EXCITATION-DEPENDENT STEPSIZE CONTROL OF ADAPTIVE VOLTERRA FILTERS FOR ACOUSTIC ECHO CANCELLATION

Cristian Contan[1], Marcus Zeller[2], Walter Kellermann[2] and Marina Topa[1]

[1]Bases of Electronics Department,
Technical University of Cluj-Napoca, Romania
{cristian.contan,marina.topa}@bel.utcluj.ro

[2]Multimedia Communications and Signal Processing,
University of Erlangen-Nuremberg, Germany
{zeller,wk}@LNT.de

ABSTRACT

This paper proposes a new updating technique for adaptive Volterra kernels employed as nonlinear models for the identification of unknown feedback paths in acoustic nonlinear echo cancellation tasks. Considering that nonlinear distortions are mainly introduced for high input levels, the effective step size is shaped according to the instantaneous envelope of the excitation signal. The loudspeaker-enclosure-microphone setup is modeled using measured linear and quadratic Volterra kernels. The efficiency of the proposed method is compared to the one of the second-order Volterra filter that uses only the Normalized Least-Mean-Square algorithm for kernel adaptation in terms of Echo Return Loss Enhancement. Simulations for different input signals show a superior behavior of the new approach in terms of convergence speed, considering the same steady-state error.

Index Terms— Volterra filters, adaptive algorithms, acoustic echo cancellation, nonlinear distortions

1. INTRODUCTION

Acoustic echo cancellation techniques are mainly used in speech processing systems such as, e.g., mobile phones, hands-free devices or in teleconferencing setups. In these applications, the communication should be facilitated by removing acoustic echoes that result from the feedback of loudspeaker signals into microphones [1, 2]. The basic structure of an acoustic echo cancellation (AEC) scenario is presented in Fig. 1. Here, the two dashed boxes indicate both the loudspeaker-enclosure-microphone (LEM) setup that includes the acoustic hardware from the enclosure together with the audio signal and local noise levels and the adaptive filter. The role of the adaptive filter is to model the LEM path as accurately as possible. The common AEC works under the assumption that the LEM system can be sufficiently well described by a linear adaptive model [3]. Regarding the very small and often low-cost hardware components in recent mobile devices, however, this assumption will not always hold in practice since their nonlinear behavior cannot be accounted for.

This work has been carried out while Cristian Contan was a visiting researcher at the Chair of Multimedia Communications and Signal Processing in Erlangen, Germany. This paper was supported by the project "Doctoral studies in engineering sciences for developing the knowledge based society-SIDOC" contract no. POSDRU/88/1.5/S/60078, project co-funded from European Social Fund through Sectorial Operational Program Human Resources 2007-2013.

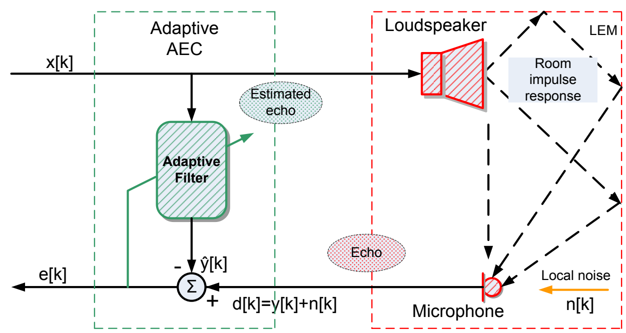

Fig. 1. Acoustic echo cancellation setup.

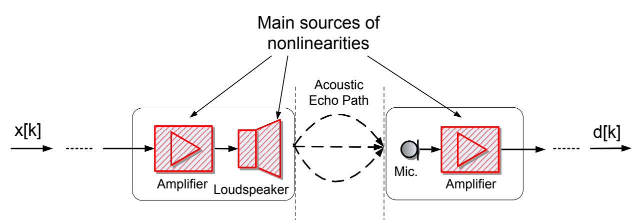

Fig. 2. Nonlinear echo path.

The main sources of nonlinearities along the acoustic echo path are illustrated in Fig. 2. One source of noticeable nonlinearities is an overdriven amplifier which shows a memoryless saturation characteristic as described in [4]. Another source is a small loudspeaker driven at high volume which causes nonlinear distortions with memory as discussed in [5]. Various structures for appropriate nonlinear AEC have been considered so far, such as adaptive Wiener and Volterra models in [6] or adaptive power filters in [7]. In this work we focus on a new nonlinear echo cancellation procedure that relies on the application of adaptive Volterra filters while also taking the excitation level into account.

The paper is organized as follows: Sec. 2 describes the main features of the second-order Volterra structure along with the new adaptation procedure, while in Sec. 3 the simulation results are presented. Conclusions are drawn in Sec. 4.

2. PROPOSED STRUCTURE

Assuming that the complete input/output characteristics of the LEM setup can be modeled sufficiently well by an N-th-order truncated Volterra series, relation (1) can be used for this purpose [6]:

$$y[k] \approx y_{NVF}[k] = \sum_{p=1}^{N} \sum_{m_1=0}^{M_p-1} \dots \sum_{m_p=m_{p-1}}^{M_p-1} \hat{h}_p[m_1, \dots, m_p] x[k-m_1] \cdot \ldots \cdot x[k-m_p], \quad (1)$$

where $x[k]$ is the discrete input signal, $y_{NVF}[k]$ is the output signal of the Volterra structure while M_p represents the memory length of each Volterra kernel. The p-th order Volterra kernel $\hat{h}_p[m_1, \dots, m_p]$ exhibits a general symmetry ($\hat{h}_p[m_1, \dots, m_p] = \hat{h}_p[m_p, \dots, m_1]$) that is reflected in (1): only terms with non-decreasing indices ($m_p \geq m_{p-1}$) are considered. The accuracy of the Volterra filter for identifying the nonlinear distortions found in the LEM structure will generally increase with the order and the memory size of the filter at the expense of increased computational complexity. As related in [8], the second-order Volterra filter is adequate to incorporate the type of nonlinearities encountered in the LEM setup without excessive computational cost. The output of a second-order Volterra filter can be written as the sum of the two kernel outputs $y_1[k]$ and $y_2[k]$:

$$y_{2VF}[k] = y_1[k] + y_2[k], \quad (2)$$

where the two kernel outputs are computed as:

$$\begin{cases} y_1[k] = \hat{\mathbf{h}}_1^T[k]\mathbf{x}_1[k] = \sum_{m=0}^{M_1-1} \hat{h}_1[m] x[k-m] \\ y_2[k] = \hat{\mathbf{h}}_2^T[k]\mathbf{x}_2[k] = \sum_{m_1=0}^{M_2-1} \sum_{m_2=m_1}^{M_2-1} \hat{h}_2[m_1, m_2] \underbrace{x[k-m_1] x[k-m_2]}_{\text{products of input samples}} \end{cases} \quad (3)$$

using the following vector definitions:

$$\mathbf{x}_1[k] = (x[k], x[k-1], \dots, x[k-M_1+1])^T;$$
$$\hat{\mathbf{h}}_1[k] = (\hat{h}_1[0], \hat{h}_1[1], \dots, \hat{h}_1[M_1-1])^T;$$
$$\mathbf{x}_2[k] = (x^2[k], x[k]x[k-1], \dots, x[k]x[k-M_2+1], x^2[k-1], x[k-1]x[k-2], \dots, x^2[k-M_2+1])^T;$$
$$\hat{\mathbf{h}}_2[k] = (\hat{h}_2[0,0], \hat{h}_2[0,1], \dots, \hat{h}_2[0, M_2-1], \hat{h}_2[1,1], \hat{h}_2[1,2], \dots \dots, \hat{h}_2[M_2-1, M_2-1])^T.$$

The number of terms embedded in the kernel vectors is determined by the order and the memory length of the actual filter as follows:

$$length(\hat{\mathbf{h}}_p) = \frac{(M_p + p - 1)!}{(M_p - 1)!\, p!}. \quad (4)$$

When dealing with second-order Volterra filters for nonlinear acoustic echo cancellation, the goal is to optimally approximate the distortions caused by the electroacoustic converters. Using this approach, the LEM characteristics are to be modeled by the linear and quadratic Volterra kernels starting from some initial values. The Volterra kernels are updated at each iteration of the input samples using an adaptive algorithm in order to minimize the residual error signal $e[k]$ computed as:

$$e[k] = d[k] - y_{2VF}[k]. \quad (5)$$

One of the most widely used adaptive filtering algorithms in practical error reduction scenarios is the Normalized Least-Mean-Square (NLMS) algorithm which takes into account the variation of the input signal statistics. In the following set of equations, the NLMS adaptation algorithm [9] is summarized for the linear and the second-order Volterra kernels:

$$\begin{cases} \hat{\mathbf{h}}_1[k+1] = \hat{\mathbf{h}}_1[k] + \dfrac{\mu_1}{\mathbf{x}_1^T[k]\mathbf{x}_1[k] + \varphi} e[k]\mathbf{x}_1[k], \\ \hat{\mathbf{h}}_2[k+1] = \hat{\mathbf{h}}_2[k] + \dfrac{\mu_2}{\mathbf{x}_2^T[k]\mathbf{x}_2[k] + \varphi} e[k]\mathbf{x}_2[k], \end{cases} \quad (6)$$

where φ is a constant introduced to prevent division by zero. The step-size parameters μ_1 and μ_2 ($\mu_1 > \mu_2$) are constants that are chosen from the range (0, 2) so as to guarantee a stable convergence in the mean [3]. The major disadvantage of selecting constant step-size parameters is the resulting tradeoff between the speed of convergence and the steady-state error [10]. If a small step size is chosen, a reduced steady-state error will be obtained at the cost of a slower convergence. Although the convergence can be boosted by using a larger step-size value, the steady-state error will significantly increase.

The method proposed here to alleviate the above tradeoff consists in developing a new step size control that depends not only on the value of μ but also on the short-term envelope of the input signal. This dependency on the input signal is proposed because we want to increase the adaptation for the nonlinear kernel when high levels of the input are encountered. On the other hand, for small amplitudes of the excitation signal, small adaptation steps should be used for the quadratic kernel coefficients under the assumption that the LEM behaves mostly linear in these situations.

The proposed excitation-dependent step-size control (ESC) is then given by (7) which represents the update equation of the second-order Volterra kernel:

$$\hat{\mathbf{h}}_2[k+1] = \hat{\mathbf{h}}_2[k] + \alpha |x[k]|^{\beta} \frac{\mu_2}{\mathbf{x}_2^T[k]\mathbf{x}_2[k] + \varphi} e[k]\mathbf{x}_2[k], \quad (7)$$

while the update of the linear kernel is still carried out as given by (6).

In (7), the parameter α is computed as $\alpha := \frac{1}{|x_0|^{\beta}}$, where x_0 is set as a threshold for the input samples and β is selected for choosing the shape of the step-size function. The effective value of the new step size $\mu_{2,\text{eff}} = \mu_2 \alpha |x[k]|^{\beta}$ depending on the input samples is depicted in Fig. 3 for $\mu_2 = 1$. The curves for the following values of β ({0,5; 1; 2}) are illustrated in comparison to the standard NLMS (with $\mu_2 = 1$) for the chosen reference $x_0 = 0.5$. For input values larger than x_0 ($x[k] \geq x_0$) the obtained step size is larger than μ_2 in case of a usual NLMS-type adaptation according to (6). In each of the selected β cases this will speed up the convergence of the Volterra filter in the upper range of the input samples. For the lower range of input samples the ESC structure uses the common NLMS algorithm to update the second-order kernel as in (6) (using only μ_2 as step size). The convergence speed has been found to be too slow if we follow the curve of the effective step size with values smaller than μ_2. This is a consequence of the large number of quadratic coefficients to be identified. The ESC structure is depicted in Fig. 4.

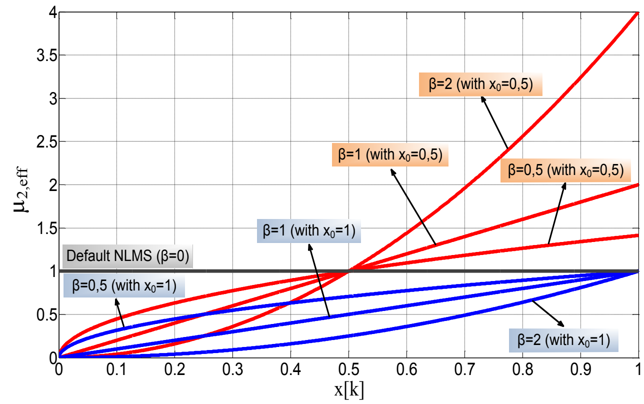

Fig. 3. Behavior of the effective step size used for the adaptation of the quadratic kernel coefficients.

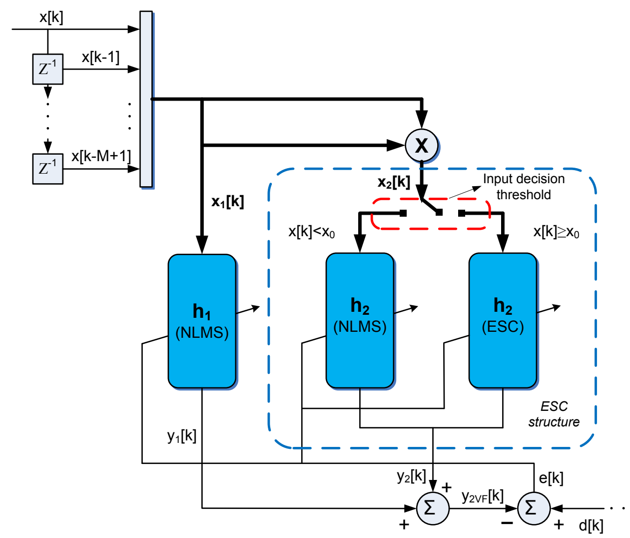

Fig. 4. Second-order Volterra structure using the ESC adaptation technique.

3. EXPERIMENTAL RESULTS

In the following section the second-order ESC Volterra method is tested in nonlinear echo cancellation setups in order to compare the speed of convergence of the new structure and the steady-state control of the error with the achieved echo reduction evolution of the second-order NLMS Volterra filter. The nonlinear LEM setup is modeled as a second-order polynomial written in Volterra form by outlining the linear and the nonlinear components of the microphone signal $d[k]$. Local noise $n[k]$ is added to the reference signal in addition to the nonlinear distortions. The Volterra expression of the LEM's output follows the model as in [11]:

$$d[k] = \underbrace{\mathbf{h}_1[k]\mathbf{x}_1^T[k]}_{\text{linear}} + \gamma[k]\underbrace{\mathbf{h}_2[k]\mathbf{x}_2^T[k]}_{\text{nonlinear}} + \delta[k]\underbrace{n[k]}_{\text{local noise}}, \qquad (8)$$

where $\mathbf{h}_1[k]$ and $\mathbf{h}_2[k]$ represent the linear and the quadratic kernels defined in vector form as in (3) that are measured from low-cost acoustic components in a low-reverberant room. The linear and the quadratic kernels employed for the design of the LEM structure are depicted in Fig. 5 and have been obtained from measurements at a sampling rate of 8 kHz, with memory lengths equal to 320 and 64x64 taps, respectively, so that they include all the coefficients with significant nonzero values.

The levels of the so–called Linear–to–Nonlinear Ratio (LNLR) and Signal-to-Noise Ratio (SNR) are kept constant throughout the simulations with the help of parameters $\gamma[k]$ and $\delta[k]$. The LNLR is defined in this case as the ratio between the power of the microphone signal's linear component and the power of its nonlinear component. The SNR is defined as the ratio of the reference signal power in the absence of local noise to the local noise power. The argument $\gamma[k]$ is set to keep the LNLR constant during simulations at 10 dB and also $\delta[k]$ is selected properly to set the SNR at 30 dB. For the input, signals with different probability density functions are chosen: white Gaussian noise and nonstationary audio signals. Additive white Gaussian noise is used as local noise. In experiments the memory lengths of the Volterra filter kernels are selected to match those of the LEM system.

The difference in the performance of the ESC and NLMS second-order Volterra structures is evaluated in terms of convergence rate and steady-state error by monitoring the Echo Return Loss Enhancement (ERLE)

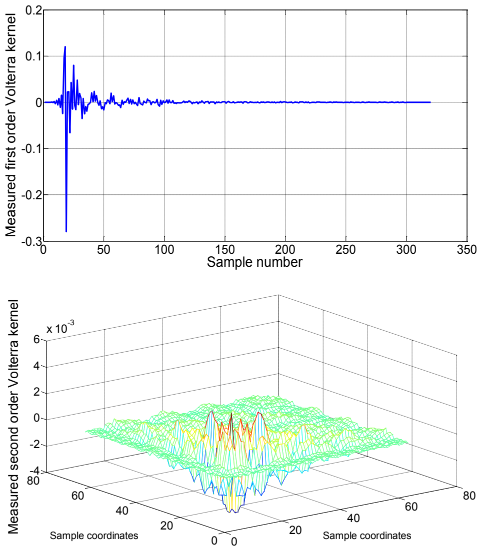

Fig. 5. Linear kernel (top); quadratic kernel (bottom).

quantity defined as:

$$\mathrm{ERLE} = 10\log_{10}\frac{\mathrm{E}\{d[k]^2\}}{\mathrm{E}\{e[k]^2\}} \quad [\mathrm{dB}], \tag{9}$$

where $\mathrm{E}\{\cdot\}$ denotes statistical expectation.

In the first set of simulations white Gaussian noise is used as input signal. The step-size parameters for the ESC and NLMS Volterra structures are chosen $\mu_1 = 0.1$ for the linear kernel and $\mu_2 = 0.05$ for updating the quadratic kernel. The value of φ is set 0.1 for input samples in the range [-1; 1]. Regarding the modeling of the ESC Volterra filter's new step-size function and the input signal's threshold above which the proposed method can be validated, the following values have been chosen: $\beta = 0.5$ and $x_0 = 0.1$. The evolution of ERLE for both nonlinear adaptive filters and for the linear filter can be observed in Fig. 6. The usage of only an adaptive NLMS linear filter for nonlinear acoustic echo cancellation involving the mentioned enclosure setup offers a high convergence speed, although the ERLE saturates at the established LNLR value of 10 dB. In contrast, for the two adaptive Volterra implementations the ERLE settles at 30 dB in each case, which is the SNR value. The difference between these two Volterra filter versions is obviously an improved convergence rate when using the ESC design without affecting the steady-state error too much.

For the results depicted in Fig. 7 the same LEM structure is used, but a steeper evolution of the ESC step size is implemented ($\beta = 2$). By using this value for the exponent β , the samples of the input signal above the mentioned threshold are emphasized even more, so that it will eventually affect the ESC structure's step size, offering a better convergence for the high-amplitude input samples. We have to consider that for large values of $\alpha|x[k]|^{\beta}$, the newly formed step size $\mu_{2,\mathrm{eff}}$ that uses the value of μ_2 can exceed the range (0; 2) such that a stable convergence in the mean of the Volterra filter cannot be guaranteed anymore. In this case the value of μ_1 is reduced to 0.01 and for μ_2 to 0.002 to form the new step size in the normal range of adaptation. The input threshold is set again to $x_0 = 0.1$. By applying the specified step-size parameters μ_1 and μ_2 , the convergence rate of the NLMS linear and second-order Volterra filters will drop in comparison to the previous example. However, when exploiting the new value of the parameter β , the difference between the convergence rates of ERLE for the two adaptive second-order Volterra filters increases; the method that uses the ESC adaptation procedure converges faster than the classic NLMS approach for updating the Volterra kernels. Again the linear NLMS filter settles at 10 dB ERLE.

In the next simulations different audio signals in the range (-1; 1) are used as input. To design the ESC's input-conditioned step size the following situation is proposed in order to benefit from the high valued samples of the input signal using $\beta = 2$ and $x_0 = 0.05$. The coefficients μ_1 and μ_2 are set to 0.1 and 0.005. Fig. 8 shows the ERLE evolution for the proposed methods and the signal intercepted by the microphone which is a nonlinearly distorted speech. In the following cases indicated in Fig. 9 and Fig. 10, instrumental music and a song fragment are used as excitations.

As it can be seen in all three cases with realistic, nonstationary input signals, the acoustic cancellation method that adapts its second-order Volterra kernel using the ESC procedure provides a better convergence of ERLE in comparison to the second-order NLMS Volterra filter. The two ERLE characteristics stabilize at a discharged echo amount of 30 dB which indicates the chosen SNR value. Moreover, the NLMS linear adaptive filter reaches a steady-state ERLE of approximately 10 dB, equal to the LNLR value.

4. CONCLUSIONS

In this work a novel nonlinear excitation-dependent step-size control for the adaptation of an adaptive second-order Volterra filter was proposed in order to enhance the convergence of the nonlinear model parts in case of high input levels. To achieve this improvement, an amplitude threshold for the input signal was selected such only excitation samples above this threshold will lead to increased adaptation step-size of the second-order kernel. The proposed method was tested for a real loudspeaker-enclosure-microphone setup that includes measured linear and quadratic distortions with specific noise conditions and level of nonlinear distortions. The effectiveness of the new nonlinear adaptation method for acoustic echo cancellation was evaluated in terms of the

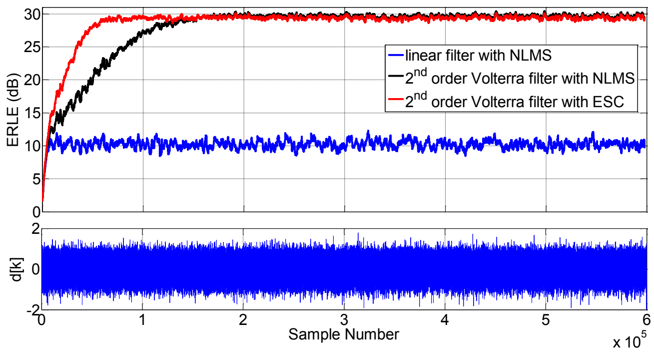

Fig. 6. ERLE evolution for white Gaussian noise with $\beta = 0.5$ and $x_0 = 0.1$ (top); microphone signal (bottom).

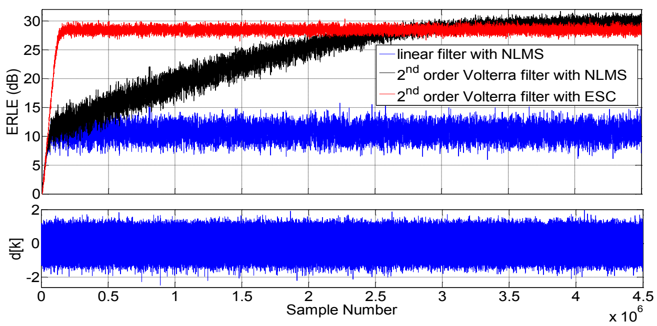

Fig. 7. ERLE evolution for white Gaussian noise with $\beta = 2$ and $x_0 = 0.1$ (top); microphone signal (bottom).

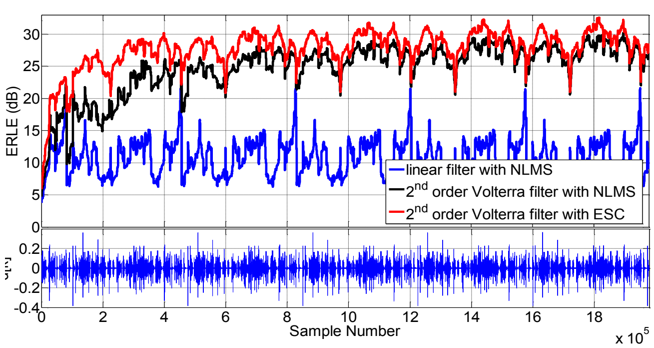

Fig. 8. ERLE evolution for speech as input (top); microphone signal (bottom).

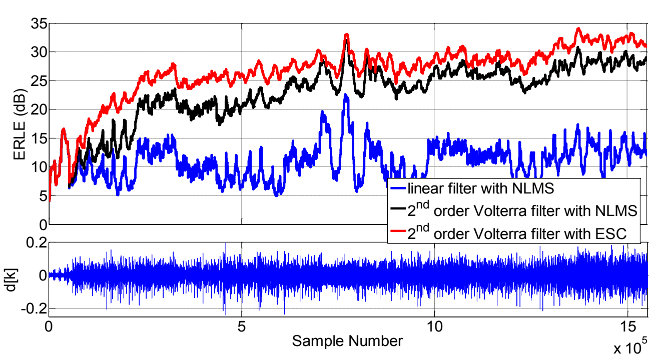

Fig. 9. ERLE evolution for instrumental music as input (top); microphone signal (bottom).

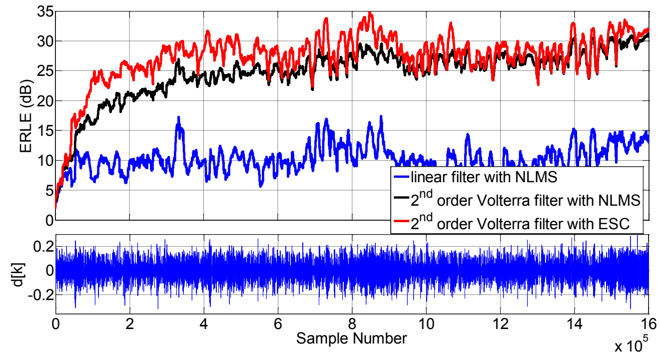

Fig. 10. ERLE evolution for a song fragment as input (top); microphone intercepted signal (bottom).

evolution of the residual error power in comparison to the second-order Volterra filter structure that uses the NLMS algorithm for adapting the kernels. The proposed structure has been shown to outperform the classic NLMS-type adaptation in terms of initial convergence. Simulations for white Gaussian noise and real nonstationary audio signals as excitations were provided.

One can conclude that the proposed low-complexity method can be applied successfully in nonlinear acoustic echo cancellation scenarios for different real LEM setups, effecting better performance for different types of input signals.

REFERENCES

[1] E. Hänsler, G. Schmidt, *Topics in Acoustic Echo and Noise Control,* Springer, 2010.

[2] F. A. Everest, *The Master Handbook of Acoustics*, fourth edition, McGraw-Hill, 2001.

[3] S. Haykin, *Adaptive Filter Theory*, Englewood Cliffs, NJ, USA, Prentice Hall, 1996.

[4] B. S. Nollet and D. L. Jones, ″Nonlinear Echo Cancellation for Hands-Free Speakerphones″, *Proc. NSIP'97*, Michigan USA, September 8-10, 1997.

[5] R. Martin et al., *Advances in Digital Speech Transmission*, Wiley, 2008.

[6] T. Ogunfunmi, *Adaptive Nonlinear System Identification: The Volterra and Wiener Model Approaches*, Springer, 2007.

[7] F. Kuech, A. Mitnacht, W. Kellermann, "Nonlinear acoustic echo cancellation using adaptive orthogonalized power filters", *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Proceedings,* vol. 3, pp.105-108, 2005.

[8] A. Guerin, G. Faucon, R. Le Bouquin-Jeannes, "Nonlinear acoustic echo cancellation based on Volterra filters", *IEEE Transactions on Speech and Audio Processing*, vol. 11, issue 6, pp.672-683, Nov. 2003.

[9] B. Farhang-Boroujeny, *Adaptive Filters Theory and Applications*, Wiley, Oct. 1998.

[10] C. Contan et al., "Nonlinear acoustic system identification using adaptive combinations of Volterra filters", *Acta Technica Napocensis*, Vol. 52, No. 2, pp. 48-53, 2011.

[11] L. A. Azpicueta-Ruiz et al., "Adaptive combination of Volterra kernels and its application to nonlinear acoustic echo cancellation:", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 19, no.1, pp.97-110, 2011.