NON-LINEAR ACOUSTIC ECHO CANCELLATION USING ONLINE LOUDSPEAKER LINEARIZATION

Moctar I. Mossi, Christelle Yemdji, Nicholas Evans

EURECOM
Sophia Antipolis, France
{mossi, yemdji, evans}@eurecom.fr

Christophe Beaugeant

Intel Mobile Communications
Sophia-Antipolis, France
christophe.beaugeant@intel.com

ABSTRACT

This paper presents an approach to non-linear acoustic echo cancellation (AEC). We first present the model of a loudspeaker enclosure microphone system which is divided into two blocks: a non-linear, power filter model for the down-link path (loudspeaker and amplifiers) and a linear model for the acoustic channel and up-link path. Using this model we propose an approach that uses loudspeaker linearization and linear AEC to improve performance of an otherwise classical approach to linear AEC. The novel contribution in this paper relates to a new on-line linearization pre-processing algorithm that adapts to long-term variations in the loudspeaker characteristics. This feature contrasts with fixed pre-processor aglorithms which have been reported previously.

***Index Terms*—** Adaptive filtering, AEC, loudspeaker linearization, non-linearity, power filter, Volterra.

1. INTRODUCTION

In typical mobile communications scenarios a part of the loudspeaker signal is generally acquired by the device's own microphone and is transmitted back to the far-end user. With the delay introduced by the network the resulting echo signal can disturb the listener. A classical approach to solve this problem involves the use of acoustic echo cancellation (AEC) [1]. In most approaches the system model is linear and acceptable performance is often obtained. However, in the highly competitive mobile device market there is a trend toward miniaturization, in addition to a preference for more economical components and manufacturing. This has given rise to a new challenge; namely the problem of non-linearity due to small transducers, especially in the case of loudspeakers.

To overcome the problem of non-linear echo two main approaches have emerged: non-linear adaptive filtering and residual non-linear echo suppression. In the first case, cascaded [2, 3], parallel [4], and loudspeaker linearization approaches [5] have all proved popular. The second case involves post-processing to suppress residual non-linear echo [6]. Both approaches have their advantages and drawbacks. Non-linear adaptive filtering typically suffers from slow convergence and high complexity, whereas residual non-linear echo suppression is less complex but often suffers from distortion in the near-end speech signal due to over-estimation or under-estimation of the residual echo in the frequency domain.

In this paper we focus on non-linear adaptive filtering based on loudspeaker linearization where the loudspeaker input is pre-processed by a non-linear filter, referred to here as a linearization pre-processor. It aims to compensate for non-linearities that are subsequently introduced by the loudspeaker so that, when combined, the linearization pre-processor and loudspeaker form a linear system. Loudspeaker linearization then permits the use of conventional linear AEC. In terms of echo reduction performance is improved and, as the linearization pre-processor relies only on the loudspeaker characteristics, it does not need to be re-initialized when the acoustic environment changes. This is a distinct benefit over alternative post-loudspeaker approaches which depend fundamentally on the acoustic path and thus suffer from convergence issues when the echo path changes.

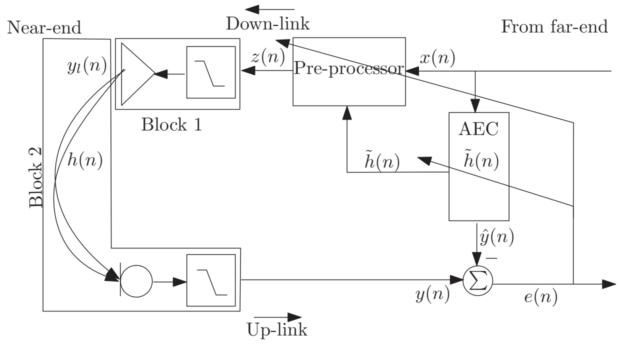


Figure 1: 'Loudspeaker linearization and acoustic echo cancellation where the LEM is divided into two parts. The first is a nonlinear model and the second is a linear model.

In practice, however, loudspeaker linearization is rarely used with AEC since there is no direct access to the loudspeaker output. A solution proposed in [5] renders the system dependent to the device and is based upon an inverse, static model of the loudspeaker. Transducer characteristics are dynamic, however, and in practice such solutions can sometimes even increase perturbation instead of reducing non-linear echo. The solution proposed in this paper uses a new on-line loudspeaker linearization approach which enables the tracking of long-term variation.

The remainder of the paper is organized as follow. In Section 2 we present the loudspeaker linearization approach. In Section 3 we describe the overall system, its operation and behaviour. Test results are presented in Section 4. Finally we present our conclusions in Section 5.

2. LEM SYSTEM MODEL

The loudspeaker enclosure microphone (LEM) model presented here is based on the most commonly used AEC model where the main source of non-linearity is assumed to be the loudspeaker and where a power filter model is further assumed to be sufficient for modelling non-linearity [3, 4, 7]. Though the implementation is not detailed her the proposed approach can be easily extended to a Volterra series [8].

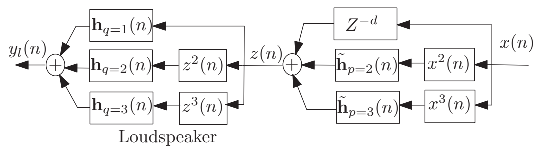


Figure 2: Loudspeaker linearization system.

For some non-linear AEC applications, it is beneficial to divide the global LEM system into two models as illustrated in Figure 1: a non-linear model (block 1) and a linear model (block 2). The first model represents the down-link path (amplifiers and loudspeaker) and is considered as a power non-linear system. It has a small impulse response and slow variability. The second model represents the acoustic channel and up-link path. It is assumed to be linear with a longer impulse response and greater variability. Thus the overall system is composed of an adaptive pre-processor that aims to linearize the loudspeaker output and a linear AEC module which tracks the acoustic path (acoustic channel + down-link devices).

The loudspeaker linearization system is illustrated in Figure 2 and corresponds to the down-link path in Figure 1, including the pre-processor and the loudspeaker. The far-end signal $x(n)$ forms the input to the pre-processor and the linear acoustic echo canceller. According to the power non-linear model the discrete loudspeaker output $y_l(n)$ can be written as:

$$y_l(n) = \sum_{q=1}^{Q} \mathbf{h}_q(n)\mathbf{z}_q^T(n) \tag{1}$$

where the vector $\mathbf{z}(n)$ is the input, $\mathbf{h}_q(n)$ is the sub-filter applied to the q^{th} power of $z(n)$ and Q is the maximum signal exponent. The vector $\mathbf{z}_q(n)$ is given by:

$$\mathbf{z}_q(n) = [z^q(n), z^q(n-1), \cdots, z^q(n-M_q-1)]^T \tag{2}$$

where M_q is the length of each filter $\mathbf{h}_q(n)$ and is always independent of q for all the work reported here.

There is no need to process the linear component of x(n) (top path) which is instead delayed by d samples and corresponds to the processing delay of the non-linear components ($p \geq 2$). Together they are used to generate an output that aims to compensate for the non-linearities which are subsequently introduced by the loudspeaker [8]. The pre-processor output signal can be written as:

$$z(n) = x(n) + \sum_{p=2}^{P} \tilde{\mathbf{h}}_p(n)\mathbf{x}_p^T(n) \tag{3}$$

where $z(n)$ is the output of the pre-processor, the vector $\mathbf{x}_p(n)$ is the far-end signal vector and $\tilde{\mathbf{h}}_p(n)$ is the sub-filter of the p^{th} power input. The objective is to obtain a loudspeaker output such that $y_l(n) \approx \mathbf{h}_1(n)\mathbf{x}^T(n)$ where $\mathbf{h}_1(n)$ is the linear loudspeaker impulse response. If this approximation is reached then the resulting echo signal can be estimated using a conventional linear adaptive AEC filter such as the least mean square (LMS) algorithm.

3. NON-LINEAR AEC

In this section we present the proposed non-linear AEC algorithm which is based on the well-known least mean square (LMS) approach.

3.1. AEC Filtering

To derive the estimate of the AEC filter and linearization pre-processor we need to express the the AEC system error ($e(n)$ in Figure 1) according to the different model parameters. According to (1) and (3) the discrete loudspeaker output can be rewritten as:

$$\begin{aligned}
y_l(n) &= \sum_{q=1}^{Q} \mathbf{h}_q(n)\left\{x(n) + \sum_{p=2}^{P} \tilde{\mathbf{h}}_p(n)\mathbf{x}_p^T(n)\right\}_q \\
&= \sum_{q=1}^{Q} \mathbf{h}_q(n)\left\{\mathbf{x}_q^T(n) + \sum_{p=2}^{P} \tilde{\mathbf{h}}_p(n)\mathbf{x}_{p*q}^T(n)\right\} \\
&= \mathbf{h}_1(n)\mathbf{x}_1^T(n) + \sum_{q=2}^{Q} \mathbf{h}_q(n)\mathbf{x}_q(n) \\
&+ \sum_{p=2}^{P} \mathbf{h}_1(n)[\tilde{\mathbf{h}}_p(n)\mathbf{X}_{p*(q=1)}^T(n)]^T \\
&+ \underbrace{\sum_{q=2}^{Q}\sum_{p=2}^{P} \mathbf{h}_q(n)[\tilde{\mathbf{h}}_p(n)\mathbf{X}_{p*q}^T(n)]^T}_{\text{neglected}(p*q\geq 4)}
\end{aligned}$$

where $\mathbf{X}_{p*q}(n)$ is an $M_p \times M_q$ matrix form of the signal $x^{(p*q)}(n)$ given by:

$$\mathbf{X}_{p*q}(n) = [\mathbf{x}_{p*q}(n), \mathbf{x}_{p*q}(n-1), \cdots, \mathbf{x}_{p*q}(n-M_p-1)]$$

where $\mathbf{x}_{p*q}(n)$ is a vector of length Mq defined as in (2). M_p and M_q are respectively the length of filters $\tilde{\mathbf{h}}_p(n)$ and $\mathbf{h}_q(n)$. We assume that the highest order terms are negligible and that the non-linearity can be modelled sufficiently with $P = 3$. Experiments performed by other authors and with real loudspeakers show that the performance benefit obtained from the inclusion of higher order terms does not justify the extra complexity [4, 5, 8]. The loudspeaker output can therefore be approximated as:

$$\begin{aligned}
y_l(n) &= \mathbf{h}_1(n)\mathbf{x}_1^T(n) + \sum_{q=2}^{Q} \mathbf{h}_q(n)\mathbf{x}_q(n) \\
&+ \sum_{p=2}^{P} \mathbf{h}_1(n)[\tilde{\mathbf{h}}_p(n)\mathbf{X}_{p\cdot(q=1)}^T(n)]^T
\end{aligned}$$

The output of the loudspeaker is convolved with the acoustic path $\mathbf{h}(n)$ (acoustic channel + up-link):

$$\begin{aligned}
y(n) &= \mathbf{h}(n)[\mathbf{h}_1(n)\mathbf{X}_1^T(n)]^T + \sum_{q=2}^{Q} \mathbf{h}(n)[\mathbf{h}_q(n)\mathbf{X}_q^T(n)]^T \\
&+ \sum_{p=2}^{P} (\mathbf{h}(n)*\mathbf{h}_1(n))[\tilde{\mathbf{h}}_p(n)\mathbf{X}_{p*(q=1)}^T(n)]^T
\end{aligned}$$

The AEC output is given by:

$$\hat{y}(n) = \tilde{\mathbf{h}}(n)\mathbf{x}^T(n)$$

and the error between the echo and its estimate is given by:

$$e(n) = y(n) - \hat{y}(n). \tag{4}$$

The error is used to obtain an adaptive estimate of the linear filter [9]. We assume that the linear echo component is dominant and thus that we have direct access to it. Using the LMS approach the adaptation of the AEC filter is given by:

$$\tilde{\mathbf{h}}(n+1) = \tilde{\mathbf{h}}(n) + \mu e(n)\mathbf{x}(n) \tag{5}$$

and, after sufficient iterations, $\tilde{\mathbf{h}}(n)$, will converge to $\mathbf{h}_l$(n), where $\mathbf{h}_l(n)$ is the linear filter such that its convolution with $x(n)$ gives the linear echo component. Note from the first term of (4), which represents the linear echo component that $h_l(n) = h_1(n) * h(n)$.

3.2. Linearization Pre-processing

In the same way as for the AEC filter the sub-filters of the linearization pre-processor are estimated using the LMS approach, leading to:

$$\tilde{\mathbf{h}}_p(n+1) = \tilde{\mathbf{h}}_p(n) + \mu \frac{\delta e^2(n)}{\delta \tilde{\mathbf{h}}_p(n)}$$

By deriving the square of the error with respect to $\tilde{\mathbf{h}}_{p=2,3}(n)$ we obtain:

$$\tilde{\mathbf{h}}_p(n+1) = \tilde{\mathbf{h}}_p(n) + \mu e(n)\mathbf{h}(n)[\mathbf{h}_1(n)\mathbf{X}^T_{p\cdot(q=1)}(n)]^T \tag{6}$$

In (6) the filter $h_l(n) = h(n) * h_1(n)$ is unknown. To overcome this problem an estimate $\tilde{\mathbf{h}}(n)$ in (5) is used and leads to:

$$\tilde{\mathbf{h}}_p(n+1) = \tilde{\mathbf{h}}_p(n) + \mu_n e(n)\tilde{\mathbf{h}}(n)\mathbf{x}^T_{p\cdot(q=1)}(n) \tag{7}$$

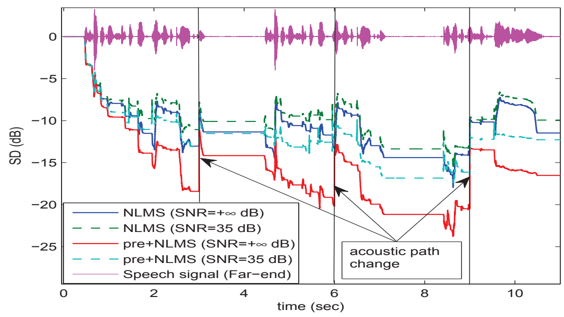
where μ_n is a normalized step-size equal to $\frac{\mu}{|\tilde{\mathbf{h}}(n)\mathbf{X}^T_{p\cdot(q=1)}(n)|^2}$ with $0 < \mu \leq 1$. Equation (7) provides a solution for the linearization of the loudspeaker in non-linear echo environments.

From (7), we see that the pre-processor updating process uses the estimate of the AEC, $\tilde{\mathbf{h}}(n)$, meaning that the linear component should be dominant. As the estimate $\tilde{\mathbf{h}}(n)$ of the AEC is used to estimate the sub-filters, $\tilde{\mathbf{h}}_{p=2,3}(n)$, it is important to ensure that the pre-processor still depends only on the loudspeaker characteristics. This means that the sub-filter estimates should converge to a fixed filter which depends only on $\mathbf{h}_{q=1,2,3}(n)$ (loudspeaker characteristics). The independency of the pre-processor to $\mathbf{h}(n)$ (acoustic path) is needed to ensure stability to changes in the echo path characteristics. We thus extend (4) to:

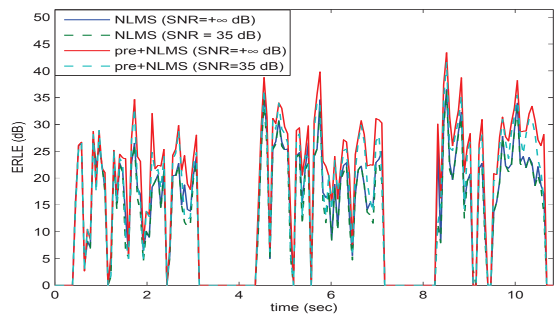
$$e(n) = \underbrace{\left(\mathbf{h}_l(n) - \tilde{\mathbf{h}}(n)\right)\mathbf{x}_1^T(n)}_{\text{linear component}} + \underbrace{\mathbf{h}_l(n)\sum_{q=2}^{Q}\left([\mathbf{h}_1^{(-1)}(n)\mathbf{h}_q(n)] + \tilde{\mathbf{h}}_p(n)\right)\mathbf{x}^T_{q\cdot(q=1)}(n)}_{\text{non-linear component}} \tag{8}$$

Equation (8) shows that, if the linear component $\tilde{\mathbf{h}}(n)$ is an estimate of $\mathbf{h}_l(n)$, the first term (linear component) in (8) goes to zero.

To minimize the second term (non-linear component) the estimate of each filter $\tilde{h}_p(n)$ should converge to $-h_1^{(-1)}(n) * h_{q=p}(n)$ with $h_1^{(-1)}(n) * h_1(n) = \delta(n)$ (where $\delta(n)$ is the Dirac function). This shows that $\tilde{\mathbf{h}}_p(n)$ is independent of the acoustic path. Thus, with a reliable estimate of the pre-processor, the updating process can be frozen without degrading the performance of the overall system,But this approach is not presented here. Thus is potentially beneficial in terms of computational complexity and for robustness in adverse environments.



a) System distance



b) Echo attenuation

Figure 3: caption should be 'NLMS performance comparison with and without linearization pre-processing under noise-free and noisy conditions with an SNR of 30dB.

4. TESTS RESULTS

In this section we report simulation results which aim to validate the new, online approach to loudspeaker linearization.

4.1. Simulation Set-up

Simulations were performed with loudspeaker sub-filters $\mathbf{h}_{q=1,2,3}(n)$ of 50 taps and with an acoustic channel filter $\mathbf{h}(n)$ of 150 taps. To evaluate convergence behavior and robustness to changes in the acoustic channel, abrupt changes in $\mathbf{h}(n)$ are introduced every 3 seconds. Thus the loudspeaker output is convolved with a sequence of three different acoustic paths $\mathbf{h}(n)$ which were all measured in real environments. Tests are conducted both in a noise-free environment and a noisy environment with a signal-to-noise-ratio (SNR) of 35 dB, where the SNR is the ratio between the echo signal and white noise. The AEC filter $\tilde{\mathbf{h}}(n)$ has 200 taps (150 for $\mathbf{h}(n)$ and 50 for $\mathbf{h}_1(n)$) whereas the pre-processor $\tilde{\mathbf{h}}(n)_{p=2,3}$ has 50 taps.

4.2. Linearization Performance

Linearization performance is assessed using the system distance between the linear filter, which results from the cascade of the loudspeaker and the acoustic path ($h_l(n) = h_1(n) * h(n)$), and the AEC filter $\hat{\mathbf{h}}(n)$. Note that the convolution of $\mathbf{h}_l(n)$ with the far-end signal gives the linear echo component. In this case the sys-

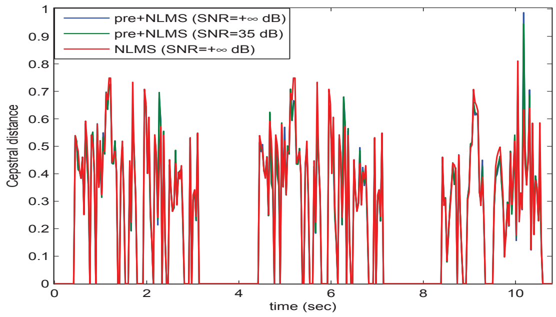


Figure 4: Cepstral distance between the far-end signal and the loudspeaker output.

tem distance is given by $SD(n) = |\frac{\mathbf{h}_l(n)-\tilde{\mathbf{h}}(n)}{\mathbf{h}_l(n)}|$. We compare the system distance of the NLMS algorithm both with and without linearization pre-processing. Results are presented in Figure 3 (a) and show the system distance against time for each configuration. For both noise-free and noisy conditions the proposed system is shown to achieve better accuracy and shows an SD reduction in the order of 3 to 7 dB. By setting the step size of the pre-processor to a value lower than that of the linear AEC filter, the pre-processor has limited influence during echo path changes. This ensures that the linear AEC re-converges quickly to the new acoustic path.

We also observe that the proposed algorithm is robust to noise; the estimated acoustic path is more accurate in noisy conditions than in noise-free conditions. This is because, on account of the correlation between non-linearities and the input signal, the conventional linear AEC system cannot track the linear component in non-linear conditions.

Of further interest in Figure 3 (a) is a slight spike in the SD profile between 4s and 5s. This is caused by a higher-level input signal which introduces more non-linearity which in turn disturbs the NLMS algorithm. Since higher order powers are neglected in (4) the linearization pre-processor is ineffective in this case. Though not reported here, this behavior can be avoided by simply controlling the output of the non-linear pre-processor component. As informative as system distance metrics are, loudspeaker linearization is not the real target of AEC and thus a comparison in terms of echo reduction is also needed.

4.3. Echo Return Loss Enhancement

Echo reduction performance is expressed in terms of the echo return loss enhancement (ERLE) in Figure 3 (b). Profiles show that better echo reduction is achieved with the linearization pre-processor in both noisy and noise-free conditions. Upon changes in the echo path (every 3s) the system still provides good performance. As the ERLE is the ratio between the echo and the error it shows that, even in noise-free environments, the linearization procedure enhances the system even though one might expect that the omission of higher-order power terms in (4) may result in increased residual echo. This result thus validates the assumption that higher order terms may be safely ignored.

4.4. Effect on Loudspeaker Output

To assess the effect of pre-processor adaptation we use the cepstral distance. Results are illustrated in Figure 4 which shows the cepstral distance between the far-end signal and the output of the loudspeaker, with and without pre-processing. We see that, for an SNR of 35 dB, profiles corresponding to noisy (green profile) and noise-free environments (blue profile) overlap significantly, thereby showing a certain robustness to moderate levels of noise. The second observation is that, even with linearization pre-processing, the loudspeaker output is not significantly distorted. Furthermore we observe that, without pre-processing, the mean cepstral distance of 0.4245 is comparable to that of 0.4175 obtained with pre-processing.

The time domain waveform shows that the level of the pre-processed output is slightly lower than that of the non-processed output. This is due to the fact that the energy of non-linearities is reduced. We also observe that, as expected, the difference is more significant for higher signal amplitudes.

5. CONCLUSIONS

This paper proposes a new, online loudspeaker linearization approach for non-linear echo cancellation. The proposed linearization pre-processor updating process relies on the estimate of the linear AEC filter but the pre-pocessor converges to a filter that is independent from the actual acoustic echo path. We also show that this approach is efficient when the linear echo component is dominant, which is the case in practice. Under this assumption the output of the loudspeaker is effectively linearized thus resulting in better linear AEC performance. Tests in noisy conditions show that the use of a pre-processor can provide better performance without distorting the loudspeaker signal. This is an important benefit in mobile communications. The pre-processor sub-filters further depend only on the loudspeaker characteristics and thus the pre-processor does not require any re-initialization. The new, online approach is thus an appealing alternative to current state-of-the-art approaches to non-linear acoustic echo cancellation.

6. REFERENCES

[1] C. Breining, P. Dreiseitd, 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] B. S. Nollett and D. L. Jones, "Adaptation of a memoryless preprocessor for nonlinear acoustic echo cancelling," *IWAENC*, 1997.

[3] A. Stenger and W. Kellermann, "Adaptation of a memoryless preprocessor for nonlinear acoustic echo cancelling," *Signal Processing*, vol. 80, pp. 1747–1760, Feb 2000.

[4] F. Kuech, A. Mitnacht, and W. Kellermann, "Nonlinear acoustic echo cancellation using adaptive orthogonalized power filters," *ICASSP*, vol. III, pp. 105–108, Mar 2005.

[5] H. Furuhashi, Y. Kajikawa, and Y. Nomura, "Realization of nonlinear acoustic echo cancellation by subband parallel cascade volterra filter," *ISPACS*, pp. 837–840, Dec 2006.

[6] O. Hoshuyama and A. Sugiyama, "An acoustic echo suppressor based on a frequency-domain model of highly nonlinear residual echo," *ICASSP*, vol. 5, May 2006.

[7] M. I. Mossi, C. Yemdji, N. W. D. Evans, C. Herglotz, C. Beaugeant, and P. Degry, "New models for characterizing non-linear distortions in mobile terminal loudspeakers," *IWAENC*, Sept 2010.

[8] W. Frank, "An efficient approximation to the quadratic volterra filter and its application in real-time loudspeaker linearization," *Signal Processing*, vol. 45, pp. 97–113, Jul 1995.

[9] S. Haykin, *Adaptive Filter Theory* 4^{th} *Ed.* Prentice Hall, 2001.