Eurécom
Department of Multimedia Communication
2229, route des Crêtes
B.P. 193
06904 Sophia-Antipolis
FRANCE

Research Report RR-11-258

A Cascaded Non-linear Acoustic Echo Canceller Combining Power Filtering and Clipping Compensation

August 22$^{\text{nd}}$, 2011

Moctar I. Mossi, Christelle Yemdji
Nicholas Evans and Christophe Beaugeant

Tel : (+33) 4 93 00 81 00
Fax : (+33) 4 93 00 82 00
Email : {mossi, yemdji, evans}@eurecom.fr, christophe.beaugeant@intel.com

[1] Eurécom's research is partially supported by its industrial members: BMW Group Research & Technology - BMW Group Company, Bouygues Télécom, Cisco Systems, France Télécom, Hitachi Europe, SFR, Sharp, STMicroelectronics, Swisscom, Thales.

A Cascaded Non-linear Acoustic Echo Canceller Combining Power Filtering and Clipping Compensation

Moctar I. Mossi, Christelle Yemdji
Nicholas Evans and Christophe Beaugeant

Abstract

This report presents a novel, cascaded approach to non-linear acoustic echo cancellation (AEC). The loudspeaker enclosure microphone (LEM) system is divided into two blocks: non-linear clipping and power filtering, and a conventional, linear AEC. They represent the non-linear amplifier and loudspeaker and linear acoustic channel and up-link path in a typical mobile communication scenario. We propose an efficient approach for clipping compensation to improve the performance of the non-linear AEC in the presence of amplifier distortion. It is shown to give a reliable estimate of quasi static clipping in both artificial and practical environments with real speech signals. The cascaded approach to clipping compensation and power filtering is also more efficient than the alternative approach where clipping compensation is integrated into a higher-order power filter.

Index Terms

Echo cancellation, non-linear distortion, power filter, clipping, NLMS, Volterra.

1 INTRODUCTION

The problem of acoustic echo arises during mobile communication when a far-end signal is picked up by a near-end microphone. With the delay in the network the far-end user will thus hear their own delayed voice which can often perturb communication. To solve this problem acoustic echo cancellation (AEC) is commonly proposed as a solution [1]. Early AEC solutions are based on the assumed linearity of the loudspeaker enclosure microphone (LEM) system. Linear AECs improve the quality of communication and have proved very popular. With the growth of the mobile communication market and the miniaturization of devices, however, the linearity assumption does not always hold since small devices such as the loudspeaker are not well modelled by a linear system.

More recently AEC algorithms have been developed to tackle the problem of non-linearity. Non-linear solutions are generally based on Volterra series [2]. Unfortunately though, Volterra-based AEC algorithms are complex and converge too slowly for real time applications such as mobile communications. To tackle these problems many alternative solutions have been proposed over recent years [2]. Among them is the cascaded approach [3–5] which divides the LEM system into two sub-systems; a non-linear system representing the loudspeaker and amplifiers and a linear system representing the acoustic channel and the up-link path. This approach has been shown to deliver improved convergence particularly in dynamic, changing acoustic environments [6] but it still combines the effects of the loudspeaker and the amplifiers within a single model. The amplifiers and loudspeaker, however, exhibit quite different characteristics and thus a joint model is somewhat sub-optimal. Amplifier effects are well modelled by a clipping function [3] whereas power filter are better suited to loudspeaker effects. Independent models are thus more appropriate. This is the motivation for the work presented in this report, which aims to improve AEC performance though the independent modelling

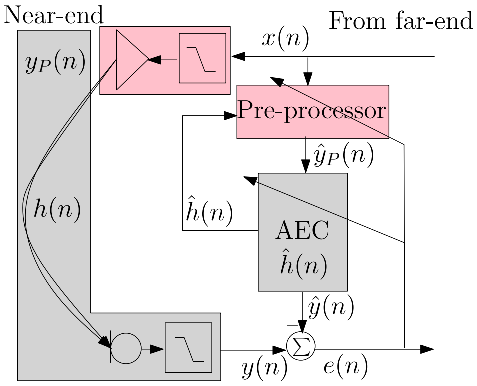

Figure 1: Cascaded non-linear acoustic echo cancellation

of amplifier and loudspeaker effects.

This report extends previous work by enhancing the cascaded model in [6] though the addition of a clipping compensation in order to model amplifier effects independently from those of the loudspeaker. The proposed approach avoid the modelling of two cascaded non-linear systems with one non-linear system which would require an increase in the order of the non-linearity, e.g. the cascade of two second-order Volterra filters requires a fourth-order Volterra filter. A second reason to use an independent clipping compensator is that the power filter, which would otherwise be used to model the clipping effect, is not very accurate even with Gram-Schmidt orthogonalization [7].

The remainder of this report is organized as follows: in Section 2 we present the new model. In Section 3 procedure to estimate the different parameters are derived. Then in Section 4 we present experimental work and analysis and finally in Section 5 we present our conclusions and perspectives.

2 NON-LINEAR AEC

In this section we present the two different processes of the proposed approach to non-linear AEC. They correspond to the model of the LEM system illustrated in

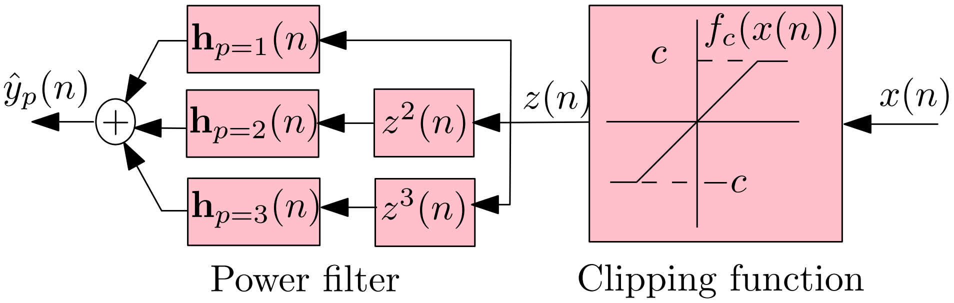

Figure 2: Pre-processor of the non-linear AEC: a concatenation of a clipping compensator model of amplifier and a power filter model of the loudspeaker.

Figure 1 which combines pre-processing and linear AEC modules.

2.1 Pre-processor

The pre-processor is used to model the characteristics of the down-link path, i.e. the amplifier and the loudspeaker. As illustrated in Figure 1 (top) the far-end signal $x(n)$ is first processed to obtain an output signal $\hat{y}_P(n)$ which is an estimate of the loudspeaker output. It is assumed here to be non-linear.

In general, due to limited power, the amplifier may introduce clipping distortion for high level signals. Clipping distortion is modelled here as in [3, 4] using a hard clipping model which is a function with a parameter c. As illustrated in Figure 2 the clipping function is given as:

$$z(n) = f_c(x(n)) = \begin{cases} sign(x(n))c & \text{if } |x(n)| \geq c \\ x(n) & \text{if } |x(n)| < c \end{cases} \tag{1}$$

where $c \geq 0$ is the absolute value of the clipping level.

The loudspeaker is also assumed to be non-linear and is modelled with a power filter as illustrated in Figure 2. The output $z(n)$ of the clipping function is processed by the power filter to obtain an estimate $\hat{y}_P$(n) of the loudspeaker output. The output $\hat{y}_P(n)$ of the power filter is a summation of the different sub-filter outputs $\mathbf{h}_{p=1,2,3}(n)$ which are filtered versions of the input signal at different powers. The pre-processor output $\hat{y}_P(n)$ is thus given by:

$$\hat{y}_P(n) = \sum_{p=1}^{P} \underbrace{\mathbf{h}_p(n)\mathbf{z}_p^T(n)}_{=\hat{y}_p(n)}$$

where $\mathbf{z}_p(n) = [z^p(n), z^p(n-1), \cdots, z^p(n-N_p)]^T$ is the input signal to the sub-filter $\mathbf{h}_p(n)$ with N_p taps and output $\hat{y}_p(n)$. The down-link path is assumed to have a low memory (short impulse response) and is static or changes slowly (compared to the acoustic channel) [4–6].

2.2 Linear AEC

The linear AEC aims to model the acoustic channel and the up-link path. As is generally assumed the acoustic path is modelled as a linear filter [1]. We furthermore suppose that it has a longer impulse response and is also more dynamic (compared to the down-link path) as described in [6].

The up-link path is also modelled as a linear filter even though its ouput can be non-linear. It generally involves only low-level signals from the loudspeaker, however, so that non-linearities can be safely neglected. The concatenation of the acoustic channel and up-link path can hence be modelled as a linear filter with a long impulse response of high variability.

The concatenation of the two linear systems is referred to collectively throughout the rest of this report as the acoustic path and is denoted by $\mathbf{h}(n)$. The filter $\mathbf{h}(n)$ provides a filtered version of the pre-processor estimate $\hat{y}_P(n)$ which is an estimate of the echo signal given by:

$$\hat{y}(n) = \mathbf{h}^T(n) \underbrace{\sum_{p=1}^{P} \mathbf{h}_p(n)\mathbf{Z}_p^T(n)}_{=\hat{\mathbf{y}}_P(n)}$$

where $\mathbf{Z}_p(n) = [\mathbf{z}_p(n), \mathbf{z}_p(n-1), \cdots, \mathbf{z}_p(n-N-1)]^T$ is an $N \times N_p$ input matrix of the filter $\mathbf{h}_p(n)$, where N is the length of the filter $\mathbf{h}(n)$. The matrix form of $z(n)$ is preferred here due to the concatenation of the two filters $\mathbf{h}(n)$ and $\mathbf{h}_p(n)$.

3 PARAMETER ESTIMATION

In this section we first present the cascade of the power filter and linear AEC algorithm according to [6]. Then we show how the clipping compensation can be efficiently incorporated into the model.

3.1 Pre-processor and linear AEC filter estimation

As the cascaded power filter and linear AEC system is presented in detail in [6] we give here the estimation procedure with minimal detail. To do so we ignore the clipping compensation in Figure 2 by assuming that $x(n) = z(n)$. The resulting system corresponds to the description given in Section 2.1 if the parameter c has a value higher than that of the maximum input signal and if no adaptation is applied. With this assumption the error of the global system is given by:

$$e(n) = y(n) - \mathbf{h}^T(n) \sum_{p=1}^{P} \mathbf{h}_p(n) \mathbf{Z}_p^T(n) \tag{2}$$

In a similar way as described in [6] the least mean square (LMS) approach is applied to minimise the error in an iterative fashion. The gradient is obtained by deriving the square of the error with respect to the filter parameters. The estimate of linear filter $\mathbf{h}(n)$ is then given by:

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \mu \hat{\mathbf{y}}_P(n) e(n), \tag{3}$$

and the pre-processor sub-filter $\mathbf{h}_p(n)$ by:

$$\hat{\mathbf{h}}_p(n+1) = \hat{\mathbf{h}}_p(n) + \mu_p \mathbf{Z}_p(n)\hat{\mathbf{h}}^T(n)e(n) \tag{4}$$

We note that even though they are sufficient to minimise (2) and thus to reliably estimate $y(n)$, equations (3) and (4) are dependent. They are thus not sufficient on their own for the identification of $\mathbf{h}(n)$ and $\mathbf{h}_p(n)$. since the system is under defined. This is not a problem, however, since we are concerned here only with the accuracy of $y(n)$ for which (3) and (4) are sufficient.

3.2 Clipping compensation

The proposed approach combines the clipping system proposed in [3, 4] with the cascaded model presented in [6]. We show here that the clipping compensation can be implemented with a complexity comparable to the system presented in [4] where no pre-processor is used. We again use the LMS approach to derive an adaptive clipping level estimator. The model presented here is based on a hard clipping model [3] (which could easily be extended to soft clipping) as given in 1. To derive a gradient for the estimator according to the LMS approach we need to incorporate the clipping function within an expression for the error $e(n)$ thus leading to:

$$e(n) = y(n) - \mathbf{h}^T(n)\sum_{p=1}^{P}\mathbf{h}_p^T(n)\underbrace{[f_c(\mathbf{X}(n))]_p}_{=\mathbf{Z}_p(n)}$$

where $[f_c(\mathbf{X}(n))]_p$ indicates that the function $f_c(x(n))$ is applied to each element of the matrix $\mathbf{X}(n) = [\mathbf{x}(n), \mathbf{x}(n-1), \cdots, \mathbf{x}(n-N)]$ where $\mathbf{x}(n) = [x(n), x(n-1), \cdots, x(n-N_p)]$.

Applying the LMS approach we can derive the clipping level estimator using the derivative of the error with respect to c which leads to:

$$\hat{c}(n+1) = \hat{c}(n) + \mu_c \mathbf{h}^T(n) \sum_{p=1}^{P} \mathbf{h}_p^T(n) \underbrace{[\dot{f}_c(\mathbf{X}(n))]_p}_{=\dot{\mathbf{Z}}_p(n)} e(n) \tag{5}$$

where $\dot{f}_c(x(n))$ is the derivative of $f_c(x(n))$ according to c. From (1) we see that $\dot{f}_c(x(n))$ is equal to:

$$\dot{f}_c(x(n)) = \begin{cases} sign(x(n)) & \text{if } |x(n)| \leq c \\ 0 & \text{elsewhere} \end{cases}$$

The gradient in (5) is highly complex due to the cascade of the pre-processor sub-filters and the linear filter. To simplify the computation of the gradient we assume that the clipping function affects only the fundamental component. We thus consider $z(n)$ to be composed of a linear component $z_l(n)$ and a non-linear distortion component $z_d(n)$ so that $z(n) = z_l(n) + z_d(n)$. We then suppose that the distortions within the power filter generated by $z_d(n)$ for $p \geq 2$ are negligible, i.e.

$$\underbrace{(|x(n)| - c)^p}_{x(n) \geq c, p \geq 2} \approx 0, \tag{6}$$

so that they can be safely ignored in the compensation.

In fact as we suppose that only the linear part ($p = 1$) is affected by the clipping, the error minimization that leads $\hat{c}(n)$ to converge to c will also minimize the error in the non-linear part ($p \geq 2$) as $\hat{c}$ is also applied to the non-linear part. This means that the approximation in (6) will be more effective when $\hat{c}(n)$ converges so that it can reach its optimal value in the minimum mean square error sense. This approximation implicitly assumes that $f_c(x(n))_{p\geq 2}$ is independent from c and leads to $(\dot{f}_c(x(n)))_{p\geq 2}$ being equal to zero. Equation (5) is thus simplified to:

$$\hat{c}(n+1) = \hat{c}(n) + \mu_c \mathbf{h}^T(n)\mathbf{h}_1^T(n)[\dot{f}_c(\mathbf{X}(n))]_1 e(n) \tag{7}$$

A second source of complexity relates to the cascade of the two filters $h(n) * h_1(n)$ in (7). In fact it is possible to use the estimates $(\hat{h}(n) * \hat{h}_1(n))$ but, in practice, they must be highly accurate otherwise (7) will be ineffective and gives poor performance. Another problem encountered using $\hat{h}(n) * \hat{h}_1(n)$ is that it leads to a more complex system since, for each iteration, $N \times N_1$ multiplications are required to compute the convolution. To overcome this problem we need to constrain one of the filters to be equal to $\delta(n)$ (Dirac function). In practice it is easier to set $\hat{h}_1(n) = \delta(n)$ as used in [4, 5] so that $h(n) * h_1(n) \approx \hat{h}(n)$. We can then rewrite (7) as:

$$\hat{c}(n+1) = \hat{c}(n) + \mu_c \hat{\mathbf{h}}^T(n)\dot{f}_c(\mathbf{x}_1(n))e(n) \tag{8}$$

which is less complex and amenable to real-time implementation. If instead we were to constrain $\hat{h}_1(n)$ to be equal to $\delta(n)$ then it will the estimate of the sub-filters $p \geq 2$ and the linear AEC. In this case the linear filter will converge to $h_1(n) * h(n)$ and the sub-filter $\hat{h}_p(n)$ will converge to $h_1^{-1}(n) * h_p(n)$.

Finally note that, in terms of implementation the pre-processor is not significantly different to the system presented in Section 3.1. The only change is that the first order sub-filter $\hat{\mathbf{h}}_1(n)$ is set to 1 and is not adaptive.

4 Experimental Work

To assess the proposed algorithm we use real speech signals, first in an artificial simulation and second with real data recorded on a mobile phone. In both cases

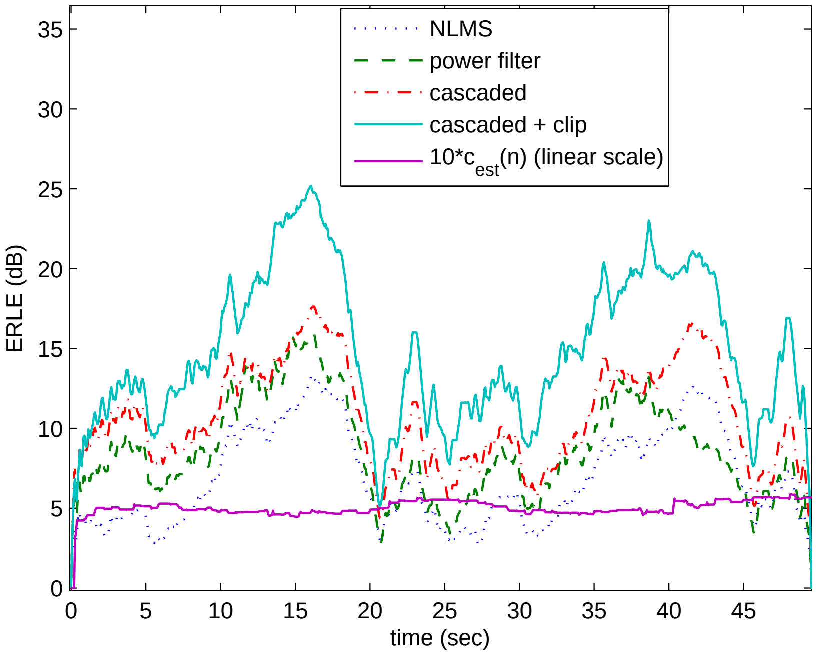

Figure 3: ERLE against time in a simulated, artificial environment. Profiles illustrated for NLMS, a power filter and the cascaded model with and without clipping compensation, and estimated clipping value (linear scale).

we compare the performance of four different AEC algorithms: a standard NLMS algorithm, the power filter alone and the cascaded filter with and without clipping compensation. The echo return loss enhancement (ERLE) metric is used to assess performance in all cases:

$$ERLE(m) = \frac{\sum_{n=m}^{m+M} y^2(n)}{\sum_{n=m}^{m+M} e^2(n)}$$

where $y(n)$ is the echo signal, $e(n)$ is the error signal and M is the frame length which is equal to 512 samples or 64ms for all experiments reported here.

4.1 Simulations

To simulate the LEM system we assume that amplifier clipping varies around the value $0.5 + \delta$ where $\delta = 0.5 - rand(1) * 0.09$. The $rand(1)$ function generates uniformally distributed noise in the range of $[0 - 1]$ and δ changes every 7s. When used only for loudspeaker modelling power filter has $P = 3$ sub-filters and where each sub-filter has $N_p = 50$ taps. The acoustic path (acoustic channel + up-link path) is simulated with echo paths measured in real environments using an impulse response with $N = 300$ taps and where the echo path changes every 10 seconds. Noise is added to the echo signal with a signal-to-noise ratio of 40dB.

The AEC algorithm is based on an NLMS approach using a filter with $N = 300$ taps. When used alone (i.e to model the full LEM system) the power filter has $N_{p=1,2,3} = 300$ taps. The cascaded model without clipping compensation has $N = 300$ taps and $N_{p=1,2,3} = 3$ taps whereas the cascaded model with clipping compensation has $N = 300$ taps, $h_1(n) = 1$ tap and $N_{p=2,3} = 5$ taps.

Results for the simulated environments are shown in Figure 3. We observe that the proposed cascaded model with clipping compensation delivers better performance than all other algorithms. This is expected since, even with Gram-Schimdt orthogonalization, the power filter cannot obtain an accurate estimate of the clipping model [7]. We also observe that the clipping level estimate ($10 * c_{est}(n)$ in Figure 3) fluctuates around 0.5 meaning that it is a good estimate of the real clipping level thus explains the observed performance with the proposed model. Upon comparison of results for the cascaded model without clipping compensation and the power filter, we observe that the cascaded model has better performance. This is explained by the fact that it has better tracking behaviour than the power filter. We observe that, upon every path change (each 10s), the cascaded model shows faster convergence. At time 40s the NLMS algorithm, however, shows better performance than the power filter. This is due to a change in the path delay so that the

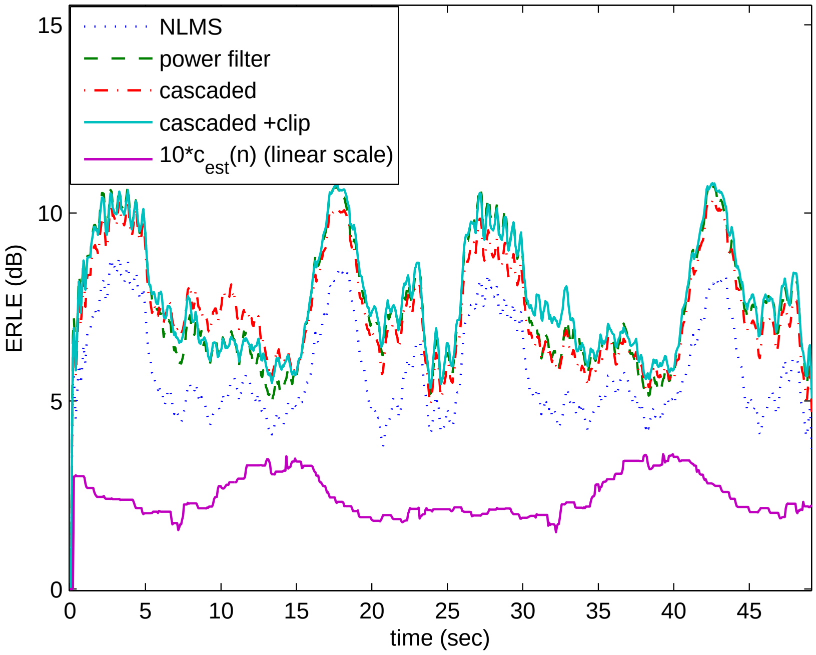

Figure 4: ERLE against time in a real environment. Profiles illustrated for NLMS, the power filter and the cascaded model with and without clipping compensation, and estimated clipping value (linear scale).

sub-filters ($\mathbf{h}_{p=2,3}(n)$ with $N_{p=2,3} = 300$) of the power filter need more time to reconverge as they necessarily use lower step-sizes to ensure stability.

4.2 Real data

Extensive tests (not reported here) show that for the real environment the best choice of acoustic path length is around 80 taps. Experiments reported here correspond to an AEC algorithm with $N = 80$ taps, a power filter with $N_{p=1,2,3} = 80$ taps, a cascaded model (without clipping compensation) with $N = 80$ taps and $N_{p=1,2,3} = 3$ taps and a cascaded model with clipping compensation with $N = 80$ taps, $N_1 = 1$ tap and $N_{p=2,3} = 5$ taps.

Results are shown in Figure 4. We observe that all the non-linear AEC algorithms have comparable results whereas the linear NLMS AEC is noticeably worse. The fact that the non-linear approaches now show similar behaviour can be explained by the shorter echo path since the cascaded model delivers better performance for longer impulse responses. During initialization we see that non-linear algorithms have comparable performance but the cascaded model without clipping compensation shows better performance thereafter and, in particular, during the clipping level changes between 8 and 13s. Generally, though, the proposed model shows better performance compare to the other algorithm even if sometimes the differences are small. This is normal due to the small length of the acoustic path which results in the power filter having similar convergence behaviour to the cascaded model and the fact that the cascaded model assumes a time invariant pre-processor.

A factor that affects the performance of the proposed model in tracking the clipping level variations is its stability for which a lower step size is required. Of interest, however, is that even when the clipping level estimator diverges it does not affect the performance of the rest of the system as $\hat{c}$ will be higher than $c(n)$, under which conditions $z(n) = x(n)$. Note also that, in higher noise environments, the proposed system will also provide a similar performance to the same model without clipping compensation, as the noise level will mask the clipping effect or the input signal level may be low so that clipping effects arise only during very short periods.

5 Conclusions

In this report we propose a new approach to combine clipping and power series non-linearity compensation for non-linear AEC. This approach is simplified so that the complexity of the clipping compensator is not adversely affected by it's use in cascaded with a power filter model of the loudspeaker which would otherwise lead to unrealistic demands of computational power. We show that the approximations used to reduce the complexity of the proposed compensator do not affect accuarcy and deliver a reliable estimate of the real clipping level. We show that the proposed approach improves non-linear AEC performance when the clipping level is quasi static.

Also shown is the difficulty in tracking changes in the clipping level. Future work should invole a comparative study of the effects from changing clipping levels and other sources of distortion (i.e. noise and echo path changes). If the clipping level is known to be changing then increased adaptation rates in such periods (with paused adaptation of the power and linear AEC filters) may give improved performance.

References

[1] C. Breining, P. Dreiseitel, E. Hänsler, k. Mader, B. Nitsch, H. Puder, T. Schertler, G. Schmidt, and J. Tilp, “Acoustic Echo Control, An Application of Very-High-Order Adaptive Filter,” *IEEE SP magazine*, pp. 42–69, July 1999.

[2] A. Fermo, A. Carini, and G. Sicuranza, “Analysis of Different Low Complexity Nonlinear Filters for Acoustic Echo Cancellation,” *IWISPA*, pp. 261–266, June 2000.

[3] B. S. Nollett and D. L. Jones, “Nonlinear Echo Cancellation for Hands-Free Speakerphones,” *NSIP*, 1997.

[4] A. Stenger and W. Kellermann, “Adaptation of a Memoryless Preprocessor for Nonlinear Acoustic Echo Cancelling,” *Signal Processing*, vol. 80, pp. 1747–1760, Feb 2000.

[5] A. Guerin, G. Faucon, and R. L. Bouquin-Jeannes, “Nonlinear Acoustic Echo Cancellation Based on Volterra Filters,” *IEEE Trans. on Speech and Audio Proc.*, vol. 11, pp. 672 – 683, Nov 2003.

[6] M. I. Mossi, C. Yemdji, N. W. D. Evans, C. Beaugeant, and P. Degry, “Robust and Low-Cost Non-linear Acoustic Echo Cancellation,” *ICASSP*, Mar 2010.

[7] S. Malik and G. Enzner, “Fourier Expansion of Hammerstein Models for Nonlinear Acoustic System Identification,” *ICASSP*, March 2011.