1 Introduction

Recently, a multi-channel frequency-domain adaptive algorithm was presented in [1]. This algorithm has been shown to work very well in the two-channel acoustic echo cancellation application, [1, 2]. It has a fairly low computational complexity compared to the fast recursive least squares algorithm (FRLS), [2]. Furthermore, it is an inherently stable algorithm well suited for a fixed point implementation. Our objective in this memorandum is to provide a complete solution, based on the multi-channel frequency-domain adaptive algorithm, which handles both echo cancellation and double-talk.

Double-talk, i.e. when the speech of the two talkers arrives simultaneously at the canceler, is best handled by a double-talk detector (DTD) combined with built-in robustness in the adaptive algorithm, [3]. A frequency-domain equivalent to the fast normalized cross-correlation DTD [3, 4], as well as a robust version of the multi-channel frequency-domain algorithm are proposed in this memorandum.

The paper is organized as follows. Section 2 introduces the multi-channel echo cancellation problem, definitions, and notations needed in the description of the adaptive algorithm and double-talk detector. Section 3 reviews the test statistic of the normalized cross-correlation DTD and presents a frequency-domain version based on a pseudo coherence estimate. Section 4 presents and comments on a robust version of the frequency-domain algorithm. A method for detecting echo path changes is discussed in Section 5. Evaluation of the proposed frequency-domain system is made in Section 6 where we study double-talk handling and tracking of echo path changes. Section 7 summarizes the results and presents the conclusions.

2 The Multi-Channel Frequency-Domain Algorithm

This section gives the definitions and notations of the multi-channel frequency-domain adaptive algorithm presented in [1]. Figure 1 shows the basic block diagram for a two-channel acoustic echo canceler. Communication is hands-free between a transmission room and a receiving room where we denote the signals picked up by the microphones in transmission room by \( x \), and the return signals picked up in the receiving room by...
Figure 1: Block diagram of a generic two-channel acoustic echo canceler.

The receiving room signal is in general composed of echo, ambient noise, and possibly receiving room speech which in this scenario, is referred to as double-talk. Thus we have the receiving room signal model:

\[ y(n) = y_e(n) + v(n) + w(n), \]  

where \( y_e(n) = \sum_{i=1}^{P} y_i(n) \) is the echo, \( v(n) \) is the receiving room speech\(^1\), and \( w(n) \) is the ambient noise.

For the general \( P \)-channel case, we assume that we have \( P \) incoming transmission room signals, \( x_i(n), i = 1, \ldots, P \). Define the \( P \) input excitation vectors as,

\[ x_i(n) = \begin{bmatrix} x_i(n) & x_i(n-1) & \cdots & x_i(n-L+1) \end{bmatrix}^T, \quad i = 1, 2, \ldots, P, \]

where the superscript \(^T\) denotes transposition.

The error signal at time \( n \) between one arbitrary microphone output \( y(n) \) in the receiving room and its estimate is

\[ e(n) = y(n) - \sum_{i=1}^{P} \hat{y}_i(n) = y(n) - P \sum_{i=1}^{P} x_i^T(n) \hat{h}_i, \]  

where

\[ \hat{h}_i = \begin{bmatrix} \hat{h}_{i,0} & \hat{h}_{i,1} & \cdots & \hat{h}_{i,L-1} \end{bmatrix}^T, \quad i = 1, 2, \ldots, P \]

are the \( P \) modeling filters. In general, we have \( P \) microphones in the receiving room as well, i.e. \( P \) return channels, and thus \( P^2 \) adaptive filters. From now on though, we will just look at one return channel because the derivation is similar for the other \( P-1 \) channels.

In order to present the frequency-domain adaptive algorithm, we need to define the following block signals,

\[ e(m) = [e(mL) \cdots e(mL + L - 1)]^T, \]  
\[ y(m) = [y(mL) \cdots y(mL + L - 1)]^T, \]

\(^1\)Sometimes referred to as near-end speech.
where the block length is chosen to be equal to the length of the adaptive filter \( L \). We can now write the estimated return signal block as,

\[
\hat{y}_i(m) = [x_i(mL) \cdots x_i(mL + L - 1)]^T \hat{h}_i, \tag{5}
\]

where \( \mathbf{X}_i \) is an \((L \times L)\) Toeplitz matrix. The block error can be written as,

\[
e(m) = y(m) - \sum_{i=1}^{P} \mathbf{X}_i^T(m) \hat{h}_i. \tag{6}
\]

We would like to compute the Toeplitz-vector product in (6) with fewer multiplications than a general matrix-vector product. This can be achieved by embedding the Toeplitz matrix in a circulant matrix, [5]. The circulant matrix \( \mathbf{C}_i \), \((2L \times 2L)\), can be formed from \( \mathbf{X}_i \) as

\[
\mathbf{C}_i(m) = \begin{bmatrix} \mathbf{X}'_i(m) & \mathbf{X}_i(m) \\ \mathbf{X}_i(m) & \mathbf{X}'_i(m) \end{bmatrix}^T, \tag{7}
\]

where \( \mathbf{X}'_i(m) \) is also a Toeplitz matrix obtained from \( \mathbf{X}_i(m) \).

We can now rewrite the block error equation, (6), as

\[
\begin{bmatrix} \mathbf{0}_{L \times 1} \\ \mathbf{e}(m) \end{bmatrix} = \begin{bmatrix} \mathbf{0}_{L \times 1} \\ \mathbf{y}(m) \end{bmatrix} - \sum_{i=1}^{P} \mathbf{W} \hat{y}'_i(m), \tag{8}
\]

where

\[
\mathbf{W} = \begin{bmatrix} \mathbf{0}_{L \times L} & \mathbf{0}_{L \times L} \\ \mathbf{0}_{L \times L} & \mathbf{I}_{L \times L} \end{bmatrix}, \tag{9}
\]

\[
\hat{y}'_i(m) = \mathbf{C}_i(m) \begin{bmatrix} \hat{h}_i \\ \mathbf{0}_{L \times 1} \end{bmatrix}. \tag{10}
\]

The fundamental property that makes it possible to reduce the computational complexity of the frequency-domain algorithms is that \( \mathbf{C} \) can be diagonalized by the Fourier matrix \( \mathbf{F} \), i.e.

\[
\mathbf{C}_i = \mathbf{F}^{-1} \mathbf{D}_i \mathbf{F}, \tag{11}
\]

where \( \mathbf{D}_i \) is a diagonal matrix whose elements are the discrete Fourier transform of the first column of \( \mathbf{C}_i \).

Transformation of the above derived equations into the frequency-domain gives

\[
e(m) = \mathbf{F} \begin{bmatrix} \mathbf{0}_{L \times 1} \\ \mathbf{e}(m) \end{bmatrix} = \mathbf{y}(m) - \sum_{i=1}^{P} \mathbf{G} \hat{y}'_i(m), \tag{12}
\]

\[
= \mathbf{y}(m) - \mathbf{G} \sum_{i=1}^{P} \mathbf{D}_i(m) \hat{h}_i = \mathbf{y}(m) - \mathbf{G} \mathbf{D}(m) \hat{h}, \tag{13}
\]

where

\[
\mathbf{y}(m) = \mathbf{F} \begin{bmatrix} \mathbf{0}_{L \times 1} \\ \mathbf{y}(m) \end{bmatrix}, \tag{14}
\]

\[
\hat{y}'_i(m) = \mathbf{F} \hat{y}'_i(m), \tag{15}
\]

\[
\mathbf{G} = \mathbf{FWF}^{-1}, \tag{16}
\]

\[
\hat{h}_i = \mathbf{F} \begin{bmatrix} \hat{h}_i \\ \mathbf{0}_{L \times 1} \end{bmatrix}, \tag{17}
\]

\[
\hat{h} = \begin{bmatrix} \hat{h}_1^T \\ \hat{h}_2^T \\ \cdots \\ \hat{h}_P^T \end{bmatrix}^T, \tag{18}
\]

\[
\mathbf{D}(m) = \begin{bmatrix} \mathbf{D}_1(m) & \mathbf{D}_2(m) & \cdots & \mathbf{D}_P(m) \end{bmatrix}. \tag{19}
\]
With this formalism, the multi-channel frequency-domain adaptive algorithm can be derived by minimizing the quadratic criterion,

$$ J(\hat{h}) = (1 - \lambda_t) \sum_{p=0}^{m} \lambda_t^{m-p} \mathbf{e}^H(p) \mathbf{e}(p) $$  

(20)

with respect to $\hat{h}$. The constant $\lambda_t$ is an exponential forgetting factor and subscript $t$ denotes “foreground,” the reason of which is explained in Section 3.

The gradient of $J$ w.r.t. $\hat{h}$ is given by:

$$ \nabla J \left[ \hat{h} \right] = \frac{\partial}{\partial \hat{h}} \left( J \left[ \hat{h} \right] \right) $$  

(21)

$$ = -(1 - \lambda_t) \sum_{p=0}^{m} \lambda_t^{m-p} \mathbf{D}^H(p) \mathbf{G}\mathbf{y}(p) $$  

(22)

$$ + (1 - \lambda_t) \sum_{p=0}^{m} \lambda_t^{m-p} \mathbf{D}^H(p) \mathbf{D}(p) \hat{h}(m). $$  

(23)

Setting this gradient to zero we obtain the normal equation:

$$ S(m)\hat{h}(m) = \mathbf{s}(m) $$  

(24)

where

$$ S(m) = (1 - \lambda_t) \sum_{p=0}^{m} \lambda_t^{m-p} \mathbf{D}^H(p) \mathbf{D}(p) $$

$$ = \lambda_t S(m-1) + (1 - \lambda_t) \mathbf{D}^H(m) \mathbf{G}\mathbf{D}(m) $$  

(25)

and

$$ \mathbf{s}(m) = (1 - \lambda_t) \sum_{p=0}^{m} \lambda_t^{m-p} \mathbf{D}^H(p) \mathbf{G}\mathbf{y}(p) $$

$$ = \lambda_t \mathbf{s}(m) + (1 - \lambda_t) \mathbf{D}^H(m) \mathbf{G}\mathbf{y}(m). $$  

(26)

It is now easy to derive an iterative algorithm by using (24), (25), and (26), [6]:

$$ \hat{h}(m) = \hat{h}(m-1) + (1 - \lambda_t) \mathbf{s}^{-1}(m) \mathbf{D}^H(m) \mathbf{g}(m). $$  

(27)

To reduce complexity, we can simplify this algorithm by approximating $S(m)$ with a block diagonal matrix $S_a(m)$,

$$ \hat{h}(m) = \hat{h}(m-1) + \mu S_a^{-1}(m) \mathbf{D}^H(m) \mathbf{g}(m), $$  

(28)

where $\mu = \mu'(1 - \lambda_t)$ with $0 < \mu' \leq 2$, is a relaxation parameter and

$$ S_a(m) = \lambda_t S_a(m-1) + (1 - \lambda_t) \mathbf{D}^H(m) \mathbf{D}(m). $$  

(29)

### 2.1 A Regularized Two-Channel Version

The simulations in this memorandum are made with a two-channel algorithm. We therefore present this version in more detail. Furthermore, it is known that adaptive filters may have large misalignment for some frequency bands if there is poor excitation in these bands. The frequency domain algorithm is sensitive to poorly excited input frequencies since poor excitation will decrease the small eigenvalues of the matrix $S_a(m)$.
in (29), thus increase the algorithm’s noise sensitivity. It is therefore important to regularize this algorithm carefully such that the update step is small for sensitive bands. Regularization of (28) is performed according to,

\[ \hat{S}(m) = S(m) + \text{diag}\{\delta_{1,0} \ldots \delta_{1,2L-1} \delta_{2,0} \ldots \delta_{2,2L-1}\}, \tag{30} \]

where \( \delta_{j,i}, j = 1, 2, i = 0, \ldots, 2L - 1 \), are chosen as a fraction of the power of the input signal. This gives us the two-channel algorithm:

\[ \begin{align*}
    \xi(m) &= y(m) - G \left[ D_1(m)\hat{h}_1(m-1) + D_2(m)\hat{h}_2(m-1) \right], \\
    \hat{h}_1(m) &= \hat{h}_1(m-1) + \mu S_1^{-1}(m) \left[ D_1^*(m) - \hat{S}_{1,2}(m)\hat{S}_{2,2}^{-1}(m) D_2^*(m) \right] \xi(m), \\
    \hat{h}_2(m) &= \hat{h}_2(m-1) + \mu S_2^{-1}(m) \left[ D_2^*(m) - \hat{S}_{2,1}(m)\hat{S}_{1,1}^{-1}(m) D_1^*(m) \right] \xi(m),
\end{align*} \tag{31-33} \]

where \( \hat{S}_{j,i}(m) \) are diagonal sub-matrices of the regularized matrix \( \hat{S}(m) \) and

\[ S_j(m) = \hat{S}_{j,j}(m) \left[ I_{2L \times 2L} - \Gamma^H(m)\Gamma(m) \right]. \tag{35} \]

\( \Gamma(m) \) is the diagonal coherence matrix:

\[ \Gamma(m) = \left[ \hat{S}_{1,1}(m)\hat{S}_{2,2}(m) \right]^{-1/2} \hat{S}_{1,2}(m). \tag{36} \]

Finally, it should be mentioned that this algorithm has been presented with no overlap of the data. Overlapping, i.e., updating the filter coefficient more often than every \( L \) samples can easily be done by computing the FFTs on overlapped data. Overlapping is described by a parameter \( \alpha_0 \) which means that the coefficients are updated every \( L/\alpha_0 \) samples. Most often, one chooses \( \alpha_0 = 1, 2, 4, \) or \( 8 \). The advantage of updating more often than once per block is to converge and track echo path changes faster.

3 Double-talk Detection Based on a Normalized Pseudo Coherence Measure

Double-talk detectors, of any kind, essentially operate in the same manner. A decision variable, \( \xi \), is formed from available signals, \( x(n), y(n), e(n) \). This variable is compared to a preset threshold, \( T \), and a decision, whether double-talk is present or not, is taken according to

\[ \begin{align*}
    \xi < T_i &\implies \text{double-talk,} \\
    \xi \geq T_i &\implies \text{no double-talk.}
\end{align*} \]

See [3] for more details on double-talk detectors.

A double-talk detection statistic was proposed in [4, 7] that has the potential of fulfilling the properties of an “optimum” statistic, [3]. This variable is defined in terms of the normalized cross-correlation vector between the excitation vector \( x(n) \) and the return signal \( y(n) \). The multi-channel normalized cross-correlation detection statistic, [3, 4], is defined as,

\[ \xi^{cc} = \sqrt{\hat{r}^T (\sigma_y^2 \mathbf{R})^{-1} \hat{r}} = ||\mathbf{c}_{xy}^{cc}||_2. \tag{37} \]

For this case we define

\[ \begin{align*}
    \mathbf{R} &= E\{x(n)x^T(n)\}, \\
    \mathbf{r} &= E\{x(n)y(n)\}, \\
    \sigma_y^2 &= E\{y^2(n)\},
\end{align*} \tag{38-40} \]

Lucent Technologies – PROPRIETARY
Use pursuant to Company Instructions
where \( \mathbf{x}(n) \) is the stacked excitation vector,

\[
\mathbf{x}(n) = \begin{bmatrix} \mathbf{x}_1^T(n) & \mathbf{x}_2^T(n) & \cdots & \mathbf{x}_p^T(n) \end{bmatrix}^T,
\]

and \( \mathbf{c}_{xy} \) is the normalized cross-correlation vector between \( \mathbf{x} \) and \( \mathbf{y} \),

\[
\mathbf{c}_{xy}^{cc} = \left( \sigma_y^2 \mathbf{R} \right)^{-1/2} \mathbf{r}.
\]

Moreover, we can rewrite \( \xi^{cc} \) as:

\[
\xi^{cc} = \frac{\sqrt{\mathbf{h}^H \mathbf{R}^L \mathbf{h}}}{\sqrt{\mathbf{h}^H \mathbf{R} \mathbf{h}} + \sigma_v^2}
\]

where \( \sigma_v^2 \) is the variance of the receiving room speaker. It is easily deduced from this form that for \( v(n) = 0 \), \( \xi^{cc} = 1 \) and for \( v(n) \neq 0 \), \( \xi^{cc} < 1 \). Note also that \( \xi^{cc} \) is independent of the echo path when \( v(n) = 0 \). Hence, our threshold, \( T \), should be chosen in the range, \( 0 < T < 1 \). Typically, \( 0.85 \leq T \leq 0.99 \). In [3], it was shown that the normalized cross-correlation detector (NCC) in fact is estimating the echo path as is done in the echo canceler. The concept of an NCC DTD and an adaptive echo canceler can therefore be regarded as a background and a foreground filter respectively, which is similar to the two-path principle for double-talk detection [8].

Evaluation of the NCC DTD shows it has very high performance, i.e. high probability of detection while maintaining a low probability of false alarm, even when echo path changes occur. The objective of this section is therefore to derive and present a corresponding frequency-domain version of the NCC DTD which can be used together with the frequency-domain adaptive algorithm presented in the previous section.

Calculation of (37) in the frequency-domain can be made from what we call the normalized pseudo coherence (NPC) vector,

\[
\mathbf{c}_{xy}^{pc} = (2L^2 \sigma_y^2 \mathbf{S})^{-1/2} \mathbf{s},
\]

where

\[
\mathbf{S} = E\{ \mathbf{D}^H(m) \mathbf{G} \mathbf{D}(m) \},
\]

\[
\mathbf{s} = E\{ \mathbf{D}^H(m) \mathbf{y}(m) \}.
\]

Looking at (44), we see that each cross-spectrum bin of (46) is normalized by the corresponding spectrum in the input signal, \( \mathbf{x} \). What differentiates (44) from being the true coherence is that it is not normalized by the corresponding spectrum of the output signal (\( \mathbf{y} \)) but by the whole power of the output signal, \( \sigma_y^2 \), hence we call it pseudo coherence. A detection statistic can be defined in the frequency-domain using (44),

\[
\xi^{pc} = ||\mathbf{c}_{xy}^{pc}||_2.
\]

A practical double-talk detection statistic can now be realized by using estimated quantities in (47) and slightly rewriting the numerator from (44),

\[
\xi^2(m) = \frac{\mathbf{s}^H(m) \mathbf{S}^{-1}(m) \mathbf{s}(m)}{\sum_{\lambda} \xi^{2}(\lambda)} = \frac{\mathbf{s}^H(m) \mathbf{h}_b^*(m)}{\sigma_{\mathbf{S}}^2(m)} = \frac{n^2(m)}{\sigma_{\mathbf{S}}^2(m)}.
\]

where we squared the statistics and dropped the superscript \( ^{pc} \) for simplicity. The echo path estimate, \( \mathbf{h}_b(m) \), called the background filter (subscript \( \lambda_b \)) should not to be confused with the foreground estimate (28). This estimate should be adapted with a smaller forgetting factor, \( \lambda_b \), than that of the foreground filter, \( \lambda_f \). By this choice, we ensure that the DTD detects double-talk fast and alerts the foreground filter before it diverges. The variables of (48) are estimated as

\[
\mathbf{s}(m) = \lambda_b \mathbf{s}(m-1) + (1 - \lambda_b) \mathbf{D}^H(m) \mathbf{y}(m),
\]

\[
\mathbf{h}_b(m) = \lambda_b \mathbf{h}_b(m-1) + (1 - \lambda_b) \mathbf{S}_u^{-1}(m) \mathbf{D}^H(m) \mathbf{y}(m),
\]

\[
\sigma_{\mathbf{S}}^2(m) = \lambda_b \sigma_{\mathbf{S}}^2(m-1) + (1 - \lambda_b) \mathbf{y}^H(m) \mathbf{y}(m).
\]
where
\[ g_r(m) = y(m) - GD(m)\hat{h}_r(m-1). \] (52)

Finally, we can show that (37) and (47) are equivalent by looking at the inner product of the cross-spectrum vector in (46) and the frequency-domain echo path vector. We also know that \( y = GD(m)\hat{h} \) in the noise-less case and \( G = G^HG \).

\[
\begin{align*}
\hat{h}^H s &= \hat{h}^H E(D^H(m)G^HGD(m))\hat{h} \\
&= 2L^2\hat{h}^TR\hat{h} \\
&= 2L^2r^TR^{-1}r,
\end{align*}
\] (53)

where
\[
C(m) = [C_1(m) \ C_2(m) \ldots C_P(m)].
\] (54)

Similar type of calculations show that \( E(\sigma^2_s) = E(y^H(m)y(m)) = 2L^2\sigma^2_r \).

### 4 An Outlier-Robust Multi-Channel Frequency-Domain Algorithm

An outlier-robust version of the algorithm (28) has been derived, i.e. this new algorithm is robust against large disturbances resulting from detection errors of the double-talk detector. Instead of minimizing the quadratic criterion in Section 2 the algorithm is derived from minimization of:

\[
J(\hat{h}) = \sum_{p=mL}^{mL+L-1} \rho \left( \frac{|e(p)|}{s_e(m)} \right),
\] (55)

where \( \rho [\cdot] \) is a convex function and \( s_e(m) \) is a real positive scale factor for block \( m \). The non-linear function \( \rho [\cdot] \) can be chosen in many ways, however, as long as it is chosen to have a bounded derivative, the resulting algorithm inherits robust properties. [9]. Robust methods have been proposed and developed for a variety of time-domain algorithms in [3, 10, 11, 12]. In this memorandum though, we look at the frequency-domain case. A variant of this frequency-domain approach was proposed in [13]. A good choice of \( \rho [\cdot] \) is, [9],

\[
\rho (|z|) = \begin{cases} 
\frac{|z|^2}{2}, & |z| \leq k_0 \\
kp_0 - \frac{k_0^2}{2}, & |z| \geq k_0
\end{cases}
\] (56)

where \( k_0 \) is a constant controlling robustness of the algorithm. Appendices A and B show the details in the derivation of the robust frequency-domain algorithm:

\[
\hat{h}(m) = \hat{h}(m-1) + \frac{\mu_s}{s_e(m)\psi_{min}}S_u^{-1}(m) D^H(m)F \psi [e(m)].
\] (57)

where
\[
\psi [e(m)] = \left[ 0_{1 \times L} \psi \left( \frac{e(mL)}{s_e(m)} \right) \text{sign} [e(mL)] \ldots \psi \left( \frac{e(mL+L-1)}{s_e(m)} \right) \text{sign} [e(mL+L-1)] \right]^T,
\]

\[
\psi (|z|) = \rho' (|z|) = \min\{|z|, k_0\}.
\]

In general, the robust algorithm tolerates more double-talk detection errors that a non-robust algorithm which simplifies the design and makes parameter choices of the DTD less critical. In the simulations section, we show the performance of the non-robust as well as the robust frequency-domain algorithm. The DTD threshold \( T \) of the non-robust algorithm is higher than that used with the robust algorithm. With a high threshold, the DTD has a very low probability of missing double-talk which effectively reduces divergence of the non-robust algorithm during double-talk.

Lucent Technologies – PROPRIETARY
Use pursuant to Company Instructions
5 A Tracking Improvement Measure

Increasing the convergence rate of the algorithm is important when there is an echo path change. The pseudo coherence based DTD in Section 3 and the foreground adaptive filter give us the possibility to detect when an echo path change has occurred. Since the DTD is a background adaptive filter and the echo canceler is a foreground adaptive filter, both estimating the same echo path, we can look at the performance of each filter and conclude when the convergence rate of the foreground filter should increase. Moreover, the background filter is updated regardless of double-talk we can deduce the following: if the background residual error power is smaller than the foreground error power, the probability that an echo path change has occurred is high and we can decide to increase the relaxation parameter $\mu$ of the foreground filter (28) thus increasing its convergence rate.

6 Evaluation and Simulations

In this section, we evaluate the performance of the pseudo coherence based DTD by estimating its receiver operating characteristic (ROC). The performance of the whole system, double-talk detector and echo canceler (both non-robust and robust), is shown when echo path changes occur in the receiving room as well as the transmission room and when the system is exposed to double-talk. All results are directly applicable to the general multi-channel case.

6.1 Receiver Operating Characteristics of the DTD

When evaluating a detector, it is customary to show the probability of detection, $P_d$, or probability of miss, $P_m$, versus the probability of false alarm, $P_f$. This has been previously suggested in [7],[14] and was used to evaluate the fast NCC DTD in [3]. One problem that has to be kept in mind though, is that speech has a time-varying power level. This may give less consistent results when estimating $P_d$ and $P_f$ than what would be the case for a stationary signal. Nevertheless, for comparative studies of the DTDs, the estimated ROC is still a useful performance metric. In this section, we evaluate the ROC for the single-channel version of the DTD because of lengthy simulations. The single-channel version has comparable performance to the two-channel version.

Estimation of $P_f$, $P_d$ using a set of speech data obtained from a database, [15], was made according to the procedure presented in [7]. The files used are: sentences (040), (046), and (062) of talker M25, and sentences (012) and (075) of talkers F14, F45, M08, and M51. Furthermore, all sentences are also normalized to have the same average power level.

**Echo path.** The measured acoustic path, “sgi2cl.mat” channel-b, is used as echo path. This response is estimated between a standard SGI cardioid microphone positioned on top of the workstation and the left loudspeaker. The original impulse response has a length of 4096 coefficients, but here it is truncated to a total length of 2048 coefficients, see Fig. 2. It is also normalized so that $\sigma_y^2 = \sigma_e^2$ for the actual speech data in order to maintain a constant average echo to receiving room speech ratio. The original echo path is available in the database: /u/drrm/matlab/desktop.

**Probability of false alarm.** When estimating the probability of false alarm we use all sentences above as transmission room speech. This corresponds to a speech sequence of 21.8 seconds at 16 kHz sampling rate. The signal (echo)-to-ambient noise ratio, $\text{SNR} = 10 \log_{10} (\sigma_e^2 / \sigma_w^2)$, is set to 30 dB which is typical of many office environments.

**Probability of miss.** The three sentences of talker M25, a total of 5 seconds of data, are used as transmission room speech when the probability of a miss is estimated. As receiving room speech, the remaining sentences are used, each about 2 seconds long. In our case we investigate the performance when the average echo-to-near-end ratio, $\text{ENR} = 10 \log_{10} (\sigma_e^2 / \sigma_v^2)$ is 0 dB. This is natural in a

Lucent Technologies – PROPRIETARY
Use pursuant to Company Instructions
desktop situation where we can assume equally strong talkers and distances between microphone and loudspeaker/talker.

The tested thresholds of the NPC detector are chosen such that its probability of false alarm go from approximately 0 to about 0.12. For these thresholds, we then estimate the corresponding probability of miss (see [7] for details of estimation of the probabilities). Results from the simulations are shown as the ROC in Fig. 3.

6.2 Data for the Two-Channel Case

For the two-channel simulations, we use real-life speech data described in [16]. The source data is from the same database as the data used in the previous section, [15]. These two-channel recordings however, are made in the HuMaNet room B [17] and are available on a CD (“Recording2”). When double-talk behaviour and changes in the receiving room are studied, the transmission room signals have been processed with a nonlinearity so that the echo canceler converges to a stable unique solution, [18].

Recorded receiving room data are used. In these situations, the microphone and loudspeaker setup in the receiving room is shown in Fig. 4, and estimated acoustic responses and magnitude functions are shown in Fig. 5 and 6.

Data and general algorithm settings are:

**Transmission Room Speech:** Stereo recordings from “Recording2” using talker M25 of [15]. The transmission room speech is pre-processed before emitted in the receiving room according to:

\[ x(\cdot) = x(\cdot) + \frac{\alpha}{2} [x(\cdot) \pm |x(\cdot)|] , \tag{58} \]

where \((\cdot)\) denotes either (left) or (right) channel. The distortion parameter of the nonlinearity is \(\alpha\). The positive half-wave is added to the left channel and the negative to the right.

**Background noise level:** \(\text{SNR} = 10 \log_{10} \left( \frac{\sigma^2_y}{\sigma^2_w} \right) \approx 38 \text{ dB} \).

**Adaptive filter parameters:** \( L = 1024 \) (64 ms), \( \lambda_f = 1 - \frac{1}{3L} \hat{\alpha}_w \), \( \alpha_0 = 4 \), \( \mu = 1 - \lambda_f \), \( \varphi = 0.99 \), \( \hat{h}(0) = 0 \).

**DTD parameters:** \( \lambda_b = (1 - \frac{2}{3L}) \hat{\alpha}_w \),

- **Non-robust adaptive filter:** \( T = 0.99 \Rightarrow P_\text{m} \approx 0.012, P_\text{f} \approx 0.06 \). (SNR = 30 dB.)
- **Robust adaptive filter:** \( T = 0.92 \Rightarrow P_\text{m} \approx 0.102, P_\text{f} \approx 0.0 \). (SNR = 30 dB.)

Table 1 presents the modified system of two-channel NPC DTD and frequency-domain algorithm used in the simulations this section. We use a constrained version where the tail coefficients are zeroed by the matrix \( W_1 \). The reason for this is that there otherwise will be a distortion induced from the implicit circular convolution of (13) when the echo path we are estimating is longer than our model filter.

The performance is measured by means of the mean square error (MSE). We estimate the MSE according to,

\[ \text{MSE}(n) = 10 \log_{10} \left\{ \frac{\text{LPF}([e(n) - v(n) - w(n)]^2)}{\text{LPF}([y_e(n)]^2)} \right\} \tag{59} \]

where LPF denotes a first-order lowpass filter with a pole at 0.999 which corresponds to a 3 dB bandwidth of 2.5 Hz.
Table 1 The two-channel NPC double-talk detector and echo canceler with modifications used in all simulations. $\varrho$ bounds the maximum allowed coherence between the channels.

<table>
<thead>
<tr>
<th>Spectrum estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_{s,j}(m) = \lambda_sS_{s,j}(m-1) + (1-\lambda_t)D_t^H(m)D_j(m), , i,j = 1,2$</td>
</tr>
<tr>
<td>$\tilde{S}<em>{s,j}(m) = S</em>{s,j}(m) + \text{diag}{\delta_{j,0}, \ldots, \delta_{j,2L-1}}, , j = 1,2$</td>
</tr>
<tr>
<td>$S_{j}(m) = \tilde{S}<em>{j,j}(m) \left[ I</em>{2L \times 2L} - \varrho^2 \left[ S_{1,1}(m)S_{2,2}(m) \right]^{-1} \right] S_{2,1}(m)S_{1,2}(m), , i,j = 1,2$</td>
</tr>
<tr>
<td>$K_1(m) = S_1^{-1}(m) \left[ D_1^H(m) - \varrho S_{1,2}(m)S_{2,2}^{-1}(m) D_2^H(m) \right], , 0 \leq \varrho \leq 1$</td>
</tr>
<tr>
<td>$K_2(m) = S_2^{-1}(m) \left[ D_2^H(m) - \varrho S_{2,1}(m)S_{1,1}^{-1}(m) D_1^H(m) \right], , 0 \leq \varrho \leq 1$</td>
</tr>
</tbody>
</table>

Double-talk detector (Background filter)

| $e_b(m) = y(m) - G \left[ D_1(m)h_{b,1}(m-1) + D_2(m)h_{b,2}(m-1) \right]$ |
| $h_{b,j}(m) = h_{b,j}(m-1) + (1-\lambda_b)K_f(m)e_b(m), \, j = 1,2$ |
| $s(m) = \lambda_b s(m-1) + (1-\lambda_b)D^H(m)y(m)$ |
| $\eta^2(m) = \left[ h_{b,1}^H(m)h_{b,2}^H(m) \right] s(m)$ |
| $\sigma_\vartheta^2(m) = \lambda_b \sigma_\vartheta^2(m-1) + (1-\lambda_b)\nu^H(m)\nu(m)$ |
| $\xi(m) = \eta(m)/\sigma_\vartheta(m) < T, \Rightarrow \text{double-talk}, \mu = 0$ |
| $\xi(m) = \eta(m)/\sigma_\vartheta(m) \geq T, \Rightarrow \text{no double-talk}, \mu = (1-\lambda_t)$ |

Echo canceler (Foreground filter)

| $e(m) = y(m) - G \left[ D_1(m)h_1(m-1) + D_2(m)h_2(m-1) \right]$ |
| $h_j(m) = h_j(m-1) + \mu F_WF^{-1}K_f(m)e(m), \, j = 1,2$ |

Definition

$W_i = \begin{bmatrix} I_{L\times L} & 0_{L\times L} \\ 0_{L\times L} & 0_{L\times L} \end{bmatrix}$

6.3 AEC Tracking and DTD Sensitivity to Transmission Room Echo Path Changes

In this section, we show the influence the nonlinear processing of the transmission signal has on the convergence to the true solution of the echo canceler. This is done by shifting the position of the transmission room speaker, thus changing the acoustic paths of the transmission room, $g_i(n), i = 1, 2$ in Figure 1. Any loss of MSE performance because of this change, is undesired and indicates that the channels have not been properly decorrelated and only a non-unique solution has been found by the echo canceler, [18].

Figure 7 shows the MSE when we have a transmission room speaker change at 5.2 seconds for unprocessed speech data ($\alpha = 0$) and for data processed with a nonlinearity with a distortion parameter of $\alpha = 0.5$. The algorithm has to track the transmission room change when $\alpha = 0$ while it is not as badly affected for $\alpha = 0.5$. Furthermore, in this case, the DTD does not false alarm because of the transmission path changes.

In the following simulations, we use $\alpha = 0.5$ so that we always have a solution close to the true receiving room echo paths when the canceler has converged in mean square error sense. In practice, a smaller $\alpha$ could be chosen.
6.4 AEC Tracking and DTD Sensitivity to Receiving Room Echo Path Changes

For the purpose of showing how the AEC/DTD system handles echo path changes in the receiving room, we have chosen to study the reconvergence when the echo path undergoes one of four changes after the algorithm has been adapting for about 2 seconds. We impose the same echo path change at two time instances so that variations of convergence rate due to input excitation variations are exemplified.

In reality, there are infinite possibilities for echo path variations. However, we have four cases that represent rather severe variations that may occur: Time delay/advance which occur if a loudspeaker or microphone is repositioned, and echo path gain variation of $+6$ or $-6$ dB which may occur if there is a corresponding increase or decrease in the loudspeaker volume. In order to have controlled echo path changes, the receiving room data has been manipulated such that we achieve the following changes at 2 and 5.2 seconds:

a) Time delay: $h_i(n) \rightarrow h_i(n-5)$, $i = 1, 2$.

b) Time advance: $h_i(n) \rightarrow h_i(n+14)$, $i = 1, 2$.

c) 6 dB increase of echo path gain: $h_i(n) \rightarrow 2h_i(n)$, $i = 1, 2$.

d) 6 dB decrease of echo path gain: $h_i(n) \rightarrow \frac{1}{2}h_i(n)$, $i = 1, 2$.

Figure 8 shows the performance of the echo canceler for the echo path situations described above. At the first time the echo path changes (2 seconds) both algorithms reconverge at the same rate. After the second echo path change at 5 seconds, the robust algorithm does not exploit the input speech as well as the non-robust algorithm and therefore shows slower reconvergence in this particular case.

Echo path changes do trigger false alarms in the double-talk detector. These false alarms are due to the rather high threshold chosen, $T = 0.99$. By studying the statistic $\xi$, (figures not shown here) it has been found that time delay/advance changes result in low values of $\xi$ during a short period of time while gain changes result in only a slight decrease of $\xi$ over a longer period of time. Choosing a lower threshold would decrease the sensitivity to echo path changes, particularly gain variations. In reality, time shifts of the whole impulse response is less probable. It is more likely that the direct path is unchanged while the tail changes slowly. These changes will not result in as many false alarms as the case studied in this section.

6.5 AEC/DTD Performance During Double-Talk

The performance of the echo canceler during double-talk is presented in this section. We show the divergence that results from presence of receiving room speech where we use two receiving room speakers (male/female) who speak the same phrase. This speech sequence is repeated twice such that two different scenarios of double-talking are shown.

Receiving room speech (double-talk): Added speech from a male talker (M51), and a female talker (F14), sentence (075): “Which tea party did Judge Baker go to?” [15].

Transmission room level: $\text{ENR} = 10 \log_{10} \left( \frac{\sigma_y^2}{\sigma_v^2} \right) \approx 0 \text{ dB}.$

Results from these simulations are shown in Figure 9. The NPC DTD does not detect double-talk in the beginning of the first burst at 4.2 seconds. This has the consequence that the non-robust algorithm starts to diverge, especially for the female talker, Fig. 9d. The robust algorithm shows much slower divergence during this difficult situation. For the second double-talk burst, the NPC DTD is accurate enough so that divergence is avoided completely for both algorithms.
7 Conclusions

We have presented a double-talk detector based on a pseudo coherence measure and a multi-channel frequency-domain adaptive algorithm suitable for acoustic echo cancellation. Together, this DTD and the frequency-domain algorithm, make a system that shows high performance with respect to convergence rate and good behaviour during double-talk. Evaluation of this system has been made for the two-channel case using real-life recorded data. Significant results with the proposed system are:

- The DTD is independent of any assumptions about the echo path e.g., the echo path gain. This is an important property since the variability of acoustic echo paths is high.
- False alarms of the DTD triggered by echo path variations can be kept low by lowering the threshold. Possible divergence due to detection misses can be significantly reduced by using the robust version of the algorithm.
- Looking at the equations for the frequency-domain algorithm, Table 1, we find that they are based on computations involving simple arithmetic and Fourier transforms. Even though we have not investigated the dynamic range of the computations, these calculations can most likely, with some effort, be successfully implemented in a fixed-point signal processing platform and thus providing a cost efficient solution for multi-channel echo cancellation.

The continuation of this work will be to implement the ideas on a real-time platform for evaluation of multi-channel audio communication in a desktop situation.

Acknowledgments

The authors would like to thank G. W. Elko and M. M. Sondhi for discussions and comments and Peter Eneroth for sharing his two-channel AEC graphic.

Tomas Gaensler

MH-10009657-TG/JB-tg

Jacob Benesty

Atts.
References
Figures 2-9
Appendices A-C
References


Figure 2: Echo path response used to simulate the acoustic path when the ROC of the NPC DTD is estimated.
Figure 3: Receiver operating characteristic (ROC) of the NPC DTD. ENR = 0 dB and SNR = 30 dB.
Figure 4: Receiving room microphone and loudspeaker setup.

Figure 5: Estimated receiving room impulse responses. (a) From left speaker to left microphone. (b) From right speaker to right microphone. (c) From left speaker to right microphone. (d) From right speaker to left microphone.
Figure 6: Magnitude of the frequency response of the receiving room. Other conditions as in Fig. 5.
Figure 7: (a) Left channel transmission speech, the speaker position changes at 5.2 seconds. (b) Mean square error when echo paths of the transmission room changes: $\alpha = 0.5$ (Solid line), $\alpha = 0$ (Dash-dotted line).
Figure 8: Mean square error of the two-channel frequency-domain algorithm with the pseudo coherence based DTD. The echo paths in the *receiving room* changes as: (a) Time delay of 5 samples. (b) Time advance of 14 samples. (c) 6 dB increase of echo path gain. (d) 6 dB decrease of echo path gain. The rectangles in the bottom of the figures indicate where double-talk has been detected (in this case, it is false alarms). This is the detection results with $T = 0.99$ used with the non-robust algorithm. Solid line: Robust algorithm. Dashed line: Non-robust algorithm.
Figure 9: Double-talk situation. (a) Left channel transmission room speech (upper), receiving room speech, male speaker (lower). (b) Mean square error of left channel, male speaker. (c) Left channel transmission room speech (upper), receiving room speech, female speaker (lower). (d) Mean square error of left channel, female speaker. Other conditions same as in Fig. 8.
A The Robust Multi-Channel Frequency-Domain Adaptive Algorithm

As in derivations of time-domain adaptive algorithms, we start by forming a criterion which is minimized with respect to its filter coefficients. Most commonly, the choice is a quadratic criterion that corresponds to a maximum likelihood estimator when the underlying noise distribution is Gaussian. However, in our case, we choose to look at the maximum likelihood criterion derived from a non-Gaussian assumption of the noise. Modeling the noise with a probability density function (PDF) that has a tail which is heavier than the Gaussian PDF gives us a non-quadratic function to minimize that results in an outlier-robust algorithm, [9].

We choose to work with the following criterion:

\[
J(\hat{h}) = \sum_{p=0}^{mL-1} \rho \left[ \frac{|e(p)|}{s_c(m)} \right],
\]

(60)

where \(\rho[\cdot]\) is a convex function given in (56), and \(s_c(m)\) is a real positive scale factor for block \(m\) discussed below.

Adaptive Newton-type algorithms, [19], minimize the criterion (60) by using the recursion:

\[
\hat{h}(m) = \hat{h}(m-1) - \mu S_{\psi^{-1}} J \left[ \hat{h}(m-1) \right],
\]

(61)

where \(\nabla J(\hat{h})\) is the gradient of the optimization criterion w.r.t. \(\hat{h}\) defined as \(\nabla J(\hat{h}) = \frac{\partial}{\partial \hat{h}} \left( J(\hat{h}) \right)\). \(S_{\psi^{-1}}\) is (an approximation of) \(E \left\{ \nabla^2 J(\hat{h}(m-1)) \right\}\), and \(\mu\) is the relaxation parameter.

In order to proceed we thus need to calculate the gradient and Hessian of (60). These calculations are made in detail in Appendix B where it is shown that,

\[
\nabla J(\hat{h}) = -\frac{1}{s_c} D^H(m) G^H F^{-H} \psi(e(m)),
\]

(62)

where

\[
\psi(e(m)) = \begin{bmatrix}
0_{L\times 1} \\
\psi \left[ \frac{e(mL)}{s_c(m)} \right] \text{sign}[e(mL)] \\
\vdots \\
\psi \left[ \frac{e(mL + L - 1)}{s_c(m)} \right] \text{sign}[e(mL + L - 1)]
\end{bmatrix},
\]

(63)

and

\[
\psi(|z|) = \rho'(|z|) = \min \{|z|, k_0\}.
\]

(64)

The Hessian is given by,

\[
\nabla^2 J(\hat{h}) = \frac{1}{s_c} D^H(m) G^H F^{-H} \Psi' \left[ e(m) \right] F^{-1} G D(m),
\]

(65)

where

\[
\Psi' \left[ e(m) \right] = \text{diag} \left\{ 0_{1\times L} \psi' \left[ \frac{e(mL)}{s_c(m)} \right] \ldots \psi' \left[ \frac{e(mL + L - 1)}{s_c(m)} \right] \right\},
\]

(66)

By recursive averaging of (65) we find an estimate of \(E \left\{ \nabla^2 J(\hat{h}(m-1)) \right\}\),

\[
S_{\psi^{-1}} = (1 - \lambda_t) \frac{1}{s_c} \sum_{p=0}^{m} \lambda_t^{m-p} D^H(p) G^H F^{-H} \Psi' \left[ e(p) \right] F^{-1} G D(p),
\]

(67)
where $\lambda_f (0 < \lambda_f \leq 1)$ is an exponential forgetting factor.

Combining (61), (62), and (67), we find the robust frequency-domain algorithm:

$$
\hat{h}(m) = \hat{h}(m-1) + \frac{H}{s_c} S^{-1}_{\psi'}(m) D^H(m) F^{-H} \psi [e(m)].
$$

(68)

$S_{\psi'}$ is in general a non-diagonal matrix, hence (67) is complex from a computational point of view. To simplify this algorithm we can use the same approximation as for the non-robust frequency-domain method in Section 2. Even though $\Psi'[e(m)]$ is diagonal, it is practical to further approximate is as,

$$
\Psi'[e(m)] = \min_p \psi'_p \left[ \frac{|e(p)|}{s_c} \right] \mathbf{I}_{2L \times 2L} = \psi'_\text{min} \mathbf{I}_{2L \times 2L}.
$$

(69)

This approximation reduces complexity and gives freedom for control of the algorithm during convergence and tracking as discussed in Section 5.

We now write $S_{\psi'}$ as,

$$
S_{\psi'} = \frac{\psi'_\text{min}}{s_c^2} S_u(m),
$$

(70)

where $S_u(m)$ given by (29).

A simplified version of (68) then becomes:

$$
\hat{h}(m) = \hat{h}(m-1) + \frac{H s_c}{\psi'_\text{min}} S^{-1}_u(m) D^H(m) F \psi [e(m)].
$$

(71)

The value of the derivative of (64) is either 1 or 0. Obviously, it is not appropriate to let $\psi '[\cdot]$ become zero since it is a divisor in (68). We therefore bound it in such a way that (68) always remains stable.

### A.1 Scale Factor Estimation

An important part of the algorithm is the estimation of the scale factor $s_c$. Traditionally, the scale is used to make the robust algorithm invariant to the noise level, [9]. In our approach however, it should reflect the minimum mean square error, be robust to short burst disturbances (double-talk in our application) and track slow changes of the residual error (echo path changes). We have chosen the scale factor estimate as

$$
s_c(m+1) = \lambda_s s_c(m) + (1 - \lambda_s) \frac{m L + L - 1}{L \beta} \sum_{p=mL} S_u(m) \psi \left[ \frac{|e(p)|}{s_c(m)} \right],
$$

(72)

$$
s_c(0) = \sigma_z,
$$

which is very simple to implement. $\beta$ is a constant related to $k_0$ in (64) chosen such that the scale factor is an unbiased estimate of the standard deviation of the ambient noise when the noise is assumed to be Gaussian (see Appendix C for more details). The choice of this method of estimating $s_c$ is justified in [10, 12]. With this choice of scale factor estimate, the current estimate of $s_c$ is governed by the level of the error signal in the immediate past over a time interval roughly equal to $L/(1 - \lambda_s)$ samples. The trade off between robustness and tracking rate of the robust frequency-domain adaptive algorithm is governed by the tracking rate of the scale estimator which is controlled by one single parameter $\lambda_s$. 
B Calculation of the Gradient and Hessian

In this appendix we derive the gradient and Hessian of (60).

\[
\nabla J [\mathbf{h}] = \frac{\partial}{\partial \mathbf{h}} \left( J [\mathbf{h}] \right) = \sum_{p=mL}^{mL+L-1} \frac{\partial}{\partial \mathbf{h}} \rho \left[ \frac{|e(p)|}{s_c} \right]
\]

\[
= \sum_{p=mL}^{mL+L-1} \frac{\partial}{\partial \mathbf{h}} \left[ e^*(p) \frac{\partial}{\partial e(p)} |e(p)| \frac{\partial}{\partial |e(p)|} \rho \left[ \frac{|e(p)|}{s_c} \right] \right]
\]

\[
= \sum_{p=mL}^{mL+L-1} \frac{\partial}{\partial \mathbf{h}} \left[ e^*(p) \rho' \left[ \frac{|e(p)|}{s_c} \right] \text{sign} [e(p)] \right] \frac{1}{s_c}
\]

\[
= \sum_{p=mL}^{mL+L-1} -D^H (m) G^H F^H 1_{p-mL+L} \Psi \left[ \frac{|e(p)|}{s_c} \right] \text{sign} [e(p)] \frac{1}{s_c}
\]

\[
= -\frac{1}{s_c} D^H (m) G^H F^H \Psi [e(m)]. \tag{73}
\]

where 1_{p-mL+L} is a 2L × 1 vector containing a 1 in position p − mL + L and zeros in all other positions. \(\Psi [e(m)]\) is defined in (63).

For the Hessian we have,

\[
\nabla^2 J = \left[ \frac{\partial}{\partial \mathbf{h}} \left( \frac{\partial}{\partial \mathbf{h}} J \right) \right]^H = \left[ -\frac{\partial}{\partial \mathbf{h}} \Psi^H [e(m)] F^{-1} G D \right], \tag{74}
\]

Look at one non-zero element of \(\Psi [e(m)]^H\),

\[
\frac{\partial}{\partial \mathbf{h}} \left( \Psi \left[ \frac{|e(p)|}{s_c} \right] \text{sign}^* [e(p)] \right) = \frac{\partial}{\partial \mathbf{h}} \left[ e^*(p) \frac{\partial}{\partial e(p)} |e(p)| \frac{\partial}{\partial |e(p)|} \psi \left[ \frac{|e(p)|}{s_c} \right] \text{sign}^* [e(p)] \right]
\]

\[
= -D^H (m) G^H F^H 1_{p-mL+L} \frac{\text{sign} [e(p)]}{s_c} \psi' \left[ \frac{|e(p)|}{s_c} \right] \text{sign}^* [e(p)]. \tag{75}
\]

We can now write

\[
\frac{\partial}{\partial \mathbf{h}} \Psi^H [e(m)] = \frac{1}{s_c} D^H (m) G^H F^H \Psi' [e(m)]. \tag{76}
\]

Equation (74) and (76) gives us the final result,

\[
\nabla^2 J = \frac{1}{s_c^2} D^H (m) G^H F^H \Psi' [e(m)] F^{-1} G D \tag{77}
\]
C Calculation of the Constant $\beta$

The constant $\beta$ in (72) is a function of the truncation point $k_0$. $\beta$ can be chosen such that the scale factor is an unbiased estimate of the standard deviation of the ambient noise when it is Gaussian distributed, hence,

$$
\beta = \int_{-\infty}^{k_0} \min\{|z|, k_0\} \frac{1}{\sqrt{2\pi}} e^{-\frac{|z|^2}{2}} dz
$$

$$
= \sqrt{\frac{2}{\pi}} \left[ 1 - e^{-\frac{k_0^2}{2}} \right] + k_0 \sqrt{\frac{2}{\pi}} \int_{k_0}^{\infty} e^{-\frac{|z|^2}{2}} dz.
$$
Title: Multi-Channel Acoustic Echo and Double-Talk Handling: A Frequency-Domain Approach

Authors
Tomas Gaensler
gaensler@bell-labs.com
Jacob Benesty
jbenesty@bell-labs.com

Electronic Address
Location
MH 2D-514
MH 2D-518

Phone
(908) 582-7287
(908) 582-7643

Company (if other than Lucent–BL)

Document No.
10009657-000525-09TM
Filing Case No.
38794-43
Work Project No.
MA30033002

Keywords:
Adaptive Filter, Frequency-Domain, Robustness, Double-Talk Detector, Acoustic Echo Cancellation, Multi-Channel

MERCURY Announcement Bulletin Sections
CMM– Communications
MAS– Math and Statistics

Abstract

Multi-channel audio communication is going to be the future challenge in hands-free multi-media, network and wireless communication systems. Multi-channel audio significantly increases the naturalness and perceived quality in a conversation. A key component for hands-free, full-duplex, communication technology is the echo control device, primarily the echo canceler. In this memorandum, we describe a system for echo cancellation and double-talk control based on frequency-domain algorithms. The advantages of this approach are among others: algorithm stability, fast convergence and tracking, low computational complexity, and simplicity of implementation. In detail, we derive a new way of calculating a statistic for double-talk detection, based on an adaptive multi-channel frequency-domain algorithm and a robust multi-channel frequency-domain adaptive algorithm.
Future Lucent Distribution

Release to any Lucent employee (excluding contract employees)

Author Signatures

_________________________  _______________________
Tomas Gaensler            Jacob Benesty

Organizational Approvals

_________________________  _______________________
M. M. Sondhi              G. W. Elko

B. H. Juang - DIR

For Use by Recipient of Cover Sheet:

Computing network users may order copies via the library -l command; for information, type man library after the UNIX system prompt.

Otherwise:
Enter your Company ID number._________________________
Return this sheet to any ITDS location.