SIGNAL-BASED LATE RESIDUAL ECHO SPECTRAL VARIANCE ESTIMATION

María Luis Valero, Edwin Mabande and Emanuël A. P. Habets

International Audio Laboratories Erlangen*
Am Wolfsmantel 33, 91058 Erlangen, Germany
{maria.luis.valero,emanuel.habets}@audiolabs-erlangen.de

ABSTRACT

A commonly used technique to attenuate acoustic echo signals in hands-free devices is acoustic echo cancellation (AEC). In practice, the AEC is unable to provide sufficient echo reduction due to system misalignment, modeling errors and the insufficient length of the estimated acoustic echo path. Hence, AEC is usually used in conjunction with one or more postfilters to suppress the remaining echo. For the estimation of the postfilter the residual echo is required. In the past, several approaches have been proposed which estimate the late residual echo spectral variance using the far-end signal and parameters that are obtained from the estimated echo path. In this work, we propose a signal-based algorithm to estimate the parameters. Thus, among other advantages, the postfilter can also be used in combination with acoustic echo suppression.

***Index Terms*—** Late Residual Echo Estimation, Acoustic Echo Cancellation, Late Reverberation Estimation

1. INTRODUCTION

In typical hands-free communication scenarios, the desired near-end speech acquired by the microphone is often distorted by the acoustic echo of the far-end speech and background noise. As a result, the quality of speech and its intelligibility are degraded, thus communication is hindered. The most commonly used technique to reduce the echo is acoustic echo cancellation (AEC), [1, 2]. In order to do this, the acoustic echo path is estimated using adaptive filtering algorithms, see for example [3, 4]. Nonetheless, AEC is often unable to provide enough attenuation and a residual echo remains. The residual echo is caused by 1) the misalignment between the true and the estimated echo path, 2) the insufficient length of the estimated echo path (i.e., the under-modeling of the echo path), and 3) nonlinear signal components. Hence, one or more postfilters are necessary to suppress the remaining echo signal.

In this work we focus on the late residual echo (LRE) caused by the insufficient length of the estimated echo path. Several models have been proposed to address the problem of LRE suppression, see for example [5], [6] and [7]. For the postfilter design all of them employ an estimate of the LRE spectral variance. For instance, in [5] and [6] a recursive estimator is proposed that makes use of the far-end echo signal. Both approaches derive the parameters for the estimator based on the estimated acoustic echo path; we therefore refer to these as channel-based approaches. In contrast, the authors in [7] design a filter to suppress the acoustic echo. In this case, the model parameters are estimated directly from the signals; we refer to this as a signal-based approach.

This work is based on the signal model presented by Habets et al. in [5]. Our goal is to derive the parameters for the estimator from the available signals in the system instead of from the estimated echo path. To this end, the signal model is reformulated using the fact that the late residual echo is the late reverberation of the acoustic echo.

2. PROBLEM FORMULATION

Given a loudspeaker-enclosure-microphone environment the signal captured by the microphone, $y(n)$, is described by

$$y(n) = d(n) + s(n) + v(n) = h(n) * x(n) + s(n) + v(n), \quad (1)$$

where the discrete time index is given by n, $h(n)$ is the acoustic echo path, $x(n)$ is the far-end speech, $s(n)$ is the near-end speech and $v(n)$ is the background noise. The goal of the AEC is to obtain an estimate of the acoustic echo signal $d(n)$ which is denoted by $\hat{d}(n)$. Yet, due to complexity and convergence constraints, only N_e coefficients of the acoustic echo path can be estimated. The output signal of the AEC is obtained by subtracting $\hat{d}(n)$ from $y(n)$. This signal is referred to as the error signal and can be expressed as

$$e(n) = y(n) - \hat{d}(n) = e_\mathrm{m}(n) + e_\mathrm{r}(n) + s(n) + v(n), \quad (2)$$

where $e_\mathrm{m}(n)$ denotes the residual echo due to the misalignment of the estimated echo path and $e_\mathrm{r}(n)$ denotes the LRE signal, which is a consequence of the under-modeling of the acoustic echo path. Using the short time Fourier transform (STFT), we express the error signal in the time-frequency domain as

$$E(l,k) = E_\mathrm{m}(l,k) + E_\mathrm{r}(l,k) + S(l,k) + V(l,k), \quad (3)$$

where $E(l,k)$ denotes the STFT of $e(n)$ and l and k are the time and frequency indexes, respectively. Throughout the paper we assume that the AEC has converged, such that the term $E_\mathrm{m}(l,k) \ll E_\mathrm{r}(l,k)$ can be neglected. In practice $E_\mathrm{m}(l,k)$ can be significantly reduced by means of a residual echo suppressor such as the one proposed in [8]. We also assume that no near-end speech is present. The LRE in the time-frequency domain can be defined as in [5],

$$E_\mathrm{r}(l,k) = \sum_{i=0}^{\infty} H(i + N_{R\mathrm{e}}, k) X(l - i - N_{R\mathrm{e}}, k), \quad (4)$$

where $N_{R\mathrm{e}} = N_\mathrm{e}/R$ and R is the frame shift for the analysis window of N_w samples. Taking the same assumptions into account as in [5], the following approximation for the LRE spectral variance holds

$$\lambda_{e_\mathrm{r}}(l,k) \equiv E\{|E_\mathrm{r}(l,k)|^2\} \quad (5)$$
$$\approx \sum_{i=0}^{\infty} E\left\{|H(i + N_{R\mathrm{e}}, k)|^2\right\} E\left\{|X(l - i - N_{R\mathrm{e}}, k)|^2\right\},$$

*A joint institution of the Friedrich-Alexander-University Erlangen-Nürnberg (FAU) and Fraunhofer IIS, Germany.

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

where $E\{\cdot\}$ is the mathematical expectation operator. The envelope of the acoustic echo path, $E\{|H(l,k)|^2\}$, based on the generalized reverberation model in [9], can be expressed as

$$\mathrm{E}\{|H(l,k)|^2\} \approx \tilde{\lambda}_h(l,k) = \begin{cases} c_\mathrm{d}(k), & \text{for } l = 0; \\ c_\mathrm{r}(k)\alpha^l(k), & \text{for } l \geq 1. \end{cases} \quad (6)$$

This simplified model does not take the early reflections into account. Thus, the echo path's envelope is characterized only by the direct power, $c_\mathrm{d}(k)$, the initial reverberant power, $c_\mathrm{r}(k)$, and the exponential decay rate, $\alpha(k) \in [0,1)$. The decay rate is related to the reverberation time, $T_{60}(k)$, by $\alpha(k) = e^{-2\rho(k)R}$, where

$$\rho(k) \equiv \frac{3\ln(10)}{T_{60}(k) f_s}, \quad (7)$$

and f_s is the sampling frequency. In the following, the notation $\tilde{\cdot}$ highlights that (6) was used to estimate the spectral variances. Consequently, the estimated LRE spectral variance can be expressed as

$$\tilde{\lambda}_{e_\mathrm{r}}(l,k) = \sum_{i=0}^{\infty} c_\mathrm{r}(k)\alpha^{i+N_{Re}}(k)\lambda_x(l-i-N_{Re},k), \quad (8)$$

where $\lambda_x(l,k) = E\{|X(l,k)|^2\}$ is the far-end speech spectral variance. Finally, analogous to [5], (8) can be expressed recursively, i.e.

$$\tilde{\lambda}_{e_\mathrm{r}}(l,k) = \alpha(k)\tilde{\lambda}_{e_\mathrm{r}}(l-1,k) + c_\mathrm{r}(k)\alpha^{N_{Re}}(k)\lambda_x(l-N_{Re},k). \quad (9)$$

As in [5–7], we can now use a spectral enhancement technique to reduce the LRE and thereby enhance the near-end speech. To this end, the estimated LRE spectral variance is used for the estimation of the postfilter gains. To accomplish this, given (9), we need an estimate of the parameters $c_\mathrm{r}(k)$ and $\alpha(k)$. The aim of this work is to derive an algorithm to estimate these parameters based on the far-end signal, the microphone signal and the error signal.

3. CHANNEL-BASED PARAMETER ESTIMATION

In this section we recapitulate the estimators for the model parameters given in [5]. This approach is referred to as channel-based as the model parameters are obtained from the estimated echo path, $\hat{h}(n)$.

3.1. Estimation of the reverberation time

In [5] the reverberation time, $T_{60}(k)$, is estimated using the energy decay curve (EDC) [10] of band-pass filtered versions of $\hat{h}(n)$ [11]. For brevity, we only define the fullband form, i.e.

$$\mathrm{EDC}(m) = 10\log_{10}\left\{\frac{\sum_{j=m}^{N_\mathrm{e}-1}\hat{h}^2(j)}{\sum_{j=0}^{N_\mathrm{e}-1}\hat{h}^2(j)}\right\}; 0 \leq m \leq N_\mathrm{e}-1. \quad (10)$$

In order to estimate the reverberation time a straight line is fitted between two points of the EDC. As in [5], the range between -5 and -25 dB was used. Between these points, the EDC can be modeled as $\mathrm{EDC}(m) = p - q\cdot m$, where p is the offset, and the regression factor q is related to the estimated reverberation time by $\hat{T}_{60} = \frac{60}{q\cdot f_s}$.

3.2. Estimation of the initial power

As in [5], the power of the acoustic echo path's envelope, $c(l,k)$, can be defined as

$$c(l,k) = \left|\sum_{j=0}^{N_\mathrm{w}-1} w(j)\hat{h}(j+lR)e^{-i\frac{2\pi k}{N_\mathrm{DFT}}j}\right|^2, \quad (11)$$

where $i = \sqrt{-1}$, N_DFT denotes the length of the discrete Fourier transform and $w(j)$ is the STFT analysis window. In terms of the model in (6), we can define $\hat{c}_\mathrm{d}(k) = c(0,k)$ and $\hat{c}_\mathrm{r}(k) = c(1,k)$. Note that in [5] $\hat{c}(N_\mathrm{e},k) = \alpha^{N_\mathrm{w}/R}(k)c(N_{Re}-N_\mathrm{w}/R,k)$ was used. Finally, $\hat{c}_\mathrm{r}(k)$ is smoothed over the frequency axis in order to avoid spectral zeros.

4. SIGNAL-BASED PARAMETER ESTIMATION

In this section, the proposed signal-based approach for the estimation of the model parameters in (8) is described. In contrast to the channel-based approach, our proposal only utilizes the available signals in the system and not $\hat{h}(n)$. The optimal model parameters are obtained by minimizing the cost-function

$$J(\alpha(k), c_\mathrm{r}(k)) = \sum_{l=0}^{N_\mathrm{T}-1}\left(\lambda_e(l,k) - \tilde{\lambda}_{e_\mathrm{r}}(l,k)\right)^2, \quad (12)$$

where N_T is the length of the loudspeaker signal in frames and $\lambda_e(l,k) = E\{|E(l,k)|^2\}$ is the spectral variance of the error signal. The parameter estimators are derived using two different models for the LRE spectral variance. While the decay rate is obtained using a similar model as for the late reverberation estimation in [9], the initial power is obtained using (8).

4.1. Estimation of the decay rate

Given the signal model in [9], a relationship between the model for the late reverberation and the LRE estimation, (8), can be derived. In contrast to [9], it should be noted that in this particular case the non-reverberated far-end signal $x(n)$ is available. After suppressing the noise, using for example the proposed method in [12] or in [13], and in absence of near-end speech, the spectral variance of $Y(l,k)$ can be expressed, using the models in (6) and (8), as

$$\begin{aligned}\tilde{\lambda}_y(l,k) &= \sum_{i=0}^{\infty}\tilde{\lambda}_h(i,k)\lambda_x(l-i,k) && (13a)\\ &= c_\mathrm{d}(k)\lambda_x(l,k) + \sum_{i=1}^{\infty}c_\mathrm{r}(k)\alpha^i(k)\lambda_x(l-i,k)\\ &= c_\Delta(k)\lambda_x(l,k) + \sum_{i=0}^{\infty}c_\mathrm{r}(k)\alpha^i(k)\lambda_x(l-i,k)\\ &= c_\Delta(k)\lambda_x(l,k) + \alpha^{-N_{Re}}\tilde{\lambda}_{e_\mathrm{r}}(l+N_{Re},k), && (13b)\end{aligned}$$

where $c_\Delta(k) = c_\mathrm{d}(k) - c_\mathrm{r}(k)$ is the direct-to-reverberant level difference. Since the microphone-loudspeaker distance is commonly small, we can assume that $c_\Delta(k) \geq 0$. To estimate $c_\Delta(k)$ we propose to minimize

$$J(c_\Delta(k)) = \sum_{l=N_{Re}}^{N_\mathrm{T}+N_{Re}-1}\left(\lambda_y(l,k) - \tilde{\lambda}_y(l,k)\right)^2, \quad (14)$$

where $\lambda_y(l,k)$ is the true spectral variance of $Y(l,k)$ and $\tilde{\lambda}_y(l,k)$ is computed using (13b). To do this, we approximate $\tilde{\lambda}_{e_\mathrm{r}}(l,k)$ by $\lambda_e(l,k)$, which is valid when the aforementioned assumptions are satisfied. Using (13b) and assuming that $\tilde{\lambda}_y(l,k) \approx \lambda_y(l,k)$, we can now express the estimated LRE spectral variance as

$$\begin{aligned}\tilde{\lambda}_{e_\mathrm{r}}(l,k) &= \alpha^{N_{Re}}(k)[\lambda_y(l-N_{Re},k) - c_\Delta(k)\lambda_x(l-N_{Re},k)]\\ &= \alpha^{N_{Re}}(k)\tilde{\lambda}'_y(l-N_{Re},k), && (15)\end{aligned}$$

where $\tilde{\lambda}'_y(l,k)$ is the spectral variance of the microphone signal after subtracting the estimated excessed direct energy. Finally, the decay rate α is obtained by substituting (15) into the following logarithmic cost function

$$J(\ln(\alpha(k))) = \sum_{l=0}^{N_{\mathrm{T}}-1} \left(\ln(\lambda_e(l,k)) - \ln(\tilde{\lambda}_{e_{\mathrm{r}}}(l,k))\right)^2, \quad (16)$$

and minimizing (16) by setting the partial derivative with respect to $\ln(\alpha(k))$ to zero. Consequently,

$$\ln(\hat{\alpha}(k)) = \frac{\sum_{l=0}^{N_{\mathrm{T}}-1}(\ln(\lambda_e(l,k)) - \ln(\tilde{\lambda}'_y(l-N_{R\mathrm{e}},k)))}{N_{R\mathrm{e}} \cdot N_{\mathrm{T}}}, \quad (17)$$

To simplify, the decay rate is assumed to be frequency invariant, i.e. $\hat{\alpha} = E\{\hat{\alpha}(k)\}$. The main advantage of estimating the decay rate instead of the reverberation time is the fact that the possible values of α are limited by and tend towards 1. This is due to the exponential relation between both parameters. For instance, at a sampling frequency $f_s = 16$ kHz, the decay rate will be larger than 0.78 for reverberation times larger than 0.25 s.

4.2. Estimation of the initial reverberant power

The estimator of the initial reverberant power is obtained by substituting (8) into (12), and setting the partial derivative with respect to $c_{\mathrm{r}}(k)$ to zero. Hence,

$$\hat{c}_{\mathrm{r}}(k) = \frac{\sum_{l=0}^{N_{\mathrm{T}}-1} \lambda_e(l,k) \sum_{i=0}^{N_{Rl}-1} \hat{\alpha}^i \lambda_x(l-i-N_{R\mathrm{e}},k)}{\hat{\alpha}^{N_{R\mathrm{e}}} \sum_{l=0}^{N_{\mathrm{T}}-1}\left(\sum_{i=0}^{N_{Rl}-1} \hat{\alpha}^i \lambda_x(l-i-N_{R\mathrm{e}},k)\right)^2}, \quad (18)$$

where $N_{Rl} \gg N_{R\mathrm{e}}$ limits the length of the under-modeled acoustic echo path, and the estimated decay rate, $\hat{\alpha}$, was calculated before. As in Sec. 3.2, it is recommended to smooth $\hat{c}_{\mathrm{r}}(k)$ over frequency.

5. ONLINE PARAMETER ESTIMATION

In the previous section, an algorithm for the parameter estimation was presented which uses the complete signal's length, i.e., N_{T} frames. For real-time applications this is not feasable. Hence,

$$J(\alpha(l,k), c_{\mathrm{r}}(l,k)) = \sum_{l'=l-N_{\mathrm{t}}+1}^{l} \left(\lambda_e(l',k) - \tilde{\lambda}_{e_{\mathrm{r}}}(l',k)\right)^2 \quad (19)$$

is used, where N_{t} is the number of frames used to update the estimations. In addition, this algorithm has to be able to track changes in the environment in a controlled way. Note that the parameters must not be updated if near-end speech is present, as if this occurs (17) does not hold. The decision whether to update can be made based on the output of a double-talk detector as the one proposed in [14]. Yet, a false negative will often lead to outliers. In order to discard them, an order statistics filter as in [15] is applied. Order statistics filters, [16, 17], are defined by

$$a_{\mathrm{os}} = \arg\left\{P(x) = \gamma_a : P(x) = \int_0^x p(a)da\right\}, \quad (20)$$

where, for example, $\gamma_a = 0.5$ is the definition of the median filter. The task remains, however, to select the appropriate γ_a.

It is widely accepted that the decay rate can only be correctly estimated in periods of free decay, as it tends to be overestimated otherwise, [15, 18]. However, it is not advisable to track the minimum as this could lead to its underestimation. Hence, $\gamma_\alpha = 0.1$ is a proper value, which is also the value proposed in [15].

In order to control the update of the initial reverberant power, $c_r(k)$, and of the direct-to-reverberant level difference, $c_\Delta(k)$, a first-order recursive filter with a forgetting factor η is applied followed by an order statistics filter. Empirically, we have found that $\eta = 0.65$, $\gamma_c(k) = 0.45$ and $\gamma_{c_\Delta} = 0.5$ give good results.

6. PERFORMANCE EVALUATION

In this section we evaluate the performance of the proposed signal-based parameter estimators. First, the proposed approach is compared against the channel-based approach proposed in [5]. Secondly, an analysis of the convergence of the decay rate estimator and its influence on the LRE spectral variance estimation is performed. Finally, the performance of the online parameter estimation algorithm under adverse conditions will also be analyzed. For the accuracy assessment of the estimated LRE spectral variance, the log-error distortion as described in [19] is used. Using this measure allows us to obtain not only the over- and underestimation of the obtained LRE spectral variance, but also the distribution of these errors. The overall log-error distortion is defined by

$$\mathrm{LogErr_{ov}} = \frac{10}{KL}\sum_{l=0}^{L-1}\sum_{k=0}^{K-1}\left|\min\left(0, \log_{10}\frac{\hat{\lambda}_{e_{\mathrm{r}}}(l,k)}{\tilde{\lambda}_{e_{\mathrm{r}}}(l,k)}\right)\right|; \quad (21)$$

$$\mathrm{LogErr_{un}} = \frac{10}{KL}\sum_{l=0}^{L-1}\sum_{k=0}^{K-1}\max\left(0, \log_{10}\frac{\hat{\lambda}_{e_{\mathrm{r}}}(l,k)}{\tilde{\lambda}_{e_{\mathrm{r}}}(l,k)}\right); \quad (22)$$

$$\mathrm{LogErr} = \mathrm{LogErr_{ov}} + \mathrm{LogErr_{un}}. \quad (23)$$

where $\hat{\lambda}_{e_{\mathrm{r}}}(l,k)$ is an estimate of the true LRE spectral variance $E\{|E_{\mathrm{r}}(l,k)|^2\}$ that is obtained using

$$\hat{\lambda}_{e_{\mathrm{r}}}(l,k) = \eta_{\mathrm{er}}\hat{\lambda}_{e_{\mathrm{r}}}(l-1,k) + (1-\eta_{\mathrm{er}})|E_{\mathrm{r}}(l,k)|^2, \quad (24)$$

for which the forgetting factor $\eta_{\mathrm{er}} = \exp(\frac{-R}{f_s \cdot 0.012})$ is known to give good results, [5].

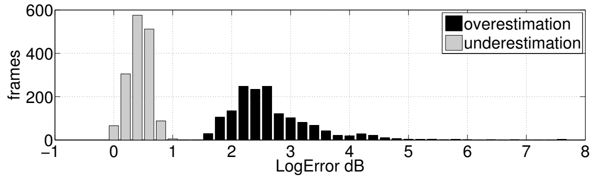
All simulations were conducted at a sampling frequency $f_s = 16$ kHz. The room impulse responses (RIR), of length $N = 4096$ taps, were generated for a 5m×4m×3m (length×width×height) room using the image method, [20]. The distance between the loudspeaker and the microphone was set to 1.4 m. During the simulations, three RIRs with reverberation times 0.25 s, 035 s and 0.45 s were used. An echo canceler of length $N_{\mathrm{e}} = 2048$ taps canceled the early echo perfectly and no background noise was present. The time domain signals were transformed to the STFT domain using a 256 points Hamming window. The overlap between successive STFT frames was set to 75% resulting in a frame shift of $R = 64$ samples. The order statistics filters were applied on 100 estimates.

6.1. Results

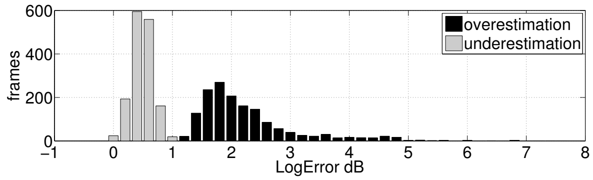
First, the log-error distortion caused by the channel-based and the signal-based approach is compared. For this, a RIR with a reverberation time $T_{60} = 0.35$ s was used. The channel-based model parameters were estimated from the first N_{e} taps of the actual RIR. The signal-based parameters were estimated as described in Sec. 4. Table 1 shows a comparison of the overall log-error distortion.

	$\mathrm{LogError_{ov}}$	$\mathrm{LogError_{un}}$	LogError
channel-based	2.6856 dB	0.4337 dB	3.1193 dB
signal-based	2.2510 dB	0.4894 dB	2.7415 dB

Table 1: Overall log-error distortion analysis



(a) channel-based approach

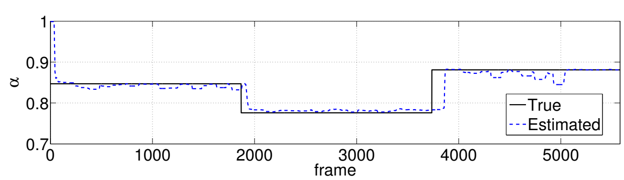


(b) signal-based approach

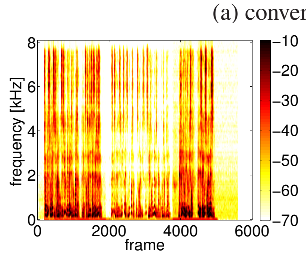
Fig. 1: Log-error distortion distribution analysis

Table 1 shows that the proposed algorithm causes less log-error distortion. Moreover, the overestimation is considerably reduced while the underestimation is only moderately increased. This is illustrated in more detail in Figs. 1a and 1b, which show that the overestimation caused by the signal-based approach is distributed towards lower values as for the channel-based approach. Hence, using the signal-based estimated LRE spectral variance for the postfilter design will cause less near-end speech distortion, as the postfilter will be less aggressive.

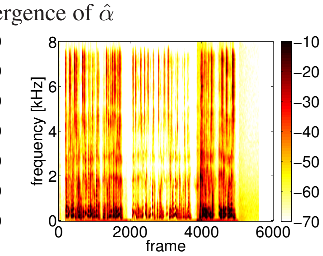
Secondly, the convergence and the ability to track changes in the environment of the proposed online algorithm were tested. To this end, the reverberation time was first set to 0.35 s, then modified to 0.25 s and finally set to 0.45 s and the parameters were updated every frame. Thus, both the ability to track the decrease and the increase of the decay rate were tested. Fig. 2a depicts the estimated decay rate compared to the true decay rate of the acoustic echo path over time. Figs. 2b and 2c compare the true LRE spectral variance and its estimation.



(a) convergence of $\hat{\alpha}$



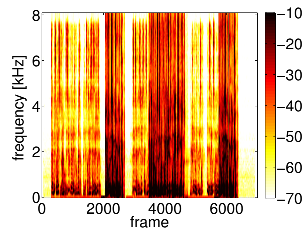
(b) $\hat{\lambda}_{e_r}(l,k)$



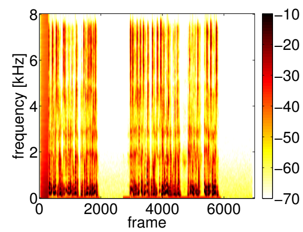
(c) $\tilde{\lambda}_{e_r}(l,k)$

Fig. 2: Convergence of the online parameter estimation algorithm

These figures show that the algorithm converges rapidly to the right solution. In addition, it is also capable of tracking changes in the environment. Nonetheless, due to the order statistics filter, it is more efficient in tracking the decrease of the decay rate as its increase. The total log-error distortion is 3.1241 dB, from which 1.9105 dB correspond to overestimation and 1.2135 dB correspond to underestimation.



(a) $\hat{\lambda}_e(l,k)$



(b) $\tilde{\lambda}_{e_r}(l,k)$

Fig. 3: LRE estimation in presence of near-end speech

Finally, the robustness of the proposed algorithm in presence of near-end speech and in double-talk situations was tested. Hence, the error signal after the AEC contains both the near-end speech and the LRE. The online parameter estimation algorithm along with a double-talk detector, similar to the one proposed in [14], was used. The RIR had a reverberation time $T_{60} = 0.35$ s. The spectral variance of the error signal is depicted in Fig. 3a and the estimated LRE spectral variance is depicted in Fig. 3b. For brevity we do not include a figure depicting the true LRE. It can be observed that the proposed online estimators need some time to converge. Yet, after they have converged, the estimated LRE spectral variance contains nearly no traces of the near-end speech. To prove the accuracy of the proposed method, Table 2 summarizes the overall log-error distortion introduced by both the channel-based and the signal-based approaches.

	$\text{LogError}_{\text{ov}}$	$\text{LogError}_{\text{un}}$	LogError
channel-based	2.5635 dB	0.4712 dB	3.0348 dB
signal-based	1.0073 dB	1.5220 dB	2.5292 dB

Table 2: Overall log-error distortion analysis

It can be concluded that the signal-based approach tends, even in adverse conditions, to introduce less overestimation, which is of advantage as the postfilter will cause less distortion to the near-end speech. Moreover, the difference in underestimation might be inaudible because the remaining LRE is likely to be masked by the near-end speech.

7. CONCLUSIONS

A signal-based parameter estimation approach for the LRE spectral variance estimator proposed in [5] has been presented. The proposed approach is able to function without having prior knowledge of the estimated acoustic echo path. In addition, an online update and control algorithm for the estimators was provided. The online algorithm allows for fast convergence to the right solution and it is also robust when using a real double-talk detector. The performance evaluation shows that the resulting LRE spectral variance presents a lower log-error distortion compared to the channel-based approach. The estimated LRE spectral variance can be used in the context of LRE suppression for the estimation of the postfilter gains.The postfiler can be used in conjunction with AEC and/or acoustic echo suppression. It must be noted that the log-error distortion due to overestimation is considerably reduced, which will lead to less near-end speech distortion caused by the LRE suppressor. Further research will include the analysis of the proposed approach in combination with adaptive AEC and in presence of background noise.

8. REFERENCES

[1] E. Hänsler and G. Schmidt, *Acoustic Echo and Noise Control: A pracical Approach*, Wiley, New Jersey, USA, 2004.

[2] G. Schmidt, "Applications of acoustic echo control – an overview," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Vienna, Austria, 2004, pp. 9–16.

[3] J. J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Process. Mag.*, vol. 9, no. 1, pp. 14–37, Jan. 1992.

[4] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, fourth edition, 2002.

[5] E. A. P. Habets, I. Cohen, S. Gannot, and P. C. W. Sommen, "Joint dereverberation and residual echo suppression of speech signals in noisy environments," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 8, pp. 1433–1451, Nov. 2008.

[6] G. Enzner, *A Model-Based Optimum Filtering Approach to Acoustic Echo Control: Theory and Practice*, Ph.D. thesis, RWTH Aachen University, Wissenschaftsverlag Mainz, Aachen, Germany, Apr. 2006, ISBN 3–86130-648–4.

[7] C. Faller A. Favrot and F. Küch, "Modeling late reverberation in acoustic echo suppression," *International Workshop on Acoustic Signal Enhancement*, Sept. 2012.

[8] G. Enzner, R. Martin, and P. Vary, "Unbiased residual echo power estimation for hands free telephony," in *Proc. IEEE ICASSP*, Orlando, Florida, USA, May 2002, IEEE, pp. 1893–1896.

[9] E. A. P. Habets, S. Gannot, and I. Cohen, "Late reverberant spectral variance estimation based on a statistical model," *IEEE Signal Process. Lett.*, vol. 16, no. 9, pp. 770–774, Sept. 2009.

[10] M. R. Schroeder, "New method of measuring reverberation time," *J. Acoust. Soc. Am.*, vol. 37, pp. 409–412, 1965.

[11] J.-M. Jot, "An analysis/synthesis approach to real-time artificial reverberation," in *Proc. IEEE ICASSP*, Mar. 1992, vol. 2, pp. 221–224.

[12] R. Martin, "Spectral subtraction based on minimum statistics," in *Proc. European Signal Processing Conf*, 1994, pp. 1182–1185.

[13] I. Cohen, "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 5, pp. 466–475, Sept. 2003.

[14] T. Gänsler, M. Hansson, C.-J. Ivarsson, and G. Salomonsson, "A double-talk detector based on coherence," *IEEE Trans. Commun.*, vol. 44, no. 11, pp. 1421–1427, May 1996.

[15] R. Ratnam, D. L. Jones, B. C. Wheeler, W. D. O'Brien, Jr., C. R. Lansing, and A. S. Feng, "Blind estimation of reverberation time," *J. Acoust. Soc. Am.*, vol. 114, no. 5, pp. 2877–2892, Nov. 2003.

[16] H. A. David, *Order Statistics*, Wiley, New York, USA, 1981.

[17] I. Pitas and A. N. Venetsanopoulos, "Order statistics in digital image processing," in *Proceedings of the IEEE*, Dec. 1992, vol. 80, pp. 1893–1921.

[18] H. W. Löllmann and P. Vary, "Estimation of the reverberation time in noisy environments," in *Proc. Intl. Workshop Acoust. Echo Noise Control (IWAENC)*, Sept. 2008, pp. 1–4.

[19] T. Gerkmann and R.C. Hendriks, "Unbiased MMSE-based noise power estimation with low complexity and low tracking delay," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 4, pp. 1383 –1393, May 2012.

[20] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.