \) by “zooming in” on the speech signal source \( v(k) \). The input vector of the \( M \)–microphone array can be written as

\[
y(k) = v(k) + x(k)
\]

(1)

\[
y(k) = \begin{bmatrix} y^{(1)}(k) \\
y^{(2)}(k) \\
\vdots \\
y^{(M)}(k) \end{bmatrix}
\]

(2)

Figure 1: An electro–acoustic loop with a feedback controller \( \hat{f}_0 \).

Figure 2: Griffiths–Jim beamformer. The coefficients in \( M \) and \( B \) are examples : \( u \) does not contain a component correlated with the broadside speech signal \( v \), and hence forms a noise reference. Correlated noise is removed from \( d \) by means of an adaptive filter.
where \(v(k)\) and \(x(k)\) have a similar structure as \(y(k)\), and where \(x(k)\) is the component in the microphone signal which stems from the noise source, while \(v(k)\) stems from the desired speech source. The fixed beamformer \(m\) already suppresses some noise coming from directions different from the speech signal direction, as well as diffuse noise. This means \(d(k)\) is a signal which contains both a speech– and a noise component. A blocking matrix \(B\) is chosen such that it suppresses the signal and hence creates a noise reference \(u(k)\). In this paper, we only use one single noise reference channel, hence the blocking matrix reduces to a blocking vector. An adaptive filter then removes the noise component in \(d(k)\) which is correlated with \(u(k)\). In low reverberation situations, this amounts to implementing a beam–pattern on the sensor array which has a zero in the direction of the noise source \(u(k)\).

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

\[
d(k) = m^T v(k-D_1) + m^T x(k-D_1),
\]

and the signal

\[
u(k) = Bx(k).
\]

The adaptive filter \(f(k)\) with length \(N_f\) now should converge to

\[
\min_f \{\|d(k)-f^T u(k)\|^2\},
\]

with

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

Hence,

\[
f(k) = E\{u(k)u^T(k)\}^{-1}E\{m^T x(k)u(k)\}.
\]

The output in Figure 2 is

\[
e(k) = m^T v(k-D_1) + m^T x(k-D_1) - \]

\[
E\{u(k)u^T(k)\}^{-1}E\{m^T x(k)u(k)\}u(k).
\]

If no signal leakage occurs through \(B\), that is the signal \(u(k)\) does not contain a component correlated with \(v(k)\) (a signal component), then \(e(k)\) contains an undistorted version of \(m^T v(k)\). This means the GSC approach works better in less reverberant environments.

3. FEEDBACK CANCELLATION

We propose to use a GSC–like structure for acoustic feedback suppression. In an electro–acoustic loop, a microphone signal is amplified, and emitted from a loudspeaker in the same room. 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 though, that in a traditional GSC scheme, the loudspeaker signal would lead to signal leakage through blocking matrix \(B\), and hence result in signal distortion.

3.1 Procedure

Still referring to Figure 2, we first assume that \(v(k)\) is a white noise signal instead of a speech signal (further on we will drop this assumption), that 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 (cf. Figure 2), then signal \(u(k)\) is not correlated with \(mv(k-D_1)\), which is one of the components in \(d(k)\). On the other hand, the component \(mx(k-D_1)\) is correlated with \(u(k)\) and will be removed by the adaptive filter. Hence \(e(k) = mv(k-D_1)\) and \(e(k)\) does not contain a component stemming from the loudspeaker, thus feedback cancellation is effectively performed. 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 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 during use. On the other hand, it may be of interest to be able to adapt to changing acoustic environments. Hence we derive an algorithm which uses the (speech) signal to perform a continuous training of the adaptive filter.

For a speech signal \(v(k)\), we propose the use of prewhitening filters, as shown in Figure 3 and adopt a procedure similar to the so–called PEM-AFROW algorithm of [8]. Here, the updating of the adaptive filter \(F\) is performed with prewhitened versions of the signals \(d(k)\) and \(u(k)\). The prewhitening filter is a linear prediction error filter \(A\), and the coefficients of \(A\) are estimated together with \(f\). The resulting adaptive filter coefficients \(f\) are copied at regular time instants to the so–called ‘controller’ \(f_0\) (dashed arrow).

For the analysis, we assume a stationary AR–input signal and a noise–free environment. In the simulations in section 4, we will show results for real speech signals, both in a noise–free and in a noisy environment. Define

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

where \(H^{(i)}(z)\) is the transfer function from the loudspeaker to the \(i\)th microphone. Similarly \(H_1(z)\) is the transfer vector from the signal source to the microphone array. Vector \(M(z)\) is an \(M\)-element vector, and \(B(z)\) is an \(1 \times M\) blocking matrix . We assume that the response of \(M(z)\) in the direction of the signal of interest is unity,

\[
M^T(z)H_1(z) = 1
\]

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

\[
B(z)H_1(z) = 0.
\]
The vector
\[ Y(z) = \begin{pmatrix} y^{(1)}(z) \\ y^{(2)}(z) \\ \vdots \\ y^{(M)}(z) \end{pmatrix} \] (14)
is the z-transform of the microphone array input signal. We have
\[
E(z) = D(z) - U(z)F_0(z), \\
Y(z) = H(z)E(z)z^{-D_1} + H(z)V(z), \\
D(z) = M^T(z)Y(z)z^{-D_1}, \\
U(z) = B(z)Y(z).
\] (15-18)
Define
\[ F(z) = D(z) - U(z)F(z) \] (19)
The whitened residual signal is
\[
E_w(z) = A(z)(D(z) - U(z)F(z)) = A(z)E(z).
\] (20)
The minimization problem that will be solved is now
\[
\min_{A,F} \varepsilon \left( \| E_w(z) \|^2 \right). \tag{21}
\]
We have
\[
E(z) = \frac{(z^{-D_1}M^T(z) - B(z)F(z))H(z)V(z)}{1 - z^{-D_1}(M^T(z)H(z)z^{-D_1} - B(z)H(z)F(z))}.
\] (22)
Defining the 1 × M vector
\[ Q(z) = z^{-D_1}M^T(z) - F(z)B(z), \]
we can write
\[
E_w(z) = A(z)\frac{Q(z)H(z)V(z)}{1 - z^{-D_1}Q(z)H(z)}.
\] (23)
With a time invariant model \( V(z) = \frac{1}{A(z)}W(z) \) and \( W(z) \) a white Gaussian noise process and \( \hat{A}(z) \) a \( P \)th order, monic polynomial (cfr. A(z)), this becomes
\[
E_w(z) = A(z)\frac{Q(z)H(z)V(z)}{A(z) - z^{-D_1}A(z)Q(z)H(z)}W(z).
\] (24)
If the minimum of \( \varepsilon \left( \| E_w(z) \|^2 \right) \) corresponds to \( E_w(z) = z^{-D_1}W(z) \), i.e., when (25) is whitened by \( A(z) \) and \( F(z) \), and if in this case \( A(z) = \hat{A}(z) \), the desired solution is found, since then (under (12) and (13)) \( E(z) = z^{-D_1}A(z)W(z) \). After copying \( F \) to \( f_0 \), the loudspeaker signal \( z^{-D_2}E(z) = z^{-D_2}\hat{D}(z) = z^{-D_2}z^{-D_1}V(z) \) (which is the delayed near end speech signal).

We note that for monic \( A(z) \) and \( \hat{A}(z) \), \( \varepsilon \left( \| W(z) \|^2 \right) \) is a lower bound for \( \varepsilon \left( \| E_w(z) \|^2 \right) \), which is reached when
\[
A(z)\frac{Q(z)H(z)}{A(z) - z^{-D_1}A(z)Q(z)H(z)} = z^{-D_1}.
\] (26)
From this expression, it can be seen that if \( D_2 > P \) (with \( P \) the order of the speech model AR process, and hence the largest exponent in \( \hat{A}(z) \)), a unique solution is found for \( A(z) \), namely \( A(z) = \hat{A}(z) \).

Using (13), we can now write (26) as
\[
\frac{M^T(z)H(z)z^{-D_1}}{1 - z^{-D_2}(M^T(z)z^{-D_1} - F(z)B(z))H(z)} = z^{-D_1}.
\] (27)
This results in
\[
F(z) = \frac{1 - z^{-D_1}D_2M^T(z)H(z) - M^T(z)H(z)}{z^{-D_1}B(z)H(z)}.
\] (28)
With the assumption of unit response in the direction of the near end signal, the last term in the numerator becomes 1 and we obtain the unique solution
\[
F(z) = \frac{z^{-D_1}D_2M^T(z)H(z)}{B(z)H(z)}.
\] (29)
In practical (PA) scenarios, usually several loudspeakers, all producing the same signal, are used. The proposed scheme also performs well in this case because the resulting signal at the microphone array can also be written as \( H(z)V(z) \) where \( H(z) \) is now the sum of all transfer functions from all the loudspeakers to the microphone array. This may seem counterintuitive, since it is well known that a traditional GSC with one single reference can only steer a zero towards a single interferer, but in our case all interference signals are correlated (i.e. the same).

3.2 Implementation

Due to the joint estimation of \( F \) and \( A \), the minimization problem (21) is nonlinear. It is similar to the minimization in [8, 12], and the same strategy (PEM–AFROW) can be applied. This technique consists of separating the nonlinear minimisation problem in two linear problems. First, \( F \) is assumed known, and for a frame of data \( e(k-L+1) \ldots e(k) \), a predictor \( A \) is computed. Note that in the ideal case (either no noise or no reverberation), and if \( F \) is correct, the estimated \( A \) corresponds to the AR coefficients \( \hat{A} \) of \( \epsilon(k) \). Then, in a second step, for the same time frame, the data \( d(k-L+1) \ldots d(k) \) and \( u_{12}(k-L+1) \ldots u_{12}(k) \) is filtered with the prediction error filter corresponding to \( A \), as shown in Figure 3, and the residuals are used to update the adaptive filter \( F \). Because of the separation of the nonlinear cost function, it is possible that convergence to a local minimum occurs. However, because of the nonstationarity of the speech signal (and the relative stationarity of \( F \), which determines the relatively constant direction of all the loudspeakers to the microphone array), the cost function constantly changes, and so it can be seen that the convergence process may also leave these local minima. This is confirmed by our simulations.

The PEM–AFROW algorithm of [8, 12] implements a different approach, where the impulse response from the loudspeaker to the microphone is estimated as in an echo–cancellation 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 compared to the direct PEM–AFROW approach in [8, 12] is that the order of the adaptive filter can be significantly lower (e.g. 40 taps) than in the room impulse modelling setup (e.g. 4000 taps for a 16 kHz sampling rate). Significantly less degrees of freedom are required to steer a zero in the direction of the loudspeaker than for modelling the exact room impulse response. 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, 12] is not required. For such a small filter order, the estimation bias due to the pitch period 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.

The computational complexity of the proposed algorithm is roughly equal to that of a PEM–AFROW implementation with a filter length of e.g. 40 taps (instead of e.g. 4000 taps).

4. SIMULATIONS

First we evaluate the scheme in the noise–free case. The feedback suppression performance is measured based on the ‘added gain’, which is the difference in maximum stable loop gain for the uncontrolled system (\( f_0 = 0 \)), and the maximum stable loop gain for the...
controlled system \( f_0 \) after convergence. As near end speech signal, a 3 second sentence sampled at 8kHz and pronounced by a male speaker is used. The near end signal is assumed to be pronounced sufficiently close to the array so that no reverberation occurs. A 500 taps simulated room impulse response [13] of a 5x4x3 meter room is used, once with wall reflection coefficients \( r = 0.5 \) and once with wall reflection coefficients \( r = 0.1 \). Three microphones are spaced 5 cm on a linear array centered at \([1,2,1]\). The loudspeaker is at \([4.5, 3.5, 1]\). An RLS adaptive filter with 40 taps is used. Values for the steering vector and the blocking matrix are \( \mathbf{m} = \left( \frac{1}{3}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right)^T \), and \( B = \left( \begin{array}{ccc} 1 & 1 & -2 \end{array} \right) \). The simulation shows that in this case a maximum added gain of 7.2 dB can be reached (maximum means that the system is still marginally stable at that point). Note that this value is lower than typical added stable gains with the PEM–AFROW approach in [8, 12]. 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 room impulse response (moving objects or persons).

If a white noise point source is added at position \([1.29, 1.5, 1]\), with a SNR of the near end signal versus the noise measured at the position of the middle microphone of 25 dB, the added gain decreases to 6.4 dB. The GSC will now also attempt to perform noise reduction. We measure the noise reduction performance by letting the adaptive filter \( f_0 \) 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) and the output energy \( \sigma_{\text{noproc}}^2 \) is measured. Then the filter coefficients in \( f \) are set to zero, and the simulation is repeated. This results in a reference output energy \( \sigma_{\text{noproc}}^2 \). The noise reduction measure is then

\[
10 \log_{10} \frac{\sigma_{\text{noproc}}^2}{\sigma_{\text{proc}}^2}, \quad \text{where a negative value corresponds to noise amplification.}
\]

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.

<table>
<thead>
<tr>
<th>SNR</th>
<th>AG</th>
<th>NR</th>
<th>AG</th>
<th>NR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.5</td>
<td></td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>0 dB</td>
<td>7.2 dB *</td>
<td>-</td>
<td>14 dB *</td>
<td>-</td>
</tr>
<tr>
<td>25 dB</td>
<td>6.4 dB *</td>
<td>-6 dB</td>
<td>-10 dB *</td>
<td></td>
</tr>
<tr>
<td>10 dB</td>
<td>3.5 dB *</td>
<td>1.89 dB</td>
<td>14 dB *</td>
<td>-10 dB</td>
</tr>
<tr>
<td>14 dB</td>
<td>6 dB</td>
<td>-</td>
<td>2 dB</td>
<td>-</td>
</tr>
<tr>
<td>10 dB</td>
<td>0 dB</td>
<td>3.1 dB</td>
<td>0 dB</td>
<td>9.3 dB</td>
</tr>
<tr>
<td>5 dB</td>
<td>1.5 dB *</td>
<td>2.62 dB</td>
<td>14 dB *</td>
<td>-10 dB</td>
</tr>
<tr>
<td>0 dB</td>
<td>0 dB *</td>
<td>7.37 dB</td>
<td>14 dB *</td>
<td>-10 dB</td>
</tr>
</tbody>
</table>

The added gains marked by an asterisk are the maximum achievable added gains. In the low reverberation case, the maximum added gain 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 minimisation criterion (high gains), then noise may even be amplified (e.g. SNR=10 dB, added gain=14 dB), but the system will remain stable. In the case with high reverberation, the maximum added gain decreases when the SNR decreases. It is found that the PEM–AFROW based GSC system performs feedback cancellation in a noise–robust fashion.

5. CONCLUSIONS

We have derived a GSC–based scheme for feedback cancellation which is robust to additional noise. In low reverberant environments, the complexity of this setup is much smaller compared to feedback–cancellation based on PEM–AFROW only, while performance is only marginally degraded.

6. ACKNOWLEDGEMENTS

This research work was carried out at the ESAT laboratory of the Katholieke Universiteit Leuven, in cooperation with Televic NV, Izegem, Belgium, in the frame of IWT project 040803: ‘SM54PA-Ii: Sound Management System for Public Address Systems’, 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 the Concerted Research Action GOA-AMBioRICS. The scientific responsibility is assumed by its authors.

REFERENCES