A NEW CROSS-DOMAIN APPROACH TO SYNCHRONIZED ADAPTIVE ECHO CANCELLATION AND ECHO POSTFILTERING

Christelle Yemdji[1], *Moctar Mossi Idrissa*[1], *Nicholas Evans*[1] *and Christophe Beaugeant*[2]

[1]EURECOM, Multimedia Department, Sophia-Antipolis, France
[2]Intel, Mobile Comunications Group, Sophia-Antipolis, France
{yemdji, mossi, evans}@eurecom.fr christophe.beaugeant@intel.com

ABSTRACT

Most acoustic echo cancellation systems are based upon an adaptive filter followed by echo postfiltering. Although the two modules have complementary functionalities, they generally operate independently. Existing approaches to synchronized echo control are somewhat limited by the need for both modules to function in the same subband or frequency domain. This paper presents our first attempt at synchronized time-domain adaptive filtering and frequency domain echo postfiltering. The new synchronization approach exploits the relationship between the system distance in the time domain and the system mismatch power spectrum. Together with a new system distance measurement, the proposed synchronous echo control system is shown to perform well in comparison to existing state of the art approaches.

Index Terms— adaptive echo cancellation, echo postfiltering, synchronized echo control, system mismatch.

1. INTRODUCTION

With the increasing need for mobility, recent years have seen the rapid evolution of telecommunications services such as mobile telephony and video-conferencing. A consistent challenge with many such services is the provision for improved speech quality.

With hands-free devices, for example, speech quality is often degraded by acoustic echo which arises from the coupling between the loudspeaker and the microphone of the communications terminal [1]. Approaches to acoustic echo control are generally based upon adaptive filtering followed by echo postfiltering [2]. Adaptive echo cancellation (AEC) is based on the assumption that the acoustic path can be modeled as a linear finite impulse response filter. The estimated acoustic path is then used to generate an estimate of the echo so that it may be subtracted from the uplink signal [3]. In practice, due to the mismatch between the acoustic path and its estimate, echo postfiltering is commonly used to further suppress residual echo [1].

Historically, adaptive echo filtering and echo postfiltering are implemented as independent modules. More recently, however, echo control systems with synchronized adaptive echo cancellation and echo postfiltering have been investigated and have shown improved performance [4, 5]. Synchronized echo control systems exploit the link between the two modules which are in this case designed to operate in the same frequency or subband domain.

In this paper we present a new approach to synchronized time domain adaptive echo filtering and subband domain echo postfiltering. The new approach exploits the relationship between the system distance in the time domain and the system mismatch power spectrum. Furthermore, we propose a modified estimate of the system distance, inspired from [6], and show how it can be used to improve global echo control performance. The approach to synchronization investigated in this paper is based upon a fullband adaptive filter but can nevertheless be extended to a subband implementation. Our approach is also advantageous for its robustness against nonlinearities [7].

The remainder of this paper is organized as follows. The proposed synchronized echo control architecture is presented in Section 2. Section 3 describes the approach used to synchronize adaptive echo filtering and echo postfiltering algorithms. An assessment of the new system is presented in Section 4 while conclusions are presented in Section 5.

2. SYSTEM OVERVIEW

Figure 1 shows an overview of the synchronized echo cancellation system proposed in this paper. The microphone signal $y(n)$ is the sum of the near-end signal $s(n)$ and the echo signal $d(n)$ which is obtained by the convolution of the loudspeaker signal $x(n)$ with the acoustic path $\boldsymbol{h}(n)$. An adaptive filter is used to generate an estimate of the echo signal $\hat{d}(n)$ which is subtracted from the microphone signal to obtain the error signal $e(n)$. The error signal is composed of residual echo $e_r(n)$ and, possibly, of near-end speech $s(n)$. The postfilter aims to suppress the residual echo. In addition to the conventional feedback used by the adaptive filter, an additional level of statistical control is applied to synchronize the adaptive filter and echo postfilter. The following details the investigated

© EURASIP, 2012 - ISSN 2076-1465

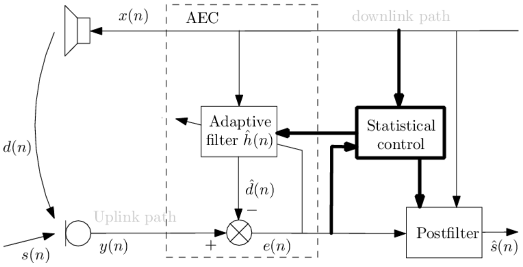

Fig. 1. System overview

adaptive filter and echo postfilter.

2.1. Adaptive echo cancellation

The adaptive filter is based upon a normalized least mean square (NLMS) algorithm where the acoustic path estimate $\hat{\boldsymbol{h}}(n)$ and its optimum stepsize $\mu(n)$ are expressed as follows [3]:

$$\hat{\boldsymbol{h}}(n+1) = \hat{\boldsymbol{h}}(n) + \frac{\mu(n)}{\mathbf{x}(n)^T \cdot \mathbf{x}(n)} \cdot \mathbf{x}(n) \cdot e(n) \quad \text{and} \quad (1)$$

$$\mu(n) = \frac{E\{e_r^2(n)\}}{E\{e^2(n)\}}, \quad (2)$$

where $\mathbf{x}(n) = \begin{bmatrix} x(n) & x(n-1) & \cdots & x(n-L+1) \end{bmatrix}^T$ is the loudspeaker signal, L is the length of the adaptive filter and $E\{.\}$ represents statistical expectation. The computation of the variable stepsize requires knowledge of the residual echo power $E\{e_r^2(n)\}$ which is not directly measurable. Instead, it is approximated as in [1, 3]:

$$E\{e_r^2(n)\} = E\{x^2(n)\} \cdot \|\boldsymbol{\beta}(n)\|^2 \quad (3)$$

where $\boldsymbol{\beta}(n)$ is the system mismatch i.e. the error between the real acoustic path $\boldsymbol{h}(n)$ and its estimate $\hat{\boldsymbol{h}}(n)$. The value $\|\boldsymbol{\beta}(n)\|^2$ is referred to as the system distance [3] and its computation is described in Section 3.

2.2. Echo postfiltering

The postfilter consists of frequency domain processing with filtering through linear convolution in the frequency domain domain [8, 9]. Prior to frequency gain computation, the postfilter input signals $x(n)$ and $e(n)$ are converted into frequency domain signals $x_i(m)$ and $e_i(m)$, respectively, where m is the frame index and i is the frequency bin index ranging from 0 to $M-1$. The i^{th} frequency signal $\hat{s}_i(m)$ is obtained through the multiplication of the gain $W_i(m)$ with $e_i(m)$. Conversion from time to frequency domain is performed on blocks of R samples through a fast Fourier transform with an overlap-add method [9].

For each frequency index i, the postfilter gains $W_i(n)$ are computed according to the Wiener rule [10]:

$$W_i(m) = \frac{\xi_i(m)}{1+\xi_i(m)}, \quad (4)$$

where $\xi_i(m)$ is the signal (near-end speech) to (residual) echo ratio (SER). In our implementation, the SER is estimated through the Ephraim and Malah approach as in [11, 10]. It requires an estimate of the residual echo power which we implement as:

$$\hat{\gamma}_i^{e_r e_r}(m) = |G_i(m)|^2 \cdot \gamma_i^{xx}(m), \quad (5)$$

where $\hat{\gamma}_i^{e_r e_r}(m)$ is the residual echo power spectral density, $\gamma_i^{xx}(m)$ is the loudspeaker power spectral density and $|G_i(m)|^2$ is the system mismatch power spectrum [5]. The computation of $|G_i(m)|^2$ is described in Section 3.

2.3. Markov model of the echo problem

Most approaches to AEC are based on a Wiener solution to the echo problem with the assumption that the echo path is stationary and deterministic. However, in a practical scenario, the acoustic path is time variant meaning it cannot be assumed to be stationary. These variations can sometimes be significant (e.g. : door opening or closing) or small. Small acoustic path changes can be modeled by a first-order Markov model [3, 6].

Such modeling of the echo path leads to Kalman filtering in the frequency domain as solution to the AEC problem [6]. Comparative assessments show that Kalman AEC converges faster than standard frequency domain adaptive filtering [12]. This solution can also be controlled synchronously with an echo postfilter [6]. In the next section, we show how similar synchronization can be achieved between the AEC and echo postfiltering algorithms presented above.

3. SYSTEM CONTROL

The architecture used here is inspired from existing synchronized approaches to echo control such as those in [4, 5, 6]. In such systems, the acoustic echo canceler (AEC) is constrained to function in the frequency or subband domains. However, a comparative study shows that subband or frequency domain AECs are less robust to non-linearities than fullband AECs [7]. The synchronization approach presented here operates with fullband AEC but can be adapted readily to operate with a subband AEC. Thus AEC and echo postfiltering are not constrained to operate in the same domain.

Our approach to synchronization is based upon the correspondence between the system mismatch $\boldsymbol{\beta}(n)$ and its spectrum $G_i(m)$ which are defined as follows

$$\boldsymbol{\beta}(n) = \boldsymbol{h}(n) - \hat{\boldsymbol{h}}(n) \text{ and } G_i(m) = H_i(m) - \hat{H}_i(m), \quad (6)$$

where $H_i(m)$ and $\hat{H}_i(m)$ are the Fourier transforms of $\boldsymbol{h}(n)$ and $\hat{\boldsymbol{h}}(n)$ respectively. From Equation 6, we note that $G_i(m)$ is the Fourier transform of $\boldsymbol{\beta}(n)$. In this case, according to

Parseval's equality, we can write the following:

$$\|\boldsymbol{\beta}(n)\|^2 = \frac{1}{M}\sum_{i=0}^{M-1} |G_i(m)|^2, \tag{7}$$

with the assumption that n is a multiple of the block size R. Equation 7 highlights the relationship between the NLMS algorithm and the echo postfilter. This relationship can be used in two different ways:

- The estimate $\boldsymbol{\beta}(n)$ can be used to compute both $\|\boldsymbol{\beta}(n)\|^2$ and $|G_i(m)|^2$. This solution is impractical because the misalignment vector $\boldsymbol{\beta}(n)$ cannot be estimated reliably. The estimation of $\boldsymbol{\beta}(n)$ requires correlation computation [3] which is highly computationally demanding. Most real time systems estimate the system distance directly [1].
- Alternatively, $|G_i(m)|^2$ can be estimated and used to derive $\|\boldsymbol{\beta}(n)\|^2$ according to Equation 7. As most echo postfilters already require the computation of $|G_i(m)|^2$, we opted for this solution. In this case there is no additional computational requirement.

Equation 7 is strictly valid in case $L \leq M$. But in most control systems $L > M$. In such cases, Equation 7 results in an underestimation of $\boldsymbol{\beta}(n)$ required by the NLMS algorithm. This underestimation can be justified by the fact that when $L > M$, $|G_i(m)|^2$ accounts for the system distance of an adaptive filter of shorter length (M taps) than the one actually considered by the AEC used (L taps). This underestimation remains very small as the error made is on the tail of the adaptive filter. Nevertheless similar relation between $\boldsymbol{\beta}(n)$ and $G_i(m)$ can be approximated through interpolation for example. So as to have a good understanding and assessment of the complete system (i.e. system without any approximation), we will only tackle the case where $L = M$ in this paper.

The system mismatch power spectrum $|G_i(m)|^2$ can be computed through the cross-correlation method [5] according to:

$$|G_i(m)|^2 = \left|\frac{\gamma_i^{xe}(m)}{\gamma_i^{xx}(m)}\right|^2, \tag{8}$$

where $\gamma_i^{xe}(m)$ is the cross spectral density between $e(n)$ and $x(n)$. However, the postfilter is updated on a frame-by-frame basis whereas the AEC requires a sample-by-sample update. In between two measurements of the system mismatch power spectrum, the system distance is updated according to the following recursion [13, 5]:

$$\|\boldsymbol{\beta}(n+1)\|^2 = \left(1 - \frac{\mu(n)}{L}\right) \cdot \|\boldsymbol{\beta}(n)\|^2. \tag{9}$$

A similar recursion can be found in [6] (Equation 64). Similarly to the solution in [6], we redefine the system distance by adding a second term $\|\Delta\boldsymbol{h}(n)\|^2$ as follows:

$$\|\boldsymbol{\beta}(n+1)\|^2 = \left(1 - \frac{\mu(n)}{L}\right) \cdot \|\boldsymbol{\beta}(n)\|^2 + \|\Delta\boldsymbol{h}(n)\|^2, \tag{10}$$

where $\|\Delta\boldsymbol{h}(n)\|^2$ accounts for changes in the acoustic path. Equivalently, $\|\Delta\boldsymbol{h}(n)\|^2$ is computed according to:

$$\|\Delta\boldsymbol{h}(n)\|^2 = (1 - A^2) \cdot \|\hat{\boldsymbol{h}}(n)\|^2, \tag{11}$$

where A is a constant set to a value lower than unity [6]. In [6], the authors carry some investigations to determine typical values of A. Therefore, in our experiments, we chose values of A to those used in [6].

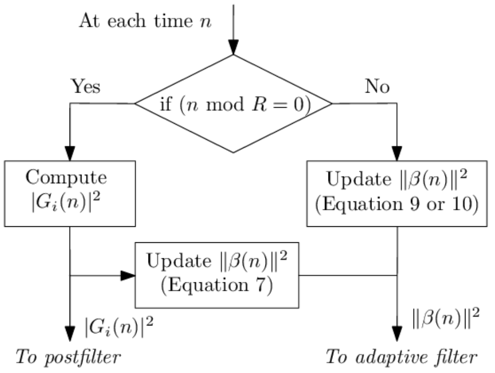

Fig. 2. Statistical control diagram

The synchronization approach that takes place within the statistical control module is summarized in Figure 2. It computes the system mismatch power spectrum according to Equation 8 on a frame-by-frame basis (i.e. when $n \bmod R = 0$). $|G_i(m)|^2$ is used within the postfilter to update the spectral gains $W_i(m)$ according to Equations 4 and 5 and within the adaptive filter for the computation of the system distance according to Equation 7. During intervals in which the postfilter is not updated, the system distance is updated according to Equations 9 or 10. We note that Equation 9 is equal to Equation 10 for $A = 1$. In the following, the implementation with Equation 9 is denoted *Sync. A = 1* while the implementation with Equation 10 is denoted *Sync. A = 0.99* since A is set to 0.99 in our experiments.

4. EXPERIMENTS

In this section we assess the synchronized echo control system proposed above and compare its performance to existing echo control systems. Section 4.1 presents the system setup while Section 4.2 presents our simulation results.

4.1. System setup

While this paper reports synchronized echo control, AEC performance and echo postfilter performance are nonetheless assessed separately. Both *Sync. A = 1* and *Sync. A = 0.99* AECs are considered. Both implementations are compared to a classical NLMS algorithm with a fixed stepsize ($\mu = 0.1$) as well as to the original Kalman AEC algorithm which inspired our work. In all cases, the adaptive filter has 256 taps.

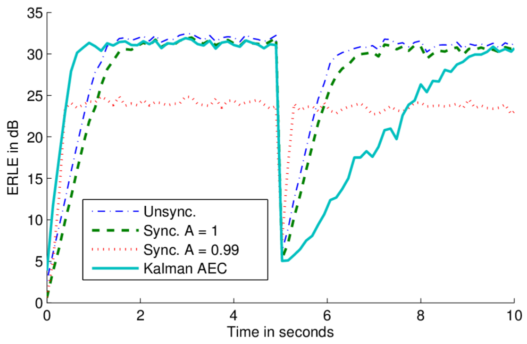

Fig. 3. ERLE against time for AEC only

The performance of the synchronized echo postfilter is compared to that of the Kalman echo control postfilter and to an unsynchronized echo postfilter. In our assessment of the postfilter, we once again consider the two proposed systems *Sync. A =1* and *Sync. A =0.99*. The unsynchronized echo postfilter consists of a postfilter placed after the NLMS AEC with fixed stepsize. In all cases, the number of frequency bins M is set to 256 while the framesize R is set to 128.

All simulations reported here were performed with either white noise or speech signals. All of which have a sampling frequency of 8kHz. White noise is used to assess the convergence while speech signals are used to simulate realistic echo cancellation conditions. Microphone speech signals contain an echo-only interval followed by a double-talk interval. The echo-only interval is long enough so that each AEC algorithm converges. The double-talk interval is used to assess the impact of near-end speech on both the AEC and postfilter. The echo signals are generated by convolving the loudspeaker signal with an acoustic path response. Four different acoustic path responses are used; they were all measured with real mobile terminals in an office environment. The resulting database of speech signals has SERs ranging from -5 dB to 10 dB with the near-end speech level set to -26 dB. Speech signal levels are set through the ITU-T speech voltmeter as in [10].

Performance is assessed in terms of echo return loss enhancement (ERLE) [1, 10], cepstral distance [8] and informal listening tests. While the ERLE is used to assess the amount of echo suppression during echo-only intervals, the cepstral distance is used to assess the amount of distortion introduced by postfiltering during double-talk intervals.

4.2. Assessment

The first set of experiments reported in Figure 3 aims to assess AEC convergence. Here, the microphone signal is composed of echo-only and corresponds to white noise. At time $t = 5s$, there is an abrupt echo path change. The curves show that *Sync. A = 0.99* converges more quickly than *Sync. A = 1*. Nevertheless, *Sync. A = 0.99* achieves less echo suppression. We also see that both proposed systems reconverge faster than the Kalman system after the abrupt acoustic path change. *Sync. A = 0.99* reconverges even faster than NLMS with fixed stepsize.

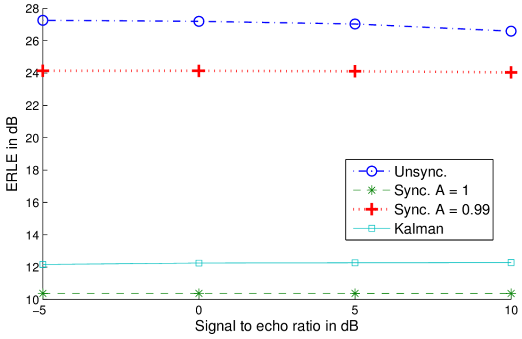

Fig. 4. Average ERLE against SER for AEC only during echo-only intervals

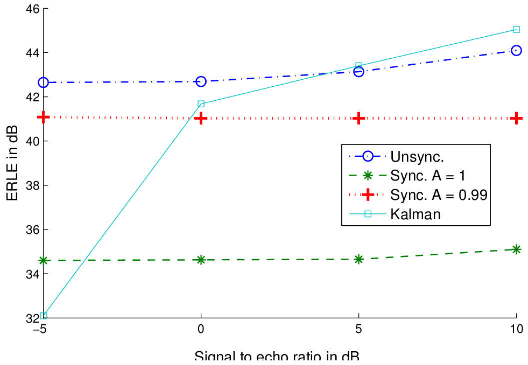

Fig. 5. Average ERLE achieved through the AEC and the echo postfilter

Figure 4 shows the mean ERLE against SER for the four different AEC implementations considered. The ERLE is measured during echo-only periods during which the AEC algorithm has converged. The unsynchronized AEC algorithm achieves the best performance in terms of ERLE. The proposed system distance (*Sync. A = 0.99*) gives better performance than the classical system distance (*Sync. A = 1*): more than 10 dB difference in ERLE across the full range of SERs. Moreover, the new system distance approach *Sync. A = 0.99* gives marginally better echo suppression than the Kalman AEC algorithm. This might be because the Kalman system is updated in the frequency domain on a frame-by-frame basis whereas the new approach is updated sample-by-sample.

Figure 5 illustrates the total amount of echo suppression achieved through combined AEC and postfiltering. The unsynchronized system and the Kalman system achieve the most echo suppression. The system with *Sync. A = 0.99* achieves slightly less echo suppression than the Kalman echo control system. This loss of performances can be attributed to the system mismatch power spectrum function estimate which is not the same in both postfilter. Once again, *Sync. A = 1* achieves worst performance in terms of ERLE: this is attributed to poor AEC performance.

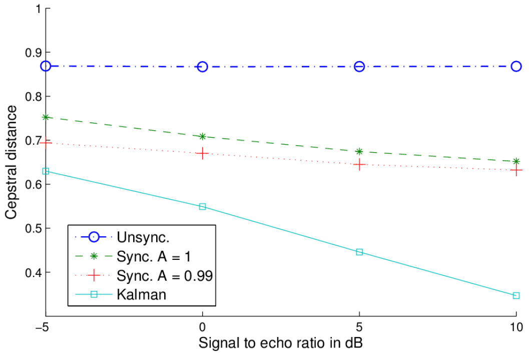

Fig. 6. Average cepstral distance against SER for the global echo control system during double-talk periods

Figure 6 shows the mean cepstral distance against SER for the four different systems. The cepstral distance is measured at the output of the postfilter during double-talk periods. We observe that the system with fixed stepsize brings the most distortion. The Kalman echo control system brings the least distortion. Although the new synchronized approaches introduce more distortion than the Kalman system, their levels remain low compared to that of the unsynchronized system. Nevertheless, the postfilter with *Sync. A =1* introduces slightly more distortion than the postfilter with *Sync. A = 0.99*. This comes from the fact that the AEC from *Sync. A = 1* achieves less echo suppression and thus places an increased demand on the postfilter than with AEC *Sync. A = 0.99*. Moreover, the complete echo control system *Sync. A = 1* achieves less echo suppression than the *Sync. A = 0.99* system (see Figure 5). The postfilter *Sync. A = 1* can be tuned in order to achieve as much echo suppression as *Sync. A = 0.99* but this results in increased distortion during double-talk intervals.

Informal listening tests reveal the presence of musical noise in signals at the output of the postfilter for both the proposed and the Kalman echo control systems. In addition to musical noise, signals processed by Kalman echo control sometimes contain crackling noise which was sometimes perceived as annoying. In signals processed by *Sync. A = 1*, echo is sometimes still audible whereas in signals processed by *Sync. A = 0.99* echo is inaudible.

5. CONCLUSION

This paper presents the first cross-domain approach to synchronized acoustic echo cancellation and echo postfiltering. The proposed approach is based on the link between the system distance and the system mismatch power spectrum. A new system distance estimate is also introduced and assessed in this paper. The performance of the new synchronized echo control system is compared to synchronized Kalman echo control system and to an unsynchronized approach.

Our approach yields a significant reduction in distortion compared to the unsynchronized echo control system. The proposed system is robust to abrupt echo path changes and is stable during intervals of double-talk. The new system distance estimate delivers significantly improved echo suppression and rapid AEC convergence while preserving a reduced level of distortion quality during double-talk intervals compared to the standard system distance. Future work should include the extension of our synchronization approach to any other variable stepsizes.

6. REFERENCES

[1] E. Hänsler and G. Schmidt, *Acoustic Echo and Noise Control: A Practical Approach*. Wiley-Interscience, 2004.

[2] ——, "Hands-free telephones - joint control of echo cancellation and postfiltering," *Signal Processing*, vol. 80, no. 11, pp. 2295–2305, 2000.

[3] S. Haykin, *Adaptive Filter Theory*. Prentice Hall, 2002.

[4] G. Enzner and P. Vary, "Robust and elegant, purely statistical adaptation of acoustic echo canceler and postfilter," in *Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Kyoto, Japan, Sep. 2003.

[5] K. Steinert, M. Schönle, C. Beaugeant, and T. Fingscheidt, "Low-delay subband echo control in automotive environment," in *Proc. Biennial on DSP for in-Vehicule and Mobile Systems*, Istanbul, Turkey, Jun. 2007.

[6] G. Enzner and P. Vary, "Frequency-domain adaptive kalman filter for acoustic echo control in hands-free telephones," *Signal Processing*, vol. 86, no. 6, pp. 1140 – 1156, 2006.

[7] M. Mossi I., N. W. D. Evans, and C. Beaugeant, "An assessment of linear adaptive filter performance with nonlinear distortions," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2010.

[8] C. Yemdji, M. Mossi I., N. W. D. Evans, and C. Beaugeant, "A scalable architecture for linear convolution in the frequency domain for speech enhancement," in *Proc. Digital Signal Processing (DSP)*, 2011.

[9] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. Prentice Hall, 1999.

[10] C. Yemdji, M. Mossi I., N. W. D. Evans, and C. Beaugeant, "Efficient low delay filtering for residual echo suppression," in *Proc. European Signal Processing Conference (EUSIPCO)*, Aalborg, Denmark, Aug. 2010.

[11] Y. Ephraim and D. Malah, "Speech enhancement using optimal nonlinear spectral amplitude estimation," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Apr. 1983, pp. 1118–1121.

[12] S. Malik and G. Enzner, "Model-based vs. traditional frequency-domain adaptive filtering in the presence of continuous double-talk and acoustic echo path variability," in *Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC)*, 2008.

[13] T. Claasen and W. Mecklenbrauker, "Comparison of the convergence of two algorithms for adaptive fir digital filters," *IEEE Trans. on Circuits and Systems*, vol. 28, no. 6, pp. 510 – 518, Jun 1981.