VI Congreso Iberoamericano de Acústica - FIA 2008
Buenos Aires, 5, 6 y 7 de noviembre de 2008

FIA2008-A024

Different strategies for nonlinear acoustic echo cancellation

Sofia Alfonso < sofia.alfonso@gmail.com >,
Adolfo Altenberg < altenberg@gmail.com >,
Guillermo Sentoni [(a)] <gsentoni@gmail.com >.

(a) UNGS

Abstract

Acoustic echoes affect the sound quality and may hamper many hands-free communications, making acoustic echo cancellers critical for enhancing the audio quality. Designing them is a challenging issue because of long room impulse responses and nonlinearities present in the power amplifier and/or the loudspeaker. This work proposes a general nonlinear digital filter structure for nonlinear acoustic echo cancellation applications. It is constructed under the assumption that the nonlinearity in typical hands-free speakerphones is of a localized nature then followed by a linear room impulse response. By doing so, it is made upon of a nonlinear discrete dynamic DABNet model cascaded with a FIR filter. This DABNet model is able to approximate the nonlinearity present in the system to any extent, while the FIR deals with the room impulse response. Comparisons of the echo canceller implemented with the DABNet + FIR show a significant performance improvement against either the Neural Network + FIR, and the linear FIR echo canceller, while contrast with an adaptive Volterra scheme, still gives a better, but comparable, performance level.

Resumen

El eco acústico afecta la calidad de sonido y puede dificultar las comunicaciones de manos libres, haciendo de los canceladores de eco acústico una parte crítica para mejorar la calidad de audio. Su diseño es una cuestión que presenta desafíos dado el largo de la respuesta al impulso de la habitación y de las no-linealidades presentes en el amplificador de potencia y/o el altavoz. Este trabajo propone una estructura de filtro no lineal digital para la cancelación de eco acústico no lineal. Se basa en la hipótesis de que las no linealidades típicos de un equipo manos libres son de una naturaleza localizada en el parlante seguidas por una respuesta al impulso lineal de la habitación. De tal forma, se utiliza un modelo dinámico no lineal discreto DABNet en cascada con un filtro FIR. Este modelo es capaz de aproximar arbitrariamente bien la no linealidad presente en el sistema, mientras que el FIR aproxima la respuesta al impulso de la habitación. Las comparaciones de los canceladores de eco llevadas a cabo con DABNet + FIR muestran una significativa mejora de rendimiento, ya sea comparada con la red neuronal + FIR o con el cancelador de eco lineal. Se muestran comparaciones con un filtro de segundo orden de Volterra con rendimientos similares.

1 Introduction

This work is focused on approximation issues regarding acoustic echo cancellation in nonlinear channels. Applications in handsfree mobile telephony, teleconference and VoIP require high quality echo cancellation devices to provide the users with an adequate service. Traditionally, acoustic echo cancellation is approached as a problem of adaptive identification of a sound channel by means of a linear system. The large length of the room impulse response gives great complexity to the problem of adaptive identification, while the fast variations of the impulse response require fast convergence. Generally a widely proven adaptive FIR structure with an on-line NRLMS algorithm is used to model the acoustic channel. Under a linear channel assumption, it could be expected that the acoustic coupling between the loudspeaker and the microphone would be accurately modeled by means of an adaptive FIR model. However, this is hardly possible due to several factors that limit the achievable amount of echo cancellation. Those factors include, ambient noise, finite precision arithmetic effects, under-modeling of the room impulse response, loudspeaker non-linearity, dynamic tracking and double-talk. Moreover, most of the nonlinear behavior of the acoustic channel is due to the loudspeaker response. In fact, suspension nonlinearity produces a sizable distortion at low frequencies, while nonuniformity in the magnetic flux density causes high distortion levels for large amplitude output values (Gao; Snelgrove, 1991). The latter is a very common situation for hands-free mobile telephony systems. For them, the channel can be characterized as a nonlinear dynamic system representing the loudspeaker response, cascaded with a linear dynamic system representing the acoustic echo path (Figure 1-a).

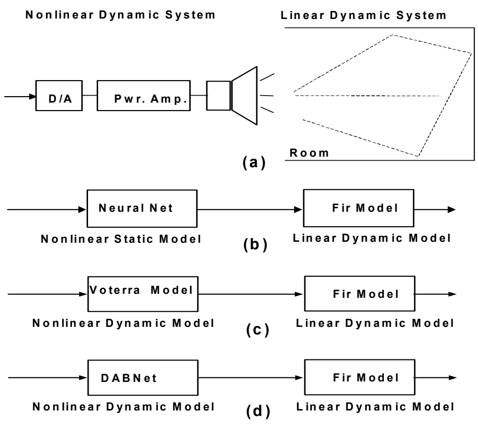

Figure 1: Systems and Models.

In this paper, we use a DABNet model (De-coupled A-B matrices Neural Network) to represent the loudspeaker dynamic non-linear behavior cascaded with a FIR to model the linear acoustic echo path (Figure 1-c). The linear dynamic stage of the DABNet model is initially spanned by a set of discrete Laguerre systems, whose states are nonlinearly combined by a single hidden layer Perceptron as can be seen in Figure 2. In previous works (Sentoni et al, 2001; Sentoni, 2003), it was shown that a nonlinear combination of discrete Laguerre systems is able to approximate any single-input nonlinear discrete system having fading memory (for fading memory concepts, see Matthews and Moschytz, 1994.). After the initial span of the linear layer by Laguerre systems, a model reduction technique is performed on the hidden nodes of the neural network as part of the identification process. The balancing is performed in such a way that the linear state space representation is de-coupled by blocks. Doing so, it is possible not only to identify the main time constants, but also to reduce the

dimensionality of the Perceptron input space. In addition, the noncontrollable/non-observable modes are removed from the model.
The final DABNet model consists of a sparse linear state space system, whose states, de-coupled by blocks, are mapped by a neural network. This de-coupled representation provides insight into the final model, especially since, in a practical identification scheme, the individual state space matrices can themselves be represented as loosely coupled first and second order sections. Once the DABNet model is trained, it is no further adapted and then, the proposed approximation structure can be completed by adding up a FIR model. Results for this structure show acoustic echo cancellation ERLE improvement of about 26 dB with respect of that of a linear FIR NRLMS structure. When compared to an adaptive Volterrea, the DABNet shows an improvement of about 9 *dB*. The model was contrasted using real speech data applied to a simulated nonlinear loudspeaker echo channel.

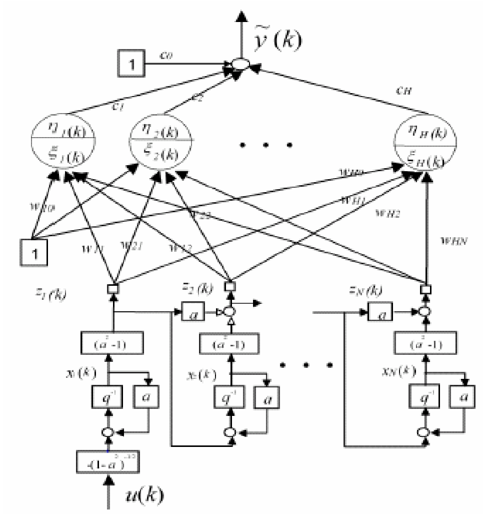

Figure 2: DABNet Model.

This paper is structured as follows. In section 2 we present the different nonlinear model structures. In section 3, we apply these structures in the equalization of a nonlinear channel. Finally, we present our conclusions and the future work arising from this study.

2 Nonlinear Models

This work proposes a nonlinear discrete approximation structure for nonlinear acoustic systems like the one depicted in Figure 1-a). In order to work with them, DABNet Model will be introduced and adaptive Volterra models will be described for comparisons purposes.

2.1 DABNet Models

This approximation structure is made upon the assumption that the nonlinearity in typical hands-free speakerphones is of a localized nature, followed by a linear room response. Therefore, the proposed structure is made of a DABNet model cascaded with a FIR filter (Figure 1-c). In this way, the DABNet model is able to approximate the nonlinearity present in the system to any extent, while the FIR deals with the linear room impulse response. The DABNet model holds for multi input single output systems but for clarity we will write the equations for single input single output (SISO) systems. The output of a DABNet model for a SISO case with n Laguerre systems generated in base to a pole a and with H hidden neurons, depicted in Figure 2, is expressed as:

$$y(k) = c^T \eta \tag{1}$$

The vector **c** is made of output weights:

$$c = [c_0, c_1 \ldots c_H]^T \quad \in \quad R^{H+1} \tag{2}$$

where **η** is the vector of outputs of the hidden layer

$$\eta = [1, \eta_1(k), \ldots, \eta_H(k)]^T, \qquad i: 1, \ldots, H \tag{3}$$

For any of the *H* hidden neurons, there exists a nonlinear sigmoidal mapping σ

$$\eta_i(k) = \sigma_i(k)(\xi i(k)), \qquad i: 1 \ldots H \tag{4}$$

applied to the pre-synaptic signals ξ_i *(k)* These signals are related to the outputs of Laguerre systems through the weights of the input layer $\tilde{w}_i$

$$\xi_i(k) = \omega_{i0} + w_i^T z(k), \qquad i = 1 \ldots H \tag{5}$$

The model is completed with the following description of the linear dynamic part, composed of a set of linear Laguerre systems whose states $x(k) = [\xi_1(k), \ldots, \xi_v(k)]^T$ for the generating pole a, are given by:

$$x_1(k+1) = a\, x_1\, (k) \;\; (1 - a^2)^{-1/2} \quad u\, (k) \tag{6}$$

$$\vdots$$

$$x_n(k+1) = a\, x_n(k) + z_{n-1}\, (k)$$

and their outputs:

$$z_1\, (k) = (a^2 - 1)\, x_1\, (k) \tag{7}$$

$$\vdots$$

$$z_n\, (k+1) = (a^2 - 1)\, x_{j+1}\, (k) + a\, z_j\, (k) \tag{8}$$

Let us now define a vector θ containing all of the parameters: the output weights c_i and the input weights w_i:

$$\theta = [c_0,\, c_1, \ldots,\, c_H, w_{10}, \ldots, w_{1n}, \ldots, w_{H0}, \ldots, w_{Hn}] \tag{9}$$

Using θ, the output of a system can be expressed as

$$y(k) = f(u(k); \theta; k) \tag{10}$$

Given a certain system, we assume that the model and the system are realizations of the same structure but characterized by a different vector θ:

$$\tilde{y}(k) = f(u(k); \tilde{\theta}; k) + \varepsilon(k) \tag{11}$$

where $\varepsilon(k)$ is noise, with variance σ^2_m. Then we have the following theorem:

Theorem: Let N be any time invariant operator with fading memory in a subset K of the input domain. Then, given any $\varepsilon > 0$, there is a set of Laguerre operators and a single hidden layer Perceptron such that for all $u \in K$

$$\left\| Nu - \tilde{N}u \right\| \leq \varepsilon \tag{12}$$

where $\tilde{N}u = \tilde{y}(k), Nu = y(k)$ are given by (11) and (10) respectively (see Sentoni et al., 1998).

DABNet models are simple to evaluate (since they require very little dynamic information), are easy to implement because of their being highly structured and, have simple derivatives formulation. The proposed approximation structure is finally conformed by cascading the DABNet with a FIR model.

2.2 Adaptive Volterra Models

Volterra models are widely applied to represent the behavior of a wide range of dynamical systems. Following the work of Kuech and Kellermann, 2004, the output of a P-th order Volterra model can be expressed by:

$$y(k) = \sum_{p=1}^{P} y_p(k) \tag{13}$$

in which,

$$y_p(k) = \sum_{n_{p,1}=0}^{N_p-1} \cdots \sum_{n_{p,p}=n_{p,p-1}}^{N_p-1} h_{n_p} \prod_{i=1}^{p} x(k - n_{p,i}) \tag{14}$$

with $n_p = [n_{p,1}, n_{p,2}, \ldots, n_{p,p}]$ being the Cartesian coordinate representation of the Volterra kernel coefficients. Another representation of a Volterra scheme can be given by means of a diagonal coordinate system. In fact, defining the index vectors

$$\underline{r}_p \left[r_{p,1}, r_{p,2}, \ldots r_{p,p-1} \right] \tag{15}$$

and

$$r_p(l) = \left[l, r_{p,1} + l, \ldots, r_{p,p-1} + l \right] \tag{16}$$

where $\underline{r}_p$ symbolizes a diagonal of the Cartesian coordinate system representation, and $r_p(l)$ a position in such diagonal, and the input for the diagonal $\underline{r}_p$ as:

$$x_{\underline{r}_p}(k) = x(k) \prod_{i=1}^{p-1} x(k - r_{p,i}) \tag{17}$$

in which $x_{\underline{r}_1}(k) = x(k)$. According to these definitions the output can be restated as:

$$y_p(k) = \sum_{r_{p,1}}^{N_p-1} \cdots \sum_{r_{p,p-1}=r_{p,p-2}}^{N_p-1} y_{\underline{r}_p}(k) \tag{18}$$

With

$$y_{\underline{r}_p}(k) = \sum_{l=0}^{L_p(\underline{r}_p)-1} h_{r_p(l)} x_{\underline{r}_p}(k - l) \tag{19}$$

And

$$L_p(\underline{r}_p) = N_p - r_{p,p-1} \tag{20}$$

representing the length of $h_{\underline{r}_p}(l)$. Therefore, $y_p(k)$ can be interpreted as the output of a linear MISO system where each diagonal with index $\underline{r}_p$ represents the coefficients and $x_{\underline{r}_p}(k)$ the input. This fact is especially important for the considered application, allowing the adaptation of the Volterra coefficients by formal use of NLMS techniques. Furthermore, the Diagonal Coordinate Representation (DCR) of the Volterra model is particularly well suited for Acoustic Echo Cancellation, as the implementation of cascaded structures with very different kernel lengths can be efficiently implemented. The output of the nonlinear model represented in Figure 1(d), can be written as:

$$z(k) = \sum_{p=1}^{P} \sum_{r_{p,1}=0}^{N_p-1} \cdots \sum_{r_{p,p-1}=r_{p,p-2}}^{N_p-1} z_{\underline{r}_p}(k) \tag{21}$$

with

$$z_{\underline{r}_p}(k) = c_k * h_{r\,p(k)} * x_{\underline{r}_p}(k) \tag{22}$$

which can be restated as:

$$z_{\underline{r}_p}(k) = v_{r\,p(k)} * x_{\underline{r}_p}(k), \tag{23}$$

that represents the output of a P-th.orderVolterra filter to an input $x(k)$ with the same DCR structure as that of the original scheme but with the length of the diagonals incremented by $\widetilde{L}_p(\underline{r}_p) = L_p(\underline{r}_p) - N_c - 1$, in which N_c represents the length of the linear filter. In this way, the adaptation for the cascaded filters can be performed in a single operation. Efficient computation of the output of a Volterra scheme given in DCR can be achieved by means of partitioned block methods allowing for fast convolution in the DFT domain. In fact, $h_{\underline{r}_p}(l)$ can be partitioned into $B(\underline{r}_p)$ blocks of length N giving,

$$h_{r_p(i),b} = h_{r_p(l)}\Big|_{l=i+bN}, \tag{24}$$

which correspond to an input signal

$$x_{\underline{r}_{p,b}}(k) = x_{\underline{r}_p}(k - bN)x(k) \quad . \tag{25}$$

The output of the Volterra filter becomes

$$y_{\underline{r}_p}(k) = \sum_{b=0}^{B(\underline{r}_p)-1} h_{r_p(k),b}\, x_{\underline{r}_p,b}(k) \quad . \tag{26}$$

Defining the overlapping signal blocks

$$y_{\underline{r}_p}(m) = [y_{\underline{r}_p}(mR),\dots,y_{\underline{r}_p}(mR+N-1)]^t, \tag{27}$$

$$x_{\underline{r}_{p,b}}(m) = [x_{\underline{r}_{p,b}}(mR-N),\dots,x_{\underline{r}_{p,b}}(mR+N-1)]^t, \tag{28}$$

$$h_{\underline{r}_{p,b}} = [h_{r_p(bN)},\dots,h_{r_p(bN+N-1)}]^t, \tag{29}$$

and,

$$y(m) = [y(mR),\dots,y(mR+N-1)]^t, \tag{30}$$

where, $R = N/\alpha$, and α is known as the overlapping factor. Applying overlap and save method for the convolutions, results in the DFT-domain vectors of length $M=2N$

$$Y_{\underline{r}_p}(m) = F_M[0^t_{Nx1}\, y^t_{\underline{r}_p}(m)], \tag{31}$$

$$X_{\underline{r}_{p,b}} = F_M\, x_{\underline{r}_{p,b}}(m), \tag{32}$$

$$H_{\underline{r}_{p,b}} = F_M[h^t_{\underline{r}_{p,b}}\, 0^t_{Nx1}], \tag{33}$$

and,

$$Y(m) = \sum_{p=1}^{P}\sum_{r_{p,1}}^{N_p-1}\cdots\sum_{r_{p,p-1}=r_{p,p-2}}^{N_p-1} Y_{\underline{r}_p}(m), \tag{34}$$

in which F_M represents the *MxM* DFT matrix, and

$$Y_{\underline{r}_p}(m) = \sum_{b=0}^{B(\underline{r}_p)-1} G\, diag\{X_{\underline{r}_{p,b}}(m)\} H_{\underline{r}_{p,b}}, \tag{35}$$

where,

$$G = F_M\begin{bmatrix} 0_{NxN} & 0_{NxN} \\ 0_{NxN} & 1_{NxN} \end{bmatrix} \tag{36}$$

As stated above, the adaptation of the Volterra coefficients is performed by formal use of NLMS techniques performed in the DFT-domain. Defining the vectors

$$d(m) = [d(mR),\dots,d(mR+N-1]^T \tag{37}$$

$$e(m) = [e(mR),\dots,e(mR+N-1]^T \tag{38}$$

where $d(m)$ is the output of the system, and $e(m)$ the approximation error of the Volterra filter, with DFT-domain components

$$D(m) = F_M[0^t_{Nx1}\, d^t(m)] \tag{39}$$

$$E(m) = F_M[0^t_{Nx1}\, e^t(m)] \tag{40}$$

From equations (34-36) it can be inferred that $Y(m)$ is linear with respect to the Volterra coefficients $H_{\underline{r}_{p,b}}$, therefore the following NLMS-type equation can be applied to update them

$$H_{\underline{r}_{p,b}}(m+1) = H_{\underline{r}_{p,b}}(m) + \mu_p \tilde{G}_{\underline{r}_{p,b}} \Delta_{\underline{r}_{p,b}}(m) \tag{41}$$

where μ_p is an adaptation rate coefficient,

$$\Delta_{r_{p,b}}(m) = S^{-1}(m)E(m)diag\{X^{*}_{r_{p,b}}(m)\}, \quad (42)$$

Where

$$S(m) = diag\{S^{(0)}(m),\ldots,S^{(M-1)}(m)\} \quad (43)$$

is a diagonal normalization matrix, and

$$\widetilde{G}_{r_{p,b}} = F_M \; diag\{[1_{r_{p,b}} \; 0_{r_{p,b}}]\}F_M^{-1} \quad (44)$$

is a constraint matrix where $1_{\underline{r}_{p,b}}$ is a vector of ones with a length given by

$$length\{1_{\underline{r}_{p,b}}\} = \begin{cases} N & ,b < B(\underline{r}_p)-1 \\ L_p(\underline{r}_p)-(b-1)N & ,b = B(\underline{r}_p)-1 \end{cases} \quad (45)$$

and $0_{\underline{r}_{p,b}}$ a vector of zeros with a length given by:

$$length\{1_{\underline{r}_{p,b}}\} + length\{0_{\underline{r}_{p,b}}\} = M \quad (46)$$

If $X_v(m)$, and H_v are the components of $X_{\underline{r}_{p,b}}$ and $H_{\underline{r}_{p,b}}$ that verify $Y^v(m) = H_v^T X_v(m)$, with $Y^v(m)$ is the ν-th. element of $Y(m)$, the corresponding update term is given by

$$\Delta_\nu(m) = \frac{E^{(\nu)}(m)X_\nu^*(m)}{S^{(\nu)}(m)}. \quad (47)$$

When

$$S^{(v)}(m) = \varepsilon\left(X_v^*(m)X_v(m)\right), \quad (48)$$

(in which $\varepsilon(.)$ is the expectation operator) the adaptation coefficient has to verfy $0 < \mu_p < 2$. For the acoustic echo cancellation implementation the parameters of the adaptive DCR Volterra model filter were taken as: $P = 2$ (quadratic Volterra), for the linear kernel: $N_1 = 64$, and $B(r_1) = 6$, $L_1(r_1) = B(r_1)N=384$; for the quadratic kernel $N_2 = 20$, $L_2(r_2) = N_1 = 64$, and $B(r_2) = 1$ (no partitioning). The overlapping factor for the blocks was taken as $\alpha = 4$.

3 Experimental Settings

Real speech data applied to a simulated nonlinear loudspeaker echo channel was used to assess the Performance of the proposed model structure. The speech input signal was obtained by sampling the phrase: "The Discrete Fourier Transform of a real valued signal is conjugate symmetric" at a rate of 8kHz. The following equations describe the echo channel and the nonlinear behavior of the loudspeaker. This is an electro-mechanic device, which transforms an electric current circulating through an electromagnetic coil in sound pressure waves by means of the movement of a cone attached to it. The differential equation representing the dynamic behavior of the mechanical components can be written as (Birkett; Goubran, 1995):

$$m\frac{d^2x}{dt^2}+r_M\frac{dx}{dt}+\frac{x}{C_M}=Bli \tag{49}$$

where ***m*** represents the total mass off the cone, the coil and the air load; $\boldsymbol{r_M}$ is the total mechanical resistance produced by dissipation in the air load and suspension system; $\boldsymbol{C_M}$ indicates the suspension compliance; ***B*** is the magnetic flux density in the air gap of the coil, ***l*** represents the length of the electromagnetic coil conductor; ***i*** indicates the intensity of the current in the coil; ***x*** is the loudspeaker displacement and ***t*** indicates time.

The dynamic behavior of the electric circuit can be represented by means of the following differential equation:

$$e=ir+L\frac{di}{dt}+Bl\frac{dx}{dt} \tag{50}$$

where, ***e*** represents the internal voltage in the generator and ***L*** indicates the inductance of the electromagnetic coil.

The force in the loudspeaker suspension system can be approximated by a polynomial of the form: $f_M=\alpha x+\beta x^2+\gamma x^3$, with α, β, γ constants. The electromagnetic force generated in the coil, f_M, is the force that produces the displacement x, therefore, the compliance of the suspension system can be expressed as:

$$C_M=\frac{x}{f_M}=\frac{1}{\alpha+\beta x+\gamma x^2} \tag{51}$$

Replacing (16) in (13) gives:

$$m\frac{d^2x}{dt^2}+r_M\frac{dx}{dt}+\alpha x+\beta x^2+\lambda x^3=Bli \tag{52}$$

This equation represents the loudspeaker dynamic behavior due to suspension non-linearity, which presents a highly nonlinear character at low frequencies. Another source of dynamic distortion is the lack of uniformity in the magnetic flux density as a function of the loudspeaker (coil-cone) displacement, which can be approximated as: $B(x)=B_0+B_1x+B_2x^2$ in which, B_0, B_1, B_2, are constants. The state-space differential equation representing the nonlinear dynamic behavior of the loudspeaker can be expressed as:

$$\begin{aligned}
\frac{dx_1}{dt}&=\frac{1}{L}\left(-rx_1-B_0lx_3+e-B_1lx_2x_3-B_2lx_2{}^2x_3\right)\\
\frac{dx_2}{dt}&=x_3\\
\frac{dx_3}{dt}&=\frac{1}{m}\left(\begin{matrix}B_0lx_1-\alpha x_2-r_Mx_3-\beta x_2{}^2-\lambda x_2{}^3+lB_1x_1x_2+\\+lB_2x_1x_2{}^2\end{matrix}\right)
\end{aligned} \tag{53}$$

Here $x_1=i, x_2=x, x_3=dx_2/dt$. The values for the simulation parameters were taken as: *L* = 2,5*, r* = 2,75*, B_0l* = 0,5*, $B_1 l$* = 0,1*, $B_2\ l$* = 0,125*, m* = 0,5 / 0,6*,* α = 0,25 / 0,6, β= 0, γ = 0,08 (0,5 / 0,6), r_M = 1,15 (0,5 / 0,6). The input of the system, *e*, is proportional (for a linear amplifier) to the signal amplitude. Thus, the original speech signal was passed through the nonlinear model representing the loudspeaker dynamic behavior to obtain its output. The resulting signal was then filtered with a FIR of 1024 coefficients representing the room impulse response, whose coefficients were obtained as (Asharif et al, 1987):

$$h(n) = \begin{cases} (R/B)\exp^{-An}, 4 \le n \le Length \\ 0, otherwise \end{cases} \tag{54}$$

with, R being a uniform random number between minus one and one, Length = 1024, A = 0,004, B = 1. A DABNet model with a Laguerre system of order nine, pole a = 0, 4, and a 9 neuron Perceptron was set up transversely between the speech signal and the simulated loudspeaker output (Figure 3-a). The data for training was the set comprised between samples 11405 and 12760 of both signals, with the remaining part of the data used as a test set. There were no attempts to optimize the size of the DABNet model. Finally, a linear adaptive FIR model of length 1000 was connected to the output of the DABNet model as its input to approximate the simulated nonlinear echo signal (Figure 1-c). The NRLMS algorithm corresponding to the linear adaptive filter was used with an adaptation rate of 0.9. For performance comparison purposes, a widely used linear adaptive FIR with the same parameters was connected transversely to the speech and the nonlinear echo signals to simulate the echo path. Besides that, another nonlinear model was used: a Single Hidden Layer Perceptron + FIR (Figure 1-b). In this case, the neural network was made of 9 neurons and, and then cascaded with a 1000 taps linear FIR. The next section presents the results.

4 Analysis of results

As stated in the section on experimental settings, the data used for training of the DABNet model was the set comprised between samples 11405 and 12760. The same data set was used to train the Single Hidden Layer Perceptron of the Neural Net+FIR model. Analysis of Figure 3-b, shows that the DABNet model not only represents accurately the dynamic behavior of the loudspeaker system in that range, but also in the entire sample span of the simulation. Therefore, the DABNet model represents precisely the nonlinear dynamic behavior of the loudspeaker, the same conclusion can be drawn for the quadratic Volterra model. A widely used measure of the performance of an Acoustic Echo Cancellation model is the Echo Return Loss Enhancement (ERLE), which is defined as:

$$ERLE(dB) = \lim_{k\to\infty} 10\log\frac{E\left[p^2(k)\right]}{E\left[e^2(k)\right]} \tag{55}$$

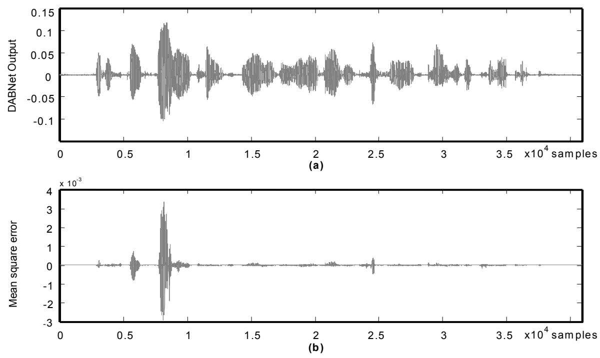

Figure 3: DABNet Approximation