

A comparative assessment of noise and non-linear echo effects in acoustic echo cancellation

Moctar I. Mossi, Christelle Yemdji and Nicholas Evans
EURECOM
06560 Sophia-Antipols, France
{mossi, yemdji, evans}@eurecom.fr
www.eurecom.fr

Christophe Beaugeant
Infineon Technologies
06560 Sophia-Antipolis, France
{christophe.beaugeant}@infineon.com
www.infineon.com

Abstract—This paper addresses the problem of adaptive filtering for acoustic echo cancellation in noisy and non-linear environments. The first contribution relates to a new analysis on the comparative impact of additive noise and non-linear echo on the performance of adaptive filtering for linear acoustic echo cancellation (AEC). A comprehensive performance assessment is reported, including echo return loss enhancement (ERLE), convergence time and system distance metrics. This work better highlights differences between algorithm performance than previously published work and sheds new light on algorithm behavior. Results show that, in non-linear and noisy environments, the normalized-least mean square (NLMS) algorithm gives similar performance to the more complex affine projection algorithm (APA). The more computationally efficient frequency block least mean square (FBLMS) algorithm is particularly adversely effected and gives poorer performance than the basic least mean square (LMS) approach. These observations question the pursuit of increased computational efficiency and reduced convergence time over robustness to distortions. The second contribution relates to an original account of the effects of non-linear echo and noise which, perhaps surprisingly, are greater for the latter. This observation highlights the need for more comprehensive studies on the effects of non-linear distortion and supports continuing efforts to tackle non-linear echo.

Keywords: echo cancellation, non-linear distortion, noise, AEC, LMS, NLMS, APA, FBLMS.

I. INTRODUCTION

There is a vast volume of work in the literature related to adaptive filtering in noisy and non-linear environments. Noisy environments are usually accommodated by restricting the rate of filter adaptation so as to limit divergence, e.g. [1], [2], [3]. For non-linear environments, i.e. involving non-linear echo, two dominant solutions have emerged. The first is based on the Volterra filter [4] and the second involves post-filtering [5] in combination with acoustic echo cancellation (AEC) adaptive filtering. The Volterra solution is generally slow to convergence and is highly computationally complex. Post-filters are less complex but rely on the performance of linear adaptive filters that are still disturbed by non-linear echo. Linear adaptive filtering is still popular in these situations and it is thus of interest to assess their performance in such environments, as reported in our previous work [6].

In [6] the performance of five adaptive filters was assessed by comparing the echo return loss enhancement (ERLE) in the presence of (i) linear echo and (ii) linear and non-linear echo. This work showed that, with higher levels of non-linear distortion, the NLMS algorithm can perform as well as APA and that the lower complexity FBLMS algorithm gives the worst performance in the same environment. However, the assessment protocol in [6] perhaps does not best reflect the true robustness of each algorithm to non-linear echo alone, since the comparison was made independently of the ERLE that is achieved with each algorithm under conditions involving only linear echo.

The first contribution in this paper is a new, comparative assessment of the same linear adaptive filters, but here with a different focus on their relative robustness to non-linear and noisy environments with an improved, fairer experimental setup. Compared to [6], where the level of steady state echo reduction is different for all adaptive filters, in all experiments reported here we tuned the different AEC algorithms so that they each obtain the same ERLE after convergence

under linear echo conditions. Also, given the well-known trade-off between steady state ERLE and the adaptation step size, which is used to control the rate of converge, the same step size is used for all algorithms. With the same amount of ERLE under linear echo conditions and a fixed step size, the assessment reported here can more reliably attribute differences in performance to differences in the prevailing environment, i.e. non-linear echo or noise, and less so due to differences in algorithm configurations. The second contribution relates to a comparison of the two environments to determine whether or not the effects of non-linearities resemble those of noise and thus to validate, or otherwise, the common assumption that the two effects are practically equivalent.

The remainder of the paper is organized as follows. In Section II we describe a general echo cancellation system and the different approaches to AEC that are investigated here. Our experimental work is presented in Section III and our conclusions in Section IV.

II. ACOUSTIC ECHO CANCELLATION

We here describe a typical system/echo model and a general framework for AEC with adaptive filtering. Also described are the four approaches to AEC that are investigated in this paper.

A. System/echo model

A general system/echo model, which was used for all experiments reported in this paper, is illustrated in Figure 1. The terminal receives a down-link (or loudspeaker) signal $x(n)$ from a far-end speaker, and transmits an uplink (or microphone) signal $y(n)$. In addition to near-end speech $s(n)$ and additive background noise $n(n)$ the uplink signal potentially includes an additional echo component $d(n)$, which is a result of the acoustical coupling between the loudspeaker and the microphone. It is generally modelled with a linear convolution, $d(n) = x(n) * h(n)$, where $h(n)$ is the impulse response which characterizes the acoustical coupling. AEC may thus be implemented by estimating $h(n)$ with a filter $\hat{h}(n)$ in order to give an estimate of the coupled echo signal $\hat{d}(n) = x(n) * \hat{h}(n)$. The echo is attenuated simply by subtracting $\hat{d}(n)$ from the uplink signal. Since the acoustical coupling is generally time varying $\hat{h}(n)$ is usually an adaptive filter. Near-end speech disturbs the adaptive filter and so $\hat{h}(n)$ is usually updated during echo-only periods, i.e. when $s(n) = 0$. Noise can also disturb the adaptive filter but, if we also suppose that the noise is negligible, i.e. $n(n) = 0$, then $y(n) = d(n)$ and thus the resulting error signal, $e(n)$ is simply the difference between the echo signal and its estimate, i.e. $e(n) = d(n) - \hat{d}(n)$. The error $e(n)$ is used to update the filter $h(n)$ whose goal is to drive $e(n)$ to zero.

AEC rarely operates under such ideal conditions, however, and thus it is interesting to study the robustness under more realistic conditions. i.e. with near-end speech, non-linear echo and additive background noise. As adaptation is simply paused during intervals of near-end speech, only disturbances from non-linear echo and background noise are considered here. Each of the approaches to AEC that are considered are described below.

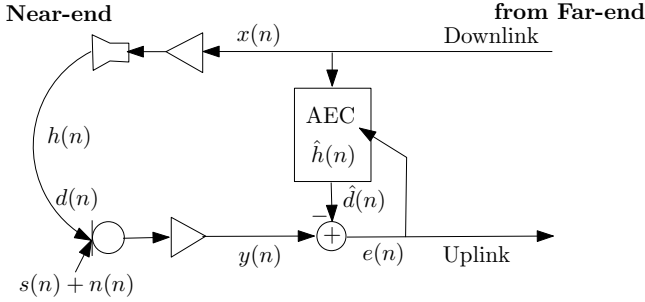


Fig. 1. System/echo model illustrating the acoustical coupling between the loudspeaker and microphone and a general approach to adaptive AEC.

B. Linear adaptive filtering algorithms

The adaptive AEC filters considered in this paper are updated according to a general adaptation recursion given by:

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \Delta\mathbf{h}(n), \quad (1)$$

where $\hat{\mathbf{h}}(n)$ is the vector of filter taps at time n and where $\Delta\mathbf{h}(n)$ is the gradient used to update the filter. Everywhere in this paper boldface denotes vectors whereas boldface capitals denote matrices. The gradient is different for each algorithm but, in all cases, should ensure that $\hat{\mathbf{h}}(n)$ converges to the optimal Wiener solution \mathbf{h}_{opt} after sufficient iterations. Only the barest of details for each approach considered are given below as full details can be found in the open literature [7].

Least Mean Square (LMS): The LMS filter update $\Delta\mathbf{h}(n)$ is equal to $\mu\mathbf{x}(n)e(n)$, where μ is a scalar or step size which aims to control the rate of adaptation (and hence convergence/divergence), $\mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-L+1)]^T$ is the input vector of the filter and L is the filter length (256 for all algorithms used here).

Normalized-LMS (NLMS): The NLMS algorithm uses a normalized step size μ . Here the update $\Delta\mathbf{h}(n)$ is equal to $\frac{\mu}{\|\mathbf{x}(n)\|^2}\mathbf{x}(n)e(n)$.

Affine Projection Algorithm (APA): The update $\Delta\mathbf{h}(n)$ is here given by $\mu\mathbf{X}(n)[\mathbf{X}^T(n)\mathbf{X}(n) + \epsilon\mathbf{I}_N]^{-1}\mathbf{e}(n)$ where $\mathbf{X}(n) = [\mathbf{x}(n)\mathbf{x}(n-1)\dots\mathbf{x}(n-N+1)]$, an $L \times N$ matrix. L is the length of the filter, N is the order of the APA, \mathbf{I}_N is the identity matrix and $\mathbf{e}(n)$ is now a vector. For all experiments represented here we use $N = 2$.

Frequency Block-LMS (FBLMS): FBLMS is an implementation of a block-by-block LMS using fast convolution. In the time domain the update $\Delta\mathbf{h}(n)$ is given by $\mu\sum_{m=0}^{B-1}e(nB+m)\mathbf{x}(nB+m)$ where n is now a block index, m is the block sample index and B is the block length. We use $B = 256$.

III. EXPERIMENTAL WORK

A. Non-linear model

The direct comparison of AEC performance with (i) linear echo and (ii) linear and non-linear echo is made with otherwise identical conditions and, for all experiments reported here, both linear and non-linear echo is added artificially. In practice, non-linearities are introduced by component imperfections, i.e. from the miniaturization of components, and can be divided into two groups: those which arise in the downlink path and those which arise in the uplink path. Previous work [4] has shown that non-linearities coming from the loudspeaker and amplifiers in the downlink path dominate those in the uplink path due to the fact that microphone and uplink

amplifiers generally operate on lower-level signals. As in [6] a third-order polynomial model is used here to simulate non-linearities and is an approximation to the Volterra model. The output of the loudspeaker is given by $x_{nl}(n) = x(n) + \alpha x^2(n) + \beta x^3(n)$, where $x(n)$ is the far-end signal and $\alpha x^2(n) + \beta x^3(n)$ are the non-linear components introduced by the downlink loudspeaker and amplifiers. The parameters α and β are used to control the relative levels of second and third order non-linear distortions. A full description of this setup is given in our previous article [6].

B. Test set-up and metrics

We report different tests on each of the adaptive filters and compare the effects of non-linearities and white noise. The assessment is based on ERLE, convergence time and system distance. A 60-second speech signal is used as the far-end signal $x(n)$ and is sufficiently long to ensure the convergence of each algorithm. In all cases ERLE measurements relate to intervals in which the algorithms are deemed to have converged. Non-linear artefacts are introduced into the downlink signal according to the model described in Section III-A. The loudspeaker output is composed of the original speech signal $x(n)$ and a non-linear component which are both convolved with the echo path $h(n)$. This leads to a linear echo component $x(n) * h(n)$ and a non-linear echo component $[\alpha x^2(n) + \beta x^3(n)] * h(n)$. Then, the mean linear echo to non-linear echo ratio ($SNeR$) is computed as in [8]:

$$SNeR = \frac{1}{K} \sum_{i=1}^K SNeR_{seg}(i), \quad (2)$$

where the segmental $SNeR_{seg}(i)$ is given by:

$$SNeR_{seg}(i) = 10 \log_{10} \frac{\sum_{m=0}^{M-1} d_i^2(n)}{\sum_{m=0}^{M-1} d_{nl,i}^2(n)} \quad (3)$$

and where $d_i(n)$ and $d_{nl,i}(n)$ are the linear and non-linear echo components respectively in the i^{th} segment of analysed signals. The $SNeR_{seg}(i)$ is computed using windows of $32ms$ ($M = 256$ for a sampling rate of $8kHz$) over which period speech is relatively stationary. A second noisy signal with linear echo is also generated, where the mean SNR is equal to the mean $SNeR$. In so doing we have two linear echo signals that are equally distorted, one with non-linear echo, and another with additive noise. The weighting factors α and β , which are used to control the level of non-linear echo, are in the range of $[0, 1]$ as in [6]. The same step size μ is used for all algorithms which are furthermore configured to give the same amount of steady state ERLE in linear conditions (no non-linear echo and no noise component). This condition is very difficult to satisfy, as the different algorithms have different convergence rates, and is here achieved by tuning the regularization factor (generally used to avoid dividing by zeros during normalization) in each case. The LMS algorithm, however, converges too slowly and it was thus necessary to relax the constraints on the step size in this case. The APA, FBLMS, and NLMS algorithms obtain $ERLEs$ of $\sim 110dB$ in linear echo conditions. LMS does not perform sufficiently well and gives an ERLE of $\sim 80dB$.

C. Echo Return Loss Enhancement

Figure 2 shows the maximum ERLE achieved by each algorithm in non-linear and noisy environments as a function of SNR or $SNeR$. The maximum ERLE is the mean ERLE obtained during a 10 second (50-60s) period where each algorithm has converged. We observe that whatever the perturbation (non-linear echo or noise) performance decreases for all adaptive filters. In non-linear environments the

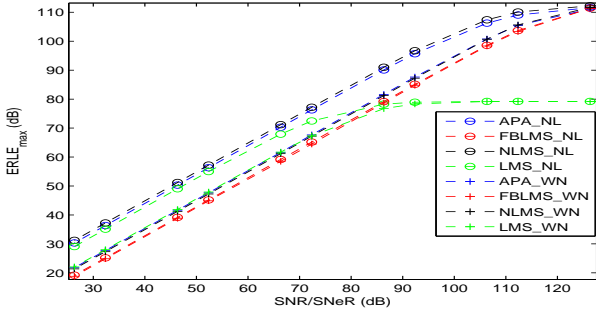


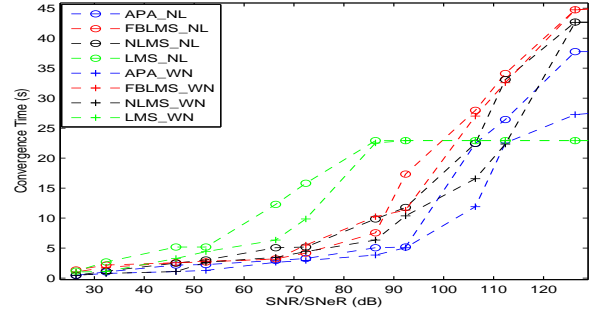
Fig. 2. Maximum ERLE (in dB) achieved after convergence as a function of $SNR/SNeR$ (also in dB). Here the SNR or $SNeR$ corresponds to added white noise (WN) or to non-linear echo (NL) as indicated. Profiles are illustrated for both perturbations and for each of the four approaches to AEC. APA, FBLMS and NLMS are all configured to give equivalent performance under linear echo conditions.

APA and NLMS algorithms show similar behaviour; decreases of approximately $80dB$ are observed between the linear echo conditions (right side of Figure 2) and non-linear echo conditions (left side of Figure 2). This shows the sensitivity of linear adaptive filters as, in this range, the non-linearities are inaudible. The FBLMS algorithm is the most affected. Performance decreases by about $90dB$ over the same range and for $SNeRs$ less than $75dB$ performance is worse than that for the standard LMS algorithm. This is explained in [6] as the effect of block-by-block processing which is more susceptible to non-linear effects than a sample-by-sample process. We see that the LMS algorithm is the most robust of all adaptive filters considered; it has the least degradation in performance as the SNR or $SNeR$ decreases. This is due to poor ERLE performance which is so low that the algorithm cannot even be configured to give equivalent performance to the other algorithms under linear echo conditions.

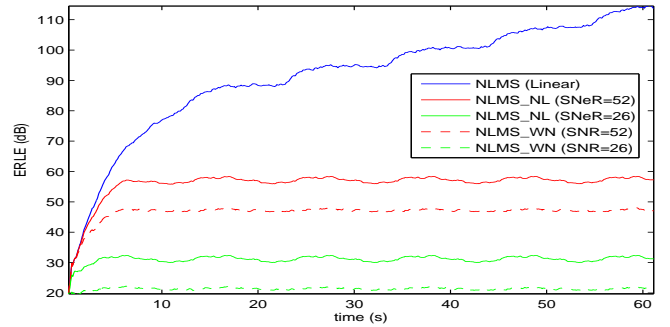
In noisy environments the performance of APA, NLMS and FBLMS algorithms decreases by approximately the same amount. For the APA and NLMS algorithms, and when the $SNR < 100dB$, the difference between the ERLE in non-linear and noisy environments is about $10dB$ for all algorithms, with better performance in non-linear environments than noisy environments. Differences in performance for non-linear and noisy environments are smallest for the FBLMS algorithm. This can again be explained by the averaging effect of block-by-block approaches. In the case of noise the perturbation is effectively averaged over the block and thus has a reduced impact on performance. This is not the case with non-linear echo, which is correlated with the input signal. In summary these results show that white noise perturbations have more effect than non-linearities on all algorithms except for the FBLMS where the performances are similar.

The difference between the effects of non-linearities and those of noise are explained by two hypotheses:

Noise spectrum: The filter frequency response depends on the differences in energy of the linear echo component and the perturbation (non-linear echo or noise). The spectrum of the non-linear echo component generally has a similar profile to that of the linear echo component whereas the white noise spectrum is flat. This means that during periods of voiced speech the amplitude of the noise signal can be much lower than the amplitude of the speech signal at low frequencies, but much higher at high frequencies. At higher frequencies the linear echo component can thus be masked by the noise spectrum, leading to significant perturbation during periods of



(a)



(b)

Fig. 3. Convergence performance with non-linear (NL) and white noise (WN) perturbations for (a) APA, FBLMS, NLMS and LMS algorithms plotted as convergence time against SNR , and (b) the NLMS algorithm plotted as ERLE against time.

voiced speech.

Non-linearities are correlated with the far-end signal: Since non-linearities are correlated with the input signal, this can result in the adaptive filter under-estimating the linear part but slightly attenuating the non-linearities. This is less so the case for noisy environments as there is no correlation between the noise and the far-end speech signal.

D. Convergence Time

The convergence time for each algorithm is computed as given in [6], and is defined as the time needed for each adaptive filter to reach 95% of its maximum ERLE value. Convergence times are determined using the same speech signals as used previously and are estimated for both conditions: linear echo with non-linear echo, and linear echo with noise. Figure 3(a) shows the convergence time in seconds against $SNR/SNeR$ for each of the four algorithms and both perturbations.

We see that, with the exception of the LMS algorithm, all profiles have a similar trend even though differences in convergence time are in the order of $25s$ at $110dB$. In addition, for each algorithm, convergence times are greater for non-linear perturbations than they are for noise. The LMS algorithm is the slowest to converge where the $SNeR$ or SNR is low but the fastest where they are high ($> 100dB$). This is explained by the fact that the ERLE obtained is lower: about $80dB$ compared to $110dB$ for all other algorithms in linear echo conditions (right side of Figure 2). We remark that, in all cases, the more the perturbations increase the lower the convergence time, since the ERLE obtained is lower.

The plots in Figure 3(a) show the absolute convergence time in

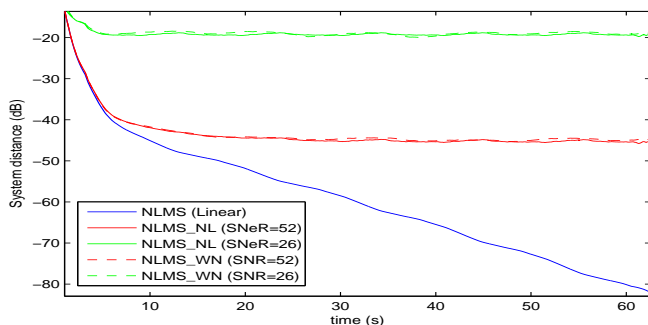


Fig. 4. Plots of system distance (in dB) against time (in seconds) for the NLMS algorithm. Profiles are illustrated for linear echo and also for linear echo with either non-linear echo (NL) or added white noise (WN) at two different levels.

seconds but do not give an impression of the dynamic performance and neither do they reflect the ERLE that is eventually achieved. They are thus potentially misleading and for this reason we present in Figure 3(b) a plot of ERLE against time (here for the NLMS algorithm only) to better illustrate the dynamic and absolute performance. Figure 3(b) shows the ERLE against time with linear echo only and added non-linear echo or noise at 52 and 26dB.

These plots show that higher levels of perturbation result in lower levels of ERLE. In the case of linear echo (top profile) convergence is slow and is not even reached during the 60s illustrated. Crucially, though, the ERLE is much higher than it is for non-linear and noise perturbations. However, in these cases the algorithm converges faster, but to a lower level (i.e. $\sim 55dB$ for non-linear echo with an $SNeR$ of 52dB and $\sim 20dB$ at 26dB $SNeR$, cf. $\sim 45dB$ for noise with an SNR of 52dB and $\sim 25dB$ at 26dB SNR). Hence consideration of the convergence time or maximum obtained ERLE are not sufficient on their own to properly appreciate performance. Similar profiles were obtained for the other adaptive filters and show an identical trend to that shown here for the NLMS algorithm albeit to different levels of ERLE. Finally, since all algorithms are shown to converge reasonably quickly in noise and non-linear environments it is of questionable advantage to focus effort on more computationally efficient algorithms; efforts are better directed toward the development of more robust algorithms. Indeed, more stable and straight forward algorithms, such as NLMS, are arguably of more interest for mobile terminal applications than their less stable and more computationally demanding alternatives.

E. System distance

The assessment of performance with linear echo is commonly measured according to the system distance which is measured according to $10\log_{10}[\frac{|h(n) - \hat{h}(n)|^2}{|h(n)|^2}]$. The system distance indicates the accuracy of $\hat{h}(n)$. It is less appropriate in the case of non-linear echo as the system distance shows only how well the linear echo path is estimated and thus does not necessarily reflect the level of echo attenuation actually achieved. Figure 4 shows the behaviour of the NLMS system distance as a function of time. Whilst there are differences in exact values of system distance, the order of the profiles and general trends are indicative of performance for all the other filters. In general, the better the system distance, the better the ERLE. However, upon comparison of Figures 3(b) and Figure 4 we observe an apparent disparity. Figure 3(b) shows that performance with non-linear echo is generally better than that under additive noise with the same SNR , whereas Figure 4 shows almost no differences. This is due to the fact that system distance is only equivalent to

ERLE under the condition of total linearity. The ERLE reflects the global performance according to the residual error, whereas the system distance reflects the accuracy of $\hat{h}(n)$. Equivalent values of system distance show that linear echo can be attenuated equally well with either non-linear echo or noise perturbations. The differences in the ERLE, however, show that non-linear echo perturbations are better attenuated than noise. This is due to the fact that in non-linear environments some of the non-linearities are indeed effectively attenuated by the adaptive filter even if the residual error is still higher than in the linear situation. This is due to the fact that adaptive filters aim to reduce the correlation (increase the orthogonality) between the error and the input signal. Since non-linear echo is correlated with the input signal it can also be attenuated, albeit only slightly. This is not the case with additive noise. This does not imply that adaptive filters are better in non-linear environments than they are in noisy environments as the adaptive filter does not aim to reduce the noise, but rather the echo signal which includes the non-linear component.

IV. CONCLUSIONS

This paper presents a new comparison of the effects of non-linearities and noise on four adaptive filters. Experimental results show that APA and NLMS have comparable behaviour in non-linear environments whereas FBLMS is badly affected. In noisy environments, however, there is little difference between each approach and, being less computationally demanding than the other approaches, FBLMS is an appealing solution in this case. Results confirm that performance decreases as the level of perturbations increase but that echo cancellation seems to be more robust to non-linearities than noise with a similar SNR (with the exception of the FBLMS algorithm). We show that the linear component of the echo path is under estimated but is as accurate in the case of non-linear echo as it is in noisy environments, again with a similar SNR . In addition, as the non-linear component is correlated with the far-end signal a fraction of non-linearities are effectively attenuated. Noise, in contrast, is neither correlated, nor attenuated.

This paper demonstrates the different nature of non-linear echo and additive noise and their effects on adaptive filtering performance. Whereas the effects of additive noise are well understood there are relatively few studies on the effects of non-linear echo in the existing literature. This contribution sheds new light on the problem and supports continuing efforts to tackle non-linear echo.

REFERENCES

- [1] E. Hänsler and G. Schmidt, *Acoustic Echo and Noise Control: A Practical Approach*. Address: Wiley-IEEE Press, 2004.
- [2] M. H. Costa and J. C. M. Bermudez, "A noise resilient variable step-size LMS algorithm," *Signal Processing*, vol. 88, pp. 733–748, 2008.
- [3] Y. Zhang, J. A. Chambers, W. Wang, P. Kendrick, and T. J. Cox, "A new variable step-size LMS algorithm with robustness to nonstationary noise," in *ICASSP 2007*, Honolulu, US, April 4 2007, pp. 257 – 260.
- [4] A. Stenger, L. Trautmann, and R. Rabenstein, "Nonlinear acoustic echo cancellation with 2nd order adaptive Volterra filters," in *ICASSP 1999*, Phoenix, AZ, USA, March 15-19, 1999, pp. 877–880.
- [5] K. Shi, X. Ma and G. T. Zhou, "A residual echo suppression technique for systems with non-linear acoustic echo paths," in *ICASSP 2008*, Las Vegas, Nevada, USA, April 1-4, 2008, pp. 257–260.
- [6] M. I. Mossi, N. W. D. Evans, and C. Beaugrant, "An assessment of linear adaptive filter performance with non-linear distortions," in *ICASSP 2010*, Dallas, Texas, March 14-19, 2010.
- [7] S. Haykin, *Adaptive Filter Theory 4th Ed.* Address: Prentice Hall, 2001.
- [8] M. Vondrasek, P. Pollak, "Methods for Speech SNR Estimation: Evaluation Tool and Analysis of VAD Dependency," *RADIOENGINEERING -PRAGUE-*, vol. 1, pp. 11–25, Jan. 1999.