Special Issue

Partitioned Residual Echo Power Estimation for Frequency–Domain Acoustic Echo Cancellation and Postfiltering

GERALD ENZNER, RAINER MARTIN, PETER VARY
Institute of Communication Systems and Data Processing
Aachen University of Technology, D-52056 Aachen, Germany
Phone: *+49-241-8026960*, E-mail: {*enzner,martin,vary*}*@ind.rwth-aachen.de*

Abstract. Residual echo arises in hands–free telephony equipment due to insufficient echo canceler convergence, but can be suppressed using a postfilter. The residual echo power spectral density is the most crucial control parameter for both frequency–domain acoustic echo cancellation and combined residual echo and noise postfiltering. In this contribution we present and compare residual echo power spectral estimation techniques. We introduce a new partitioned block-adaptive estimation technique delivering considerably improved residual echo estimates in strongly reverberant and noisy acoustic environments. We show that the adaptation loop of the frequency–domain adaptive filter (FDAF) can be used simultaneously for residual echo power estimation and tracking of the echo path impulse response. In this way, the FDAF and the postfilter concept supplement each other in a true synergy with low complexity. The resulting echo and noise control system proves to be robust in double talk situations as well.

1 INTRODUCTION

In the acoustic environment of mobile hands-free telephones we have to expect low signal–to–noise ratios and considerable acoustic feedback at the local microphone. It has been shown that a combined acoustic echo and noise reduction postfilter substantially improves the performance of the more traditional echo cancellation and noise reduction approach [1, 2].

A true synergy of acoustic echo cancellation and postfiltering can be obtained if both algorithms are implemented in the frequency domain. That leads to the concept of joint control of acoustic echo cancellation and postfiltering based on residual echo estimation. This was proposed in [3, 4] within the framework of echo compensation in sub-bands.

The control of our algorithm relies on the power spectral density (PSD) of the residual echo which is required for both frequency–domain adaptive echo cancellation [5] and postfiltering [1]. The residual echo PSD, however, cannot be directly measured and must be estimated from the available signals. Conventional block oriented approaches [1, 6] with limited DFT length (due to delay and complexity constraints) can only reflect the residual echo within the first DFT length of the residual echo impulse response, as illustrated in Figure 1. This leads to a serious bias of the residual echo estimate.

Thus, we propose a new unbiased residual echo PSD estimator, based on coherence, which conceptually takes the full length of the residual echo system as well as short–term correlations into account. The idea behind the new approach is to compute the total residual echo PSD as a sum over multiple delayed DFT frames of short length. This leads to the concept of a partitioned residual echo power estimator. The total residual echo PSD is then utilized to determine the optimum spectral weights for joint residual echo and background noise suppression, while the individual contributions can be used to control individual sections of a partitioned frequency domain adaptive filter (FDAF).

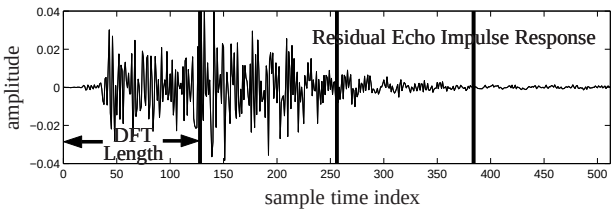

Figure 1: Partitioned residual echo impulse response.

Our full-duplex echo and noise control system with a single loudspeaker and a single microphone is depicted in

Figure 2. All signals are represented by their Fourier transforms, for example the microphone signal $y(i)$ by

$$Y(\Omega) = S(\Omega) + N(\Omega) + D(\Omega) \ , \tag{1}$$

where $S(\Omega)$, $N(\Omega)$, and $D(\Omega)$ represent clean near speech, background noise, and acoustic echo, respectively. During adaptation, the echo canceler $W(\Omega)$ is supposed to yield a robust but possibly inaccurate estimate $\hat{D}(\Omega)$ of the acoustic echo. The residual echo and background noise shall be therefore suppressed by the postfilter $H(\Omega)$ with input signal

$$\begin{aligned} E(\Omega) &= Y(\Omega) - \hat{D}(\Omega) \\ &= S(\Omega) + N(\Omega) + B(\Omega) \end{aligned} \tag{2}$$

where $B(\Omega) = D(\Omega) - \hat{D}(\Omega)$ is the Fourier transform of the residual echo signal. In the receiving and sending path of the telephone we have the far end speech $X(\Omega)$ and the estimated local speech $\hat{S}(\Omega)$, respectively.

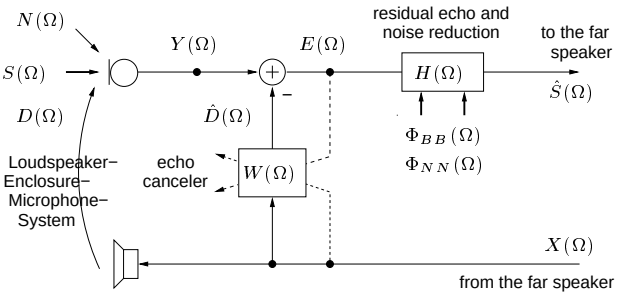

Figure 2: Combined echo and noise reduction system for mobile hands-free telephony.

The remainder of our paper is organized as follows: In Section 2 we will recall the concept of (partitioned) frequency–domain adaptive echo cancellation and combined residual echo and background noise suppression. We will observe that both the echo canceler and the postfilter rely on the same control parameter, the residual echo PSD. In Section 3 we will briefly discuss a previously proposed coherence estimator for the residual echo PSD, which serves as the basis for further algorithm development. The main part of this paper, Section 4, addresses three major problems related to residual echo PSD estimation in practice: finite DFT block length, stationary background noise, and non-stationary signals. Solutions will be proposed for each case and the corresponding benefits will be discussed. In order to substantiate our algorithm, we provide a theoretical analysis of partitioned residual echo PSD estimation (using the FDAF). Eventually, in Section 5 we will confirm our results by simulations.

2 Combined Acoustic Echo and Noise Control System

In this Section we present an overview of our acoustic echo and noise control system.

We employ the partitioned frequency–domain adaptive echo canceler which uses the overlap and save implementation [5, 7, 8, 9, 10] to reduce the numerical complexity and an appropriate sectioning of the filter impulse response [5, 9] to limit the algorithmic signal delay. The limited adaptation rate of the echo canceler in noisy environments, however, leads to insufficient echo attenuation especially during the transient phase of the adaptation process.

The postfilter is designed to achieve residual echo suppression as long as the echo canceler cannot ensure sufficient echo attenuation. As soon as the echo canceler reaches a sufficient convergence state, the responsibility for echo control is gradually taken away from the postfilter in order to maintain the highest near speech quality. In this case, the postfilter ideally performs background noise suppression only.

The interaction of echo cancellation and postfiltering is enabled on the basis of the residual echo PSD, which is the control parameter that both algorithms have in common.

2.1 Frequency Domain Adaptive Echo Cancellation

The block processing approach of the frequency–domain adaptive filter (FDAF) introduces a signal delay which increases with increasing block length (filter length). The signal delay can be reduced if the filter is divided into several filter partitions [9], and each partition is implemented with the overlap and save method.

We summarize the FDAF algorithm for a single partition as follows:

The Discrete Fourier Transform (DFT) $Y(\Omega_\ell, kR)$ of $Y(\Omega)$ at frame index $k \in \mathbb{Z}$ is obtained from the windowed time domain signal $y(i)$ with sampling time index i as

$$Y(\Omega_\ell, kR) = \sum_{i=0}^{M-1} y((k-1)R+i) w_y(i) \mathrm{e}^{-j\Omega_\ell i} \tag{3}$$

with frame shift R and the normalized discrete frequency index $\Omega_\ell = 2\pi\ell/M$ for $\ell = 0, 1, \ldots, M-1$. The rectangular window function applied to signal $y(i)$ is defined as

$$w_y(i) = \begin{cases} 1 & \text{for } R \le i \le M-1 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

and we use $M = 2R$. The same notation holds for the DFT coefficients of $\tilde{X}(\Omega_\ell, kR)$, $S(\Omega_\ell, kR)$, $N(\Omega_\ell, kR)$, and $E(\Omega_\ell, kR)$ with corresponding window functions $w_{\tilde{x}}(i) = w_s(i) = w_n(i) = w_e(i) = w_y(i)$.

Furthermore, we define the DFT of a two-frame extended excitation signal as

$$X(\Omega_\ell, kR) = Z(\Omega_\ell)\tilde{X}(\Omega_\ell, (k-1)R) + \tilde{X}(\Omega_\ell, kR) \tag{5}$$

where the vector $Z(\Omega_\ell) = (+1, -1, \ldots, +1, -1)$ of length M performs a cyclic shift in the time–domain. The spectrum $X(\Omega_\ell, kR)$ might be obtained from the analysis defined in Equation (3) as well, using the effective window function $w_x(i) = w_y(i) + w_y(i+R)$.

The FDAF algorithm now calculates the error spectrum

$$E(\Omega_\ell, kR) = DFT\{P\{IDFT\{ Y(\Omega_\ell, kR) - X(\Omega_\ell, kR)W(\Omega_\ell, kR)\}\}\}\,, \tag{6}$$

where $W(\Omega_\ell, kR)$ is the vector of adaptive weights and P is the projection operation used in the filter part of the algorithm, which returns its operand for $\ell = M/2, \ldots, M-1$ and zero otherwise.

In the adaptation part of the adaptive filter, the weight vector $W(\Omega_\ell, kR)$ is updated by

$$W(\Omega_\ell, kR+1) = W(\Omega_\ell, kR) + DFT\{Q\{IDFT\{ \mu(\Omega_\ell, kR)X^*(\Omega_\ell, kR)E(\Omega_\ell, kR)\}\}\}\,, \tag{7}$$

where Q is the projection operation used in the adaptation part of the algorithm, which returns its operand for $\ell = 0, \ldots, M/2-1$ and zero otherwise. $\mu(\Omega_\ell, kR)$ denotes the real–valued adaptive step-size factor.

The block length of the adaptive filter might be significantly shorter than the reverberation time of the acoustic environment. In that case we apply several partitions of adaptive weights in order to cover a significant length of the actual loudspeaker–enclosure–microphone (LEM) system. In particular, we adapt L sets of independent weights

$$W^{(\lambda)}(\Omega_\ell, kR)\,, \quad 0 \le \lambda \le L-1\,, \tag{8}$$

according to (7)

$$W^{(\lambda)}(\Omega_\ell, kR+1) = W^{(\lambda)}(\Omega_\ell, kR) + DFT\{Q\{IDFT\{ \mu^{(\lambda)}(\Omega_\ell, kR)X^{*(\lambda)}(\Omega_\ell, kR)E(\Omega_\ell, kR)\}\}\}\,, \tag{9}$$

using the excitation spectra

$$X^{(\lambda)}(\Omega_\ell, kR) = X(\Omega_\ell, (k-\lambda)R) \tag{10}$$

and the compound error spectrum

$$E(\Omega_\ell, kR) = DFT\{P\{IDFT\{ Y(\Omega_\ell, kR) - \sum_{\lambda=0}^{L-1} X^{(\lambda)}(\Omega_\ell, kR)W^{(\lambda)}(\Omega_\ell, kR)\}\}\}\,. \tag{11}$$

In order to obtain an expression for the individual step-size factor $\mu^{(\lambda)}(\Omega_\ell, kR)$ for partition λ, we first define the convergence state

$$|G^{(\lambda)}(\Omega_\ell, kR)|^2 = |W^{(\lambda)}_{LEM}(\Omega_\ell, kR) - W^{(\lambda)}(\Omega_\ell, kR)|^2\,. \tag{12}$$

$W^{(\lambda)}_{LEM}(\Omega_\ell, kR)$ denotes the DFT coefficients of the corresponding section of the LEM system. Then, we adopt a result of [5]: By minimizing the average convergence state of a certain partition λ of the adaptive filter, we obtain the individual step–size

$$\mu^{(\lambda)}(\Omega_\ell, kR) = \frac{\Phi^{(\lambda)}_{BB}(\Omega_\ell, kR)}{\Phi_{EE}(\Omega_\ell, kR)} \cdot \frac{1}{\Phi^{(\lambda)}_{XX}(\Omega_\ell, kR)}\,. \tag{13}$$

for this partition. $\Phi^{(\lambda)}_{BB}(\Omega_\ell, kR)$ represents a short–time estimate of the residual echo PSD at time k which is associated with the misalignment of the weights $W^{(\lambda)}(\Omega_\ell, kR)$ of partition λ. $\Phi_{EE}(\Omega_\ell, kR)$ and $\Phi^{(\lambda)}_{XX}(\Omega_\ell, kR) = \Phi_{X^{(\lambda)}X^{(\lambda)}}(\Omega_\ell, kR)$ further denote short–time PSD estimates of the error signal and the excitation, respectively.

2.2 Frequency Domain Adaptive Postfiltering

Spectral weighting of DFT coefficients

$$\hat{S}(\Omega_\ell, kR) = H_W(\Omega_\ell, kR)E(\Omega_\ell, kR) \tag{14}$$

on the basis of the Wiener rule

$$H_W(\Omega_\ell, kR) = = \frac{\Phi_{SS}(\Omega_\ell, kR)}{\Phi_{SS}(\Omega_\ell, kR) + \Phi_{NN}(\Omega_\ell, kR) + \Phi_{BB}(\Omega_\ell, kR)} \tag{15}$$

can be viewed as a simple form of DFT based speech enhancement to be possibly applied in the postfilter.

Accurate (short–time) estimates $\Phi_{NN}(\Omega_\ell, kR)$ and $\Phi_{BB}(\Omega_\ell, kR)$ of the background noise PSD and the residual echo PSD are crucial for the reliability of the spectral weights $H_W(\Omega_\ell, kR)$. The background noise PSD can be determined adaptively and accurately by the Minimum Statistics approach [11, 12], where the desired noise PSD can be tracked even during speech activity. The residual echo PSD is drawn from the new partitioned estimator to be derived in Section 4.

Instead of Wiener filtering, we actually use the more advanced MMSE-LSA spectral weighting algorithm [13] which, however, relies in a similar way on residual echo and noise PSD estimates.

2.3 Interaction of Echo Cancellation and Postfiltering

An acoustic echo canceler which is realized as an adaptive filter is naturally characterized by finite convergence

speed. The adaptation rate is specifically dependent on the echo–to–noise power ratio of the acoustic environment. During the process of adaptation, the postfilter is designed to achieve additional attenuation in the feedback loop of the hands–free telephone. The required short–time residual echo PSD can (and must) be tracked at a higher rate than the adaptive filter coefficients, consequently. This is enabled by properly choosing tracking constants of the adaptive filter and the residual echo PSD estimator. While the echo canceler converges, the residual echo PSD decreases and the responsibility for echo control is gradually taken away from the postfilter. Thereby, the near end speech quality is increased during double talk.

Ideally, after convergence, the echo canceler achieves complete feedback attenuation, while the postfilter performs background noise suppression only. In practice, however, to drive the echo canceler into the state of ideal convergence (no misalignment) we would need the perfect step-size control and, according to Equation (13), a residual echo PSD estimate with very high accuracy. That in turn becomes very difficult while approaching the state of total convergence. Therefore, the very last part of acoustic echo control must always be handled by the postfilter.

Concerning the symbiosis of the echo canceler and the postfilter, we recall that both algorithms are implemented in the frequency domain. Consequently, DFT and IDFT operations required for analysis and synthesis are efficiently shared in our echo and noise control system. In particular, the postfilter is performing speech enhancement directly upon the frequency–domain error signal $E(\Omega_\ell, kR)$ provided by the echo canceler. Moreover, the residual echo PSD estimator to be discussed in the following Sections makes use of the already available DFT coefficients as well. Obviously that results in an algorithm which is highly efficient from the viewpoint of computational complexity.

3 Coherence Analysis

The residual echo PSD estimator to be proposed in this paper as well as previously proposed algorithms make extensive use of the coherence function. We therefore briefly review coherence analysis and the relation to residual echo power estimation.

The spectrum of the residual echo is given by

$$B(\Omega) = G(\Omega)X(\Omega), \tag{16}$$

where $G(\Omega)$ is the residual echo transfer function. Under the assumption of statistically independent $S(\Omega)$, $N(\Omega)$, and $X(\Omega)$, we can write

$$G(\Omega) = \frac{\Phi_{XE}(\Omega)}{\Phi_{XX}(\Omega)} \tag{17}$$

by the use of the cross PSD $\Phi_{XE}(\Omega)$ of the signals $X(\Omega)$ and $E(\Omega)$. Then, we obtain

$$\Phi_{BB}(\Omega) = |G(\Omega)|^2 \, \Phi_{XX}(\Omega) \tag{18}$$

for the residual echo PSD. This can be expressed equivalently [6] by

$$\Phi_{BB}(\Omega) = C_{XE}(\Omega)\Phi_{EE}(\Omega) \tag{19}$$

using the magnitude squared coherence function

$$C_{XE}(\Omega) = \frac{|\Phi_{XE}(\Omega)|^2}{\Phi_{XX}(\Omega)\Phi_{EE}(\Omega)} \tag{20}$$

of the signals $X(\Omega)$ and $E(\Omega)$.

The result can be implemented approximately [6] on the basis of Welch's power spectral estimation technique [8], or recursive averaging of periodograms which accounts for the short term stationarity of speech signals. The latter one is written with $0 < \alpha < 1$ as

$$\Phi_{XE}(\Omega_\ell, kR) = \alpha \, \Phi_{XE}(\Omega_\ell, (k-1)R) + (1-\alpha) I_{XE}(\Omega_\ell, kR) \tag{21}$$

using the (cross) periodogram $I_{XE}(\Omega_\ell, kR)$ between the signals $X(\Omega_\ell, kR)$ and $E(\Omega_\ell, kR)$

$$I_{XE}(\Omega_\ell, kR) = \frac{X^*(\Omega_\ell, kR)E(\Omega_\ell, kR)}{\sum_{i=0}^{M-1} w_x(i)w_e(i)} \, . \tag{22}$$

The approach is conceptually clear, however, in practice we observed biased estimates of the residual echo PSD. This is due to insufficient coverage of the residual LEM impulse response by the DFT length and due to short term correlations of otherwise independent speech and background noise signals. Therefore, the coherence method still has potentials for considerable improvements with regard to residual echo estimation. This will be shown more detailed in the next Section and by simulations.

4 Multiple–Frame (Partitioned) Coherence Analysis

We will address in detail three practical problems associated with residual echo power spectral estimation on the basis of the coherence function:

1. Limited DFT block length
2. Stationary local background noise
3. Non-stationary local disturbances and non-stationary excitation (speech)

We will specifically discuss the impact on the quality of the residual echo estimate and subsequently propose a solution with regard to each case.

4.1 LIMITED DFT BLOCK LENGTH

The DFT length of typical speech enhancement systems is around 128 to 256 speech samples (due to signal delay and complexity constraints). Consequently, a DFT based residual echo estimator on the basis of equations (19), (20), and (21) will certainly fail to reflect the full correlation between the echo compensated signal and the far speech. Thus, one obtains systematically underestimated residual echo, especially for acoustic environments with large reverberation time and algorithms which use a relatively short echo canceler.

In order to take the full length of the residual echo system into account, while using block processing with limited DFT length, we propose the partitioned residual echo power estimation concept, based on coherence. This will be followed by a theoretic analysis of partitioned coherence estimation in order to validate the approach.

4.1.1 Partitioned Residual Echo Estimation

The residual acoustic echo in frame $E(\Omega_\ell, kR)$ is obviously correlated with the present and past frames $X(\Omega_\ell, (k-\lambda)R)$ of the excitation signal (corresponding to partitions of the residual echo system). With regard to the exponential decay of a causal residual echo impulse response, we may have to consider only a limited number L of most recent frames

$$X^{(\lambda)}(\Omega_\ell, kR) = X(\Omega_\ell, (k-\lambda)R), \quad 0 \le \lambda \le L-1\ . \tag{23}$$

A partial estimate of the residual echo PSD being due to the individual frame $X^{(\lambda)}(\Omega_\ell, kR)$ of length M is then written as

$$\Phi_{BB}^{(\lambda)}(\Omega_\ell, kR) = C_{X^{(\lambda)}E}(\Omega_\ell, kR)\,\Phi_{EE}(\Omega_\ell, kR) \tag{24}$$

according to (19). The estimator computes the total residual echo PSD by adding the contributions of L partitions λ

$$\Phi_{BB}(\Omega_\ell, kR) = \sum_{\lambda=0}^{L-1} \Phi_{BB}^{(\lambda)}(\Omega_\ell, kR) \tag{25}$$

where we assumed mutual statistical independence of the excitation frames $X(\Omega_\ell, kR)$. This is not exactly true in the case of speech excitation. Simulations, however, show that the approach can be successfully employed for frame-based acoustic echo suppression (if the DFT length is not extremely short).

Partitioned residual echo power estimation will be further justified by the analysis in the following Section.

4.1.2 Partitioned Residual Echo PSD Estimation Using the FDAF

The analysis presented in this section shows that the partitioned FDAF algorithm can be quite naturally combined with the partitioned residual echo estimator. We will prove that the proposed algorithm (24) performs exact partitioning of the residual echo system in the case of white noise excitation.

We assume that the impulse response $g(i)$ of the residual echo system can be modeled by a causal IIR filter and, thus, the output of the residual echo system is given by the linear convolution

$$e(i) = g(i) * x(i) = \sum_{m=0}^{\infty} g(m)x(i-m) \tag{26}$$

where $x(i)$ is a stationary far end excitation signal with power σ_{xx}^2 and auto-correlation $r_{xx}(p)$.

The cross-correlation in the FDAF adaptation loop (21) of partition λ yields

$$X^*(\Omega_\ell, (k-\lambda)R)E(\Omega_\ell, kR) = \sum_{q=-\infty}^{\infty}\sum_{m=0}^{\infty} g(m)\delta(q-\lambda R)X^*(\Omega_\ell, kR-q)\tilde{X}(\Omega_\ell, kR-m) \tag{27}$$

where $\delta(q) = 1$ for $q = 0$ and $\delta(q) = 0$ otherwise.

Consequently, the cross power spectral density $\Phi_{XE}^{(\lambda)}(\Omega_\ell)$, which is required to compute the coherence of partition λ, is obtained by statistical expectation from the cross periodogram $I_{X^{(\lambda)}E}(\Omega_\ell, kR)$ as

$$\Phi_{XE}^{(\lambda)}(\Omega_\ell) = E\{I_{X^{(\lambda)}E}(\Omega_\ell, kR)\} = = \sum_{p=-(M-1)}^{M-1} \big(r_{xx}(p) * g(p+\lambda R)\big) r_{w_x w_e}(p) e^{-j\Omega_\ell p}\ , \tag{28}$$

using the normalized cross-correlation $r_{w_x w_e}(p)$ of the window functions $w_x(i)$ and $w_e(i)$

$$r_{w_x w_e}(p) = \frac{\sum_{i=0}^{M-1} w_x(i)w_e(i+p)}{\sum_{i=0}^{M-1} w_x(i)w_e(i)}\ . \tag{29}$$

For the windows under consideration and $M = 256$ the cross-correlation function $r_{w_x w_e}(p)$ is shown in Figure 3.

Since the extent of $r_{xx}(p)$ is much smaller than the extent of $r_{w_x w_e}(p)$ we may approximate

$$\big(r_{xx}(p) * g(p+\lambda R)\big) \cdot r_{w_x w_e}(p) \approx \approx r_{xx}(p) * \big(g(p+\lambda R) \cdot r_{w_x w_e}(p)\big)\ . \tag{30}$$

Strict equality in (30) will hold for a white noise excitation $r_{xx}(p) = \delta(p)\sigma_{xx}^2$. With the above approximation we have

$$\Phi_{XE}^{(\lambda)}(\Omega_\ell) = \big(\Phi_{XX}(\Omega_\ell)G(\Omega_\ell)e^{j\Omega_\ell \lambda R}\big) * R_{w_x w_e}(\Omega_\ell) \approx \approx \Phi_{XX}(\Omega_\ell)\left(G(\Omega_\ell)e^{j\Omega_\ell \lambda R} * R_{w_x w_e}(\Omega_\ell)\right) \tag{31}$$

where $G(\Omega_\ell)$ is the frequency response of the residual echo system and $R_{w_x w_e}(\Omega_\ell)$ is the cross-power spectrum of the windows.

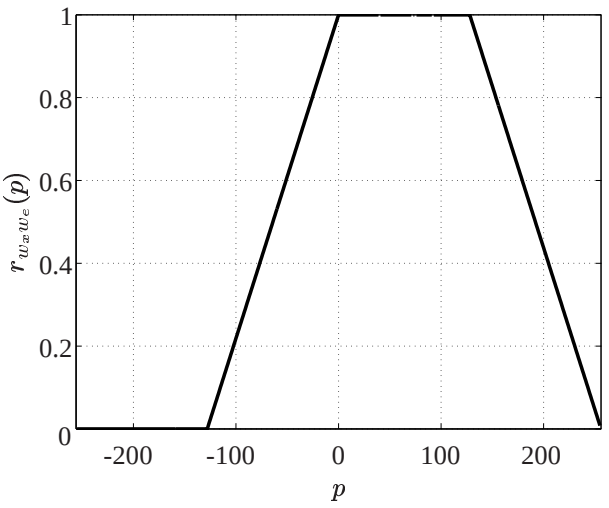

Figure 3: Window cross-correlation function $r_{w_x w_e}(p)$. $M = 256$.

Moreover, we find that the normalization of the cross power spectral density by $\Phi_{XX}(\Omega_\ell)$ removes the dependence of $\Phi_{XE}^{(\lambda)}(\Omega_\ell)$ on the input signal statistics. The additional gradient constraint operation $DFT\{Q\{IDFT\{\cdot\}\}\}$ of the FDAF yields the spectrum

$$G^{(\lambda)}(\Omega_\ell) = DFT\{Q\{IDFT\left\{\frac{\Phi_{XE}^{(\lambda)}(\Omega_\ell)}{\Phi_{XX}(\Omega_\ell)}\right\}\}\} \quad (32)$$

of a rectangular partition $g(p+\lambda R)$, $p = 0, \ldots, M/2-1$, $\lambda \in \mathbb{Z}$, of the residual echo system $g(i)$. This can be seen from

$$IDFT\left\{\frac{\Phi_{XE}^{(\lambda)}(\Omega_\ell)}{\Phi_{XX}(\Omega_\ell)}\right\} = g(p+\lambda R) r_{w_x w_e}(p) + \\ + g(p+\lambda R - M) r_{w_x w_e}(p-M), \quad p = 0 \ldots M-1\,. \quad (33)$$

in conjunction with the projection Q which zero–forces the samples for $p = M/2, \ldots, M-1$ and therefore leaves only the signal within the flat–top region of $r_{w_x w_e}(p)$ for further processing.

Replacing power spectral densities by their short–time estimates, the above procedure is ideally suited to compute the residual echo power estimate $\Phi_{BB}^{(\lambda)}(\Omega_\ell, kR)$ of partition λ at frame index k:

$$G^{(\lambda)}(\Omega_\ell, kR) = DFT\{Q\{IDFT\left\{\frac{\Phi_{XE}^{(\lambda)}(\Omega_\ell, kR)}{\Phi_{XX}^{(\lambda)}(\Omega_\ell, kR)}\right\}\}\} \quad (34)$$

$$C_{X^{(\lambda)}E}(\Omega_\ell, kR) = \frac{|G^{(\lambda)}(\Omega_\ell, kR)|^2 \Phi_{XX}^{(\lambda)}(\Omega_\ell, kR)}{\Phi_{EE}(\Omega_\ell, kR)} \quad (35)$$

The total residual echo PSD is eventually obtained from Equations (24) and (25), once more assuming white noise excitation.

Equation (33), however, also clarifies differences between residual echo PSD estimation and the adaptation part (9) of the FDAF used for echo cancellation. Firstly, the step-size $\mu^{(\lambda)}(\Omega_\ell, kR)$ differs from the simple normalization $\Phi_{XX}(\Omega_\ell, kR)$ in (33). Secondly, averaging over time takes place before the application of the gradient constraint in (33). Therefore, the constraint operation can not be shared by the echo canceler and the residual echo estimator, due to the frequency–dependent step–size control. To keep complexity low, though, we may omit the constraint for the residual echo estimation. In fact, the coherence estimator in Equation (20) can be viewed as an (approximate) unconstrained version of (34) and (35) with considerably lower complexity. Unconstrained residual echo estimation will be discussed more detailed in the Appendix of this paper.

Another approximate partitioning of the residual echo system, which in practice has proven to be sufficiently accurate, uses Hann windows $w_e(i) = w_x(i) = 0.5(1 - \cos(2\pi i/M))$, $0 \le i \le M-1$, with 50% frame overlap and unconstrained coherence estimation.

4.2 STATIONARY LOCAL BACKGROUND NOISE

Local background noise is also decisively influencing the quality of the residual echo estimator. In general, the residual echo estimates will be too high due to short–term correlations between long–term independent echo and background noise signals. The statistical expectation of the biased coherence $\hat{C}$ (20), which holds for estimates on the basis of Welch's power spectral averaging technique, is given in [14] for stationary signals:

$$E\{\hat{C}\} \approx C + \frac{1}{N}(1-C)^2\left(1 + \frac{2C}{N}\right) \triangleq f_C(C, N) \quad (36)$$

Thereby, C denotes the true coherence and N is the number of periodograms used for averaging over time. N is related to the equivalent forgetting factor α of recursive averaging (used in our algorithm) by

$$N = \frac{1+\alpha}{1-\alpha}\,. \quad (37)$$

4.2.1 Correction of the Coherence Bias

The expectation of the biased coherence estimate $\hat{C}$ is given as a function of the true coherence C in Equation (36). The proposed mechanism for bias correction directly relies on the inversion of the above formula. That implies a correction which is dependent on the background noise (and near speech) power, since the true coherence is defined as $C = \Phi_{BB}/\Phi_{EE}$.

Before, however, we have to decrease the variance of the coherence estimate in order to make the bias correction reliable. Therefore we average over several adjacent frequency components of the (cross) PSDs involved in the estimation of the coherence function (20). Frequency averaging of (cross) PSDs is written as

$$\Phi_{XE}^{(cb)}(\Omega_\ell, kR) = \frac{1}{K+1} \sum_{i=-K/2}^{K/2} \Phi_{XE}(\Omega_i, kR) \quad (38)$$

and the associated coherence estimate, Equation (20) or (35), with decreased error variance and decreased frequency resolution is denoted by $C_{XE}^{(cb)}(\Omega_\ell, kR)$.

Psychoacoustically motivated [15], we typically make a non–uniform choice for the number of frequency components to be averaged, thus avoiding noticeable performance degradations. In that respect we consider the critical bandwidth $cb(\Omega_\ell)$ at the center frequency Ω_ℓ [15]

$$cb(\Omega_\ell) = 25 + 75 \left(1 + 1.4 \left(\frac{f_a \Omega_\ell}{2\pi \text{kHz}}\right)^2\right)^{(0.69)} \text{ Hz} \quad (39)$$

where f_a is the sampling frequency. The number of DFT bins used for averaging is then determined by

$$K = \text{integer} \left(\frac{cb \cdot M}{f_a}\right) . \quad (40)$$

We proceed with the actual bias correction by the inversion of Equation (36) with the help of (37). In the following, this will be denoted in short by

$$C \approx f_C^{-1}(E\{\hat{C}\}, N(\alpha)) \quad (41)$$

which is applicable to correct the stationary bias of $\hat{C}$.

Equation (41) may be implemented by means of a look–up table or solved iteratively by

$$C^{(i+1)} = E\{\hat{C}\} - \frac{1}{N}(1 - C^{(i)})^2 \left(1 + \frac{2C^{(i)}}{N}\right) \quad (42)$$

where $C^{(0)} = E\{\hat{C}\}$. Normally, one or two iterations deliver a solution $C \approx C^{(i+1)}$ with sufficient accuracy.

Using the bias correction (41), we can now rewrite the coherence based residual echo estimator (24) as

$$\Phi_{BB}^{(\lambda,cb)}(\Omega_\ell, kR) = \\ = f_C^{-1}\left(C_{X^{(\lambda)}E}^{(cb)}(\Omega_\ell, kR), N(\alpha)\right) \Phi_{EE}(\Omega_\ell, kR) . \quad (43)$$

Consequently, we obtain the unbiased multiple–frame residual echo estimate analogously to Equation (25):

$$\Phi_{BB,new}^{(cb)}(\Omega_\ell, kR) = \sum_{\lambda=0}^{L-1} \Phi_{BB}^{(\lambda,cb)}(\Omega_\ell, kR) \quad (44)$$

The algorithm delivers an unbiased residual echo estimate even in the presence of stationary local disturbances. Note that, strictly speaking, also the residual echo due to the misalignment of filter partition $W^{(\lambda')}(\Omega_\ell, kR), \lambda' \neq \lambda$, represents a local disturbance for the estimation of $\Phi_{BB}^{(\lambda,cb)}(\Omega_\ell, kR)$. This is conceptually taken into account by the approach now.

4.3 NON-STATIONARY LOCAL DISTURBANCES AND NON-STATIONARY EXCITATION

In the case of local or far speech activity we face a severe problem with Welch's (cross) power spectral estimation technique and with recursive averaging of periodograms (21) as well.

Assume, for example, that the near speech power suddenly rises. Then the update term in Equation (21) becomes dominant and consequently the estimated coherence (20) approximately equals unity regardless of the true coherence value. This bias of the associated residual echo estimate severely impacts the control of the frequency domain adaptive echo canceler since the adaptation rate attains large values in contradiction to the actually desired behavior in the presence of local speech. Also the postfilter can not work as intended in that case.

Furthermore, consider the case of non-stationary excitation of the adaptive filter. In the case that the excitation suddenly vanishes the feedback term in Equation (21) becomes dominant and the coherence estimate (20) slowly decays with the time constant α. The associated residual echo estimate again will be too high and, therefore, drives the frequency domain adaptive filter into a state of divergence (even in the presence of only weak background noise).

4.3.1 Detection of Non–Stationarities

Our approach relies on Equation (13) for the step-size of the FDAF algorithm. In order to cope with non–stationarities we use two adaptive filters with different (hypothetical) step-sizes [16] for each partition λ of the FDAF. We compare the two adaptive filters by their resulting error signal which shall be minimized by the use of the correct step-size, which in turn yields the correct convergence state (residual echo).

In particular, an improved residual echo PSD estimate is now expressed by means of the convergence state (12) of the echo canceler

$$\Phi_{BB}^{(\lambda,cb,Sh)}(\Omega_\ell, kR) = \left|G^{(\lambda)}(\Omega_\ell, kR)\right|^2 \Phi_{XX}^{(\lambda)}(\Omega_\ell, kR) . \quad (45)$$

The estimate of the residual echo power–transfer function $\left|G^{(\lambda)}(\Omega_\ell, kR)\right|^2$ is obtained either by means of Equation

(43) as

$$\left|G^{(\lambda)}(\Omega_\ell, kR)\right|^2 = \left|G_1^{(\lambda)}(\Omega_\ell, kR)\right|^2 = \\ = \frac{f_C^{-1}\left(C_{X^{(\lambda)}E}^{(cb)}(\Omega_\ell, kR), N(\alpha)\right)\Phi_{EE}(\Omega_\ell, kR)}{\Phi_{XX}^{(\lambda)}(\Omega_\ell, kR)} \quad (46)$$

in the case that the underlying coherence estimate is useful or from the previous frame as

$$\left|G^{(\lambda)}(\Omega_\ell, kR)\right|^2 = \left|G_2^{(\lambda)}(\Omega_\ell, kR)\right|^2 = \\ = \left|G^{(\lambda)}(\Omega_\ell, (k-1)R)\right|^2 \quad (47)$$

otherwise.

In order to perform the decision on either

$$|G_1^{(\lambda)}(\Omega_\ell, kR)|^2 \quad \text{or} \quad |G_2^{(\lambda)}(\Omega_\ell, kR)|^2 \quad (48)$$

we run the two corresponding hypothetical FDAFs $W_1^{(\lambda)}(\Omega_\ell, kR)$ and $W_2^{(\lambda)}(\Omega_\ell, kR)$ with error signals $E_1(\Omega_\ell, kR)$ and $E_2(\Omega_\ell, kR)$ and hypothetical step-sizes

$$\mu_{1/2}^{(\lambda)}(\Omega_\ell, kR) = \frac{\left|G_{1/2}^{(\lambda)}(\Omega_\ell, kR)\right|^2}{\Phi_{EE,1/2}(\Omega_\ell, kR)} \quad (49)$$

according to Equations (6), (9), and (13). Assuming statistical independence of echo and local disturbances, we make the decision for that convergence state $|G_{1/2}^{(\lambda)}(\Omega_\ell, kR)|$ which minimizes the power (spectral density) of the corresponding error signal $E_{1/2}(\Omega_\ell, kR)$. Typically, we perform a global decision over the whole range of frequency bins and few iterations in time to gain in robustness. When comparing the error powers, we further apply a (heuristic) safety factor in order to avoid false detections due to the correlations within the local speech signal.

Assuming the correct decision for the convergence state $|G^{(\lambda)}(\Omega_\ell, kR)|$, we can consequently perform acoustic echo cancellation on the basis of Equations (6), (9),(13), and (45) with high reliability.

During double talk, the reliability of the coherence estimate is worst. At the same time there is only a minor improvement of the convergence state of the echo canceler possible. Thus, the above algorithm will choose $|G^{(\lambda)}(\Omega_\ell, kR)|^2 = |G_2^{(\lambda)}(\Omega_\ell, kR)|^2 = |G^{(\lambda)}(\Omega_\ell, (k-1)R)|^2$. During far end single talk, the estimate of the convergence state of the adaptive filter can be considerably improved selecting the most recent $|G^{(\lambda)}(\Omega_\ell, kR)|^2 = |G_1^{(\lambda)}(\Omega_\ell, kR)|^2$. This leads to accurate results of the residual echo estimator (45), too.

Eventually, the multiple–frame based residual echo estimator (25,44) is rewritten by means of (45) as

$$\Phi_{BB,new}^{(cb,Sh)}(\Omega_\ell, kR) = \sum_{\lambda=0}^{L-1} \Phi_{BB}^{(\lambda,cb,Sh)}(\Omega_\ell, kR) \quad (50)$$

in order to take the effect of finite DFT block lengths into account once again.

4.4 Additional Benefits

The above algorithm (50) represents our final estimate of the residual echo PSD which accounts for all kinds of estimation problems as outlined in this paper. The proposed structure for residual echo estimation entails a number of additional benefits which are briefly discussed here:

- Each coherence estimate $C_{X^{(\lambda)}E}(\Omega_\ell, kR)$ considers an individual PSD $\Phi_{XX}^{(\lambda)}(\Omega_\ell, kR)$ of the excitation signal. Thus, we make only weak assumptions with respect to the stationarity of the excitation. This is particularly meaningful for speech excitation in the presence of long reverberation times.

- The bias of each coherence estimate $C_{X^{(\lambda)}E}(\Omega_\ell, kR)$ is removed individually by the bias correction formula (41).

- Eventually, we observe the freedom to assign individual forgetting factors $\alpha^{(\lambda)}$ to the estimation processes of the coherence functions $C_{X^{(\lambda)}E}(\Omega_\ell, kR)$. This is useful, since a reasonable forgetting factor certainly depends on the individual ratio of acoustic echo and local disturbances in partition λ.

4.5 Computational Complexity

The newly proposed multiple–frame algorithm basically runs the single-frame coherence estimator (19), (20), and (21) L times in parallel in order to deliver an unbiased residual echo estimate. However, the costly spectral analysis of the input signals has to be performed only once, regardless of the parameter L. Additionally, we apply the bias correction (41) L times in parallel to cope with local disturbances. The bias correction can be implemented by means of a one–dimensional look–up table in a very simple way. The two path filter strategy (providing the robustness) can be assumed to increase the complexity by at most a factor of two. In practice, it turns out that only $L = 3$ or $L = 4$ can considerably improve the residual echo estimate in the case of car acoustics.

The computational complexity of our approach mainly depends on the number of divisions associated with coherence estimation (20). The required number of operations is significantly reduced by averaging and sub–sampling DFT bins prior to the coherence computation. Given fixed complexity constraints, we strongly recommend to design an unbiased multiple–frame residual echo estimator, if necessary at the cost of a lower frequency resolution.

Eventually, we recall the symbiosis of the acoustic echo canceler, the postfilter, and the residual echo estimator in our algorithm. As explained before, that results in a shared complexity with respect to the analysis and synthesis operations (DFT/IDFT) required in the echo and noise control system.

5 SIMULATION RESULTS

We will first show that partitioned residual echo PSD estimation with individual bias correction for each partition delivers unbiased estimates for the total residual echo PSD. Then we will demonstrate the robustness of the algorithm in the framework of our echo and noise control system, see Figure 2.

5.1 MEASUREMENT OF THE TOTAL RESIDUAL ECHO PSD

5.1.1 Log–Spectral–Mean

For the purpose of instrumental evaluation of residual echo PSD estimation techniques, we make use of a frame-oriented spectral distance measure. At frame index k, we consider the *Log–Spectral–Mean*

$$LSM(kR) = \frac{1}{M}\sum_{l=0}^{M-1} 10\log_{10}\frac{\hat{\Phi}_{BB}(\Omega_\ell, kR)}{\Phi_{BB}(\Omega_\ell, kR)} \qquad (51)$$

of the estimated–to–true residual echo power ratio at frame index k. This is a frame–based bias measure for residual echo estimators, which is ideally zero.

5.1.2 Numerical Results

We compare numerical results for several estimation techniques under consideration. In particular, these are the estimators in Equation (19) for a single DFT frame, Equation (25) for multiple DFT frames, and Equation (44) for multiple frames with individual bias correction.

We use a stationary white noise excitation $X(\Omega, k)$, various levels of local speech $S(\Omega, k)$, and car background noise $N(\Omega, k)$. The acoustic echo is generated artificially by means of a fixed car impulse response of 512 coefficients, the first 128 coefficients being canceled nearly ideally by a fixed echo compensator with 128 taps. The DFT length of 256 for residual echo estimation is made up of 128 data points for each frame plus additional zero-padding. With regard to the short term stationarity of speech, we choose the forgetting factor $\alpha = 0.8$ for the single-frame estimator. The number of partitions for the multiple–frame residual echo estimator is $L = 4$, the corresponding forgetting factors were individually chosen as $\alpha^{(0)} = 0.8$, $\alpha^{(1)} = 0.8$, $\alpha^{(2)} = 0.9$, and $\alpha^{(3)} = 0.9$.

Figure 4 depicts the results for the exact (constrained) partitioned coherence estimator as given in Equations (34, 35). We consider three different acoustic environments: In the first 300 signal frames, there is no local speech nor background noise contributing to the microphone signal, thus, acoustic echo only. In frames 300 to 600 we added local car background noise to achieve an echo–to–noise ratio of -6 dB. Eventually, in frames 600 to 900, there is local speech present (double talk) at a speech–to–noise ratio of 0 dB and car background noise at the same level as before.

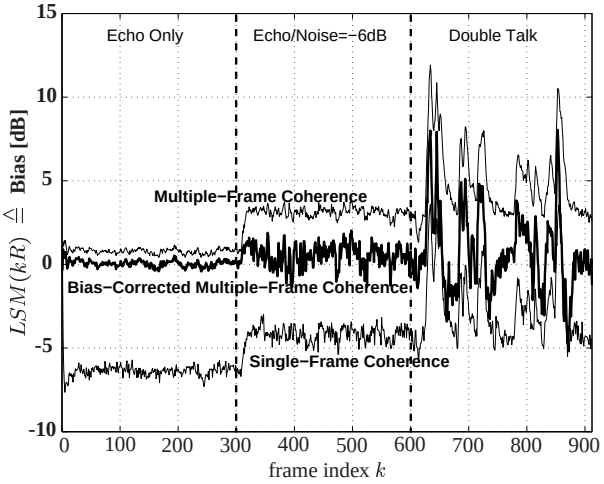

Figure 4: Bias for constrained partitioned residual echo estimators.

From Figure 4 we observe that the single-frame coherence estimator (19) does not completely reflect the residual echo. The bias of the estimator is most severe when there are no local disturbances. In the presence of local background noise and speech activity, the approach achieves better performance only because of the additional bias introduced by short–term correlations in this case. We further observe that the multiple–frame coherence estimator (25) achieves nearly unbiased residual echo estimates when there is neither speech nor background noise present. This is due to the full coverage of the LEM impulse response of length 512 by $L = 4$ estimator partitions. We can, however, see the bias of this method in the presence of local disturbances. This is circumvented by the additional coherence bias correction (41) applied in the multiple–frame estimator of (44). The latter delivers unbiased estimates with regard to various acoustic environments. Note that the variance of the estimator still depends on the local echo–to–noise/speech ratio. However, in the presence of background noise the estimate is not required to be as accurate as in noise-free environments. Hence, we conclude that the multiple–frame coherence estimator delivers consistently excellent results for the application of residual echo postfiltering.

Figure 5 refers to the same types of estimators and to the same acoustic environments as before. In contrast, this experiment shows that we can achieve nearly the same estimation performance using the approximate (unconstrained) partitioned coherence estimator, Equation (57), with much lower complexity than the exact implementation.

5.2 PERFORMANCE OF THE ACOUSTIC ECHO CONTROL SYSTEM

The performance of our algorithm in the combined echo and noise reduction system is first investigated for a single–talk situation (acoustic echo plus local background

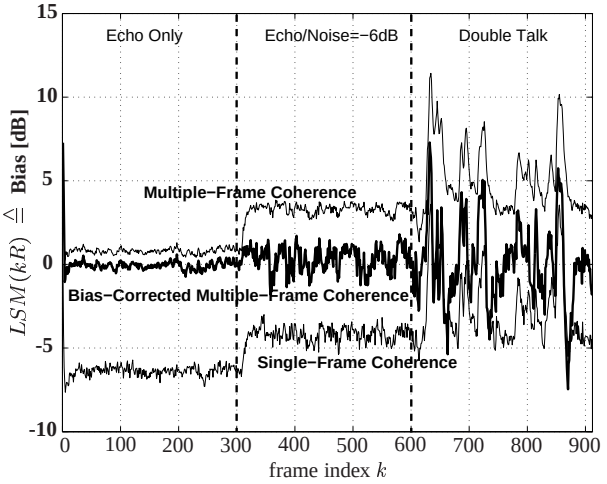

Figure 5: Bias for unconstrained partitioned residual echo estimators.

noise) at an echo–to–noise ratio of 20 dB. The underlying experiment is illustrated by Figure 6. The quality of our algorithm is expressed by means of the echo return loss enhancement (ERLE) which measures the attenuation in the feedback loop of the telephone. In particular, we consider

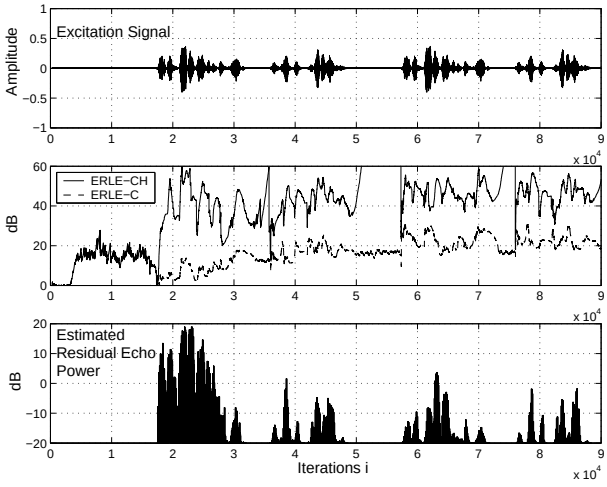

Figure 6: ERLE and estimated residual echo power for a noisy single talk (far end talk) situation.

the ERLE-C which is achieved by the echo canceler only and the ERLE-CH which is obtained from echo cancellation and additional postfiltering. Figure 6 shows that the postfilter reacts much faster with regard to the presence of acoustic feedback than the echo canceler. This is due to the short (and limited) period of averaging applied in the residual echo estimator. In the initial phase of the simulation, the postfilter achieves the ERLE-CH mostly on its own. While the echo canceler converges, the postfilter attains less additional echo attenuation, such that the total ERLE-CH is constant. That results from the decreasing estimate of the residual echo PSD which is shown in the bottom of Figure 6.

The robustness of the algorithm is analysed for the example of a difficult double talk situation. The excitation signal and the near end speech plus background noise are shown in the top two graphs of Figure 7. In the double talk

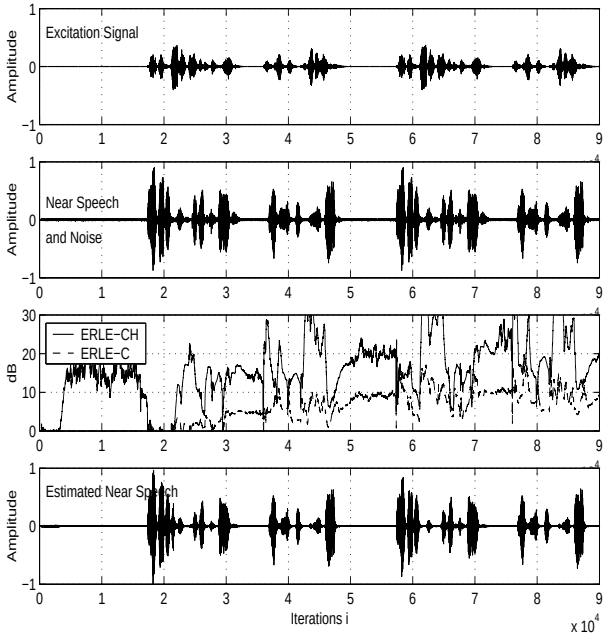

Figure 7: ERLE and estimated near speech for a noisy double talk environment.

situation under consideration, the ERLE-C and ERLE-CH is clearly not as high as in the single talk situation. But we observe once more that the postfilter attains fast and strong echo attenuation while the echo canceler requires more time to reach a good state of convergence. Moreover, in the bottom of Figure 7 we can see that the fast tracking capability of the postfilter also guarantees excellent preservation of the local speech signal. The results as shown here once more indicate the accuracy and the robustness of the residual echo estimators as presented in this paper.

6 CONCLUSIONS

In this paper we have proposed a combined acoustic echo and noise control system which is based on a partitioned FDAF with partitioned residual echo estimation for postfiltering. It was shown that an accurate residual echo PSD estimate is essential for both the control of the FDAF and the postfilter.

An accurate unbiased residual echo PSD estimator was derived by taking the full length of the residual echo impulse response as well as the bias due to stationary and non–stationary disturbances into account. The resulting partitioned bias–compensated residual echo estimator is

then used to control both the FDAF for acoustic echo cancellation and the postfilter for residual echo and background noise suppression.

Furthermore, it was shown that the FDAF, the postfilter, and the partitioned residual echo PSD estimator can be efficiently integrated into a combined system where the computational complexity of analysis and synthesis is shared between different components of the algorithm.

APPENDIX

APPROXIMATE PARTITIONED RESIDUAL ECHO PSD ESTIMATION

We aim at reducing the complexity of the exact (constrained) partitioned residual echo PSD estimator in Equations (34,35). An approximate residual echo power estimate might be obtained without applying a constraint. In this case the residual echo power estimate is biased. In what follows we derive an approximate expression for this bias.

Using (28) and (31) the magnitude squared cross PSD $|\Phi_{XE}^{(\lambda)}(\Omega_\ell)|^2$ can be normalized and computed as

$$\begin{aligned}\Phi^{*(\lambda)}_{XE}(\Omega_\ell)\Phi^{(\lambda)}_{XE}(\Omega_\ell)\Big/|\Phi_{XX}(\Omega_\ell)|^2 = \\ = \sum_{p=-(M-1)}^{M-1}\sum_{v=-(M-1)}^{M-1} g(v+\lambda R) r_{w_x w_e}(v)\cdot \\ g(p+\lambda R) r_{w_x w_e}(p) e^{-j\Omega_\ell (v-p)} \end{aligned} \quad (52)$$

and with $v = p + u$

$$\begin{aligned}|\Phi^{(\lambda)}_{XE}(\Omega_\ell)|^2\Big/|\Phi_{XX}(\Omega_\ell)|^2 = \\ = \sum_{u=-(M-1)-p}^{M-1-p}\sum_{p=-(M-1)}^{M-1} g(p+u+\lambda R) r_{w_x w_e}(p+u)\cdot \\ \cdot g(p+\lambda R) r_{w_x w_e}(p) e^{-j\Omega_\ell u} \\ = \sum_{u=-2(M-1)}^{2(M-1)}\sum_{p=-\infty}^{\infty} g(p) r_{w_x w_e}(p-\lambda R)\cdot \\ \cdot g(p+u) r_{w_x w_e}(p+u-\lambda R) e^{-j\Omega_\ell u}\,. \end{aligned} \quad (53)$$

The second equality results because $r_{w_x w_e}(p)$ is zero for $|p| \geq M$. The above equation can be interpreted as the power spectrum of a segment of the residual echo impulse response. This segment is cut out by the cross-correlation function $r_{w_x w_e}(p)$ of the DFT windows as it was shown in Figure 3. We denote this segment by $g^{(\lambda)}(p) = g(p) r_{w_x w_e}(p - \lambda R)$ and its power spectrum by $|G^{(\lambda)}(\Omega)|^2$. To achieve an unbiased residual echo power estimate we must require that the sum over all magnitude squared segments equals the power spectrum of the full residual echo impulse response $|G(\Omega)|^2$

$$\begin{aligned}\sum_{\lambda=0}^{L-1}|G^{(\lambda)}(\Omega=\Omega_\ell)|^2 = |G(\Omega=\Omega_\ell)|^2 = \\ \sum_{u=-\infty}^{\infty} r_{gg}(u) e^{-j\Omega_\ell u} = \sum_{u=-\infty}^{\infty}\sum_{p=-\infty}^{\infty} g(p) g(p+u) e^{-j\Omega_\ell u}\end{aligned} \quad (54)$$

Comparing the inner sum in (53) and (54) we find that the above condition can be fulfilled if $r_{gg}(u)$ decays to zero within $2M$ samples and if

$$\begin{aligned}r_{gg}(u) = \sum_{p=-\infty}^{\infty} g(p) g(p+u)\cdot \\ \sum_{\lambda=0}^{L-1} r_{w_x w_e}(p-\lambda R) r_{w_x w_e}(p+u-\lambda R)\end{aligned} \quad (55)$$

holds for all $u \in [-2(M-1), 2(M-1)]$. Hence, for a bias free estimate

$$\begin{aligned}R_w(p,u) = \sum_{\lambda=0}^{L-1} r_{w_x w_e}(p-\lambda R) r_{w_x w_e}(p+u-\lambda R) \stackrel{!}{=} 1 \\ \forall p, \forall u \in [-2(M-1), 2(M-1)]\end{aligned} \quad (56)$$

needs to be fulfilled. For the windows $w_x(i)$ and $w_e(i)$ used in the FDAF algorithm this is clearly not the case.

Fig. 8 plots $R_w(p, 0)$ for $M = 256$, $L = 4$, and the windows used by the FDAF algorithm. The combined support of all partitions extents over approximately 512 samples of the residual echo system impulse response. It is evident that the estimate obtained by this method is too large by a factor of approximately $F_w = 1.5$ to 1.7. Therefore, an approximately unbiased echo estimate can be constructed by dividing $|\Phi_{XE}^{(\lambda)}(\Omega_\ell)|^2$ by this factor and using the modified short–term coherence estimate

$$C_{X^{(\lambda)}E}(\Omega_\ell, kR) = \frac{|\Phi^{(\lambda)}_{XE}(\Omega_\ell, kR)|^2}{\Phi^{(\lambda)}_{XX}(\Omega_\ell, kR)\,\Phi_{EE}(\Omega_\ell, kR) F_w} \quad (57)$$

in (43) to compute the residual echo PSD estimate.

ACKNOWLEDGMENT

This work is supported by the Nokia Research Center (NRC), Tampere, Finland, and Nokia Mobile Phones (NMP), Bochum, Germany.

Manuscript received on . . .

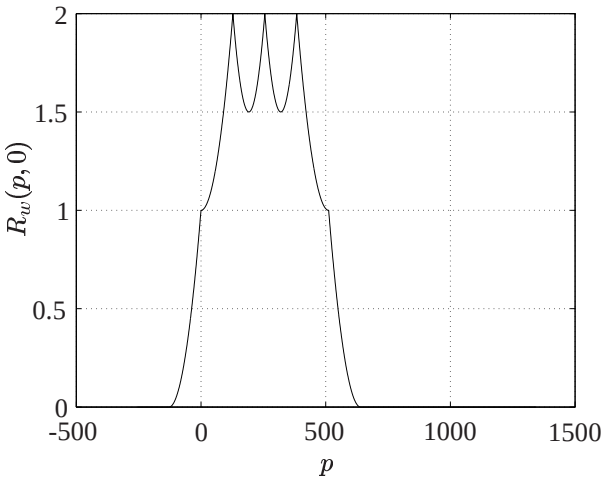

Figure 8: Region of support for $M = 256$, $L = 4$, $u = 0$ and the rectangular windows $w_x(i)$ and $w_e(i)$ of the FDAF algorithm. Ideally, this should be a flat top window of amplitude one.

REFERENCES

[1] S. Gustafsson, R. Martin, and P. Vary. Combined acoustic echo control and noise reduction for hands-free telephony. *Signal Processing*, 64:21–32, 1998.

[2] S. Gustafsson. *Enhancement of Audio Signals by Combined Acoustic Echo Cancellation and Noise Reduction*. PhD thesis, Aachen University of Technology, 1999.

[3] E. Hänsler, G. U. Schmitt. Hands-free telephones - joint control of echo cancellation and postfiltering. *Signal Processing*, 80:2295–2305, 2000.

[4] G. U. Schmidt. *Entwurf und Realisierung eines Multiratensystems zum Freisprechen*. PhD thesis, Fortschritt–Berichte VDI Verlag, Reihe 10, Nr. 674, Düsseldorf, Germany, 2001.

[5] B. H. Nitsch. A frequency-selective stepfactor control for an adaptive filter algorithm working in the frequency domain. *Signal Processing*, 80:1733–1745, 2000.

[6] C. Beaugeant. *Réduction de Bruit et Contrôle d'Echo pour les Applications Radiomobiles*. PhD thesis, University of Rennes 1, 1999.

[7] J. Benesty and D.R. Morgan. Multi–channel frequency–domain adaptive filtering. *In: S.L. Gay, J. Benesty (Eds.), Acoustic Signal Processing for Telecommunication, Kluwer Academic Publishers*, pages 121–133, 2000.

[8] S. Haykin. Adaptive Filter Theory. Prentice Hall, 1996.

[9] P. C. W. Sommen. *Adaptive Filtering Methods*. PhD thesis, TU Eindhoven, 1992.

[10] E. R. Ferrara. Frequency–domain adaptive filtering. *In: C.F.N. Cowan, P.M. Grant (Eds.), Adaptive Filters, Prentice–Hall, Englewood Cliffs, N.J.*, pages 145–179, 1985.

[11] R. Martin. Spectral Subtraction Based on Minimum Statistics. Proc. EUSIPCO-94, Edinburgh, pp. 1182-1185, September 12-16, 1994.

[12] R. Martin. Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics. *IEEE Trans. Speech and Audio Processing*, 9(5), July 2001.

[13] Y. Ephraim and D. Malah. Speech Enhancement Using a Minimum Mean-Square Error Log-Spectral Amplitude Estimator. *IEEE Trans. Acoustics, Speech and Signal Processing*, 33(2):443–445, April 1985.

[14] G. C. Carter. Coherence and time delay estimation. *Proceedings of the IEEE*, 75:236–255, 1987.

[15] E. Zwicker and H.Fastl. Psychoacoustics, Facts and Models. Springer–Verlag, New York, 1990.

[16] A. Mader, H. Puder, G. Schmidt. Step-size control for acoustic echo cancellation filters – an overview. *Signal Processing*, 80(9):1697–1719, September 2000.