PART III

NONLINEAR ADAPTIVE ALGORITHMS

—The best material model of a cat
is another, or preferably the same, cat.
Norbert Wiener

7

Consequence of Nonlinearities on Hands-free Acoustic Applications

Contents

ONE of the most important limitations of acoustic interfaces in hands-free environments is their inability to effectively cancel or reduce nonlinear interfering signals which impair the speech intelligibility. Nonlinearities in acoustic applications are mainly caused by loudspeakers during large signal peaks; this is the reason why, in this chapter and in the following ones, we focus on applications of nonlinear acoustic echo cancellation where the loudspeaker distortions may affect the echo signal. In this chapter we introduce the problem of nonlinearities and how to address it in acoustic echo cancellation.

7.1 LIMITATIONS OF ACOUSTIC INTERFACES DUE TO NONLINEAR INTERFERING SIGNALS

As said in Chapter 3, the limitations of acoustic interfaces for hands-free applications include circuit and DSP noise, acoustic reverberation, nonstationary signal sources, under-modelling of the AIR, double talk, and in Chapters 5-6 we investigates some algorithms able to tackle these limitations. However, another important limitation is caused by nonlinear interfering sources which draws a significant line at the achievable sound quality. Nonlinearities can be generated by loudspeakers during large signal peaks or by the vibration of the loudspeaker shell which often may be a plastic enclosure; this is the reason why the acoustic application most subjected to nonlinearities is the acoustic echo cancellation due to the acoustic coupling between a microphone and a loudspeaker.

The presence of nonlinearities in acoustic echo paths affects the performance of a conventional AEC compromising the quality requirements of speech communications. In recent years, this topic has become even more sensible matter of interest, due to the growing spread of low-cost commercial hands-free systems, which are often composed of poor quality elements, most of all electronic components, such amplifiers and loudspeakers, and covering materials, such as plastic shells. These devices may cause significant nonlinearities in AIRs leading to perceptual quality degradation of speech [18, 147]. In order to tackle this problem, nonlinear acoustic echo cancellers (NAECs) are employed, thus resulting in nonlinear path modelling and speech enhancement.

In recent years, different structures have been investigated in order to model the nonlinearities rebounding on acoustic echo paths. A prevalent technique is based on the use of nonlinear transformations, able to compensate different kinds of distortions [63, 46, 106, 147]. A *raised-cosine function* is used in [63] to model both *soft-clipping* and *hard-clipping* nonlinearities. In [46], a two-parameter *sigmoid function* is proposed, whose slope and amplitude can be updated during the learning process. Another adaptive sigmoid function is used in [106] to evaluate NAEC performance as reverberation time changes. A more flexible solution is proposed in [147] by using *spline functions*, that are smooth parametric curves defined by interpolation of properly control points collected in a look-up table [148]. Block-based *Wiener-Hammerstein models* using nonlinear functions are also investigated [32, 123, 121].

Even if NAECs using nonlinear functions provide good performance, the most popular nonlinear model for echo cancelling applications is based on *adaptive Volterra filters* (VFs). The generic structure of VFs derives from the well-known Taylor series, and it can be considered as a straightforward generalization of linear adaptive filters [86]. Thus, due to its nature, VFs can model a large range of nonlinearities, both with memory and memoryless [137, 56]. However, acoustic echo cancellation, as well as other hands-free applications, requires large adaptive filter order to model the AIR [120]. Therefore, since computational complexity is proportional to the number of filter coefficients, the adaptation of VFs can become prohibitively expansive, compromising real-time implementation. Moreover, the limitation of Volterra series expansion are similar to those of the Taylor series expansion, thus some types of nonlinearities cannot be modelled by Volterra series, e.g. hard clipping nonlinearities. In recent years Volterra models with reduced computational complexity have been investigated to make real-time implementation possible [43, 44, 138, 47, 10]. However, even in this case an expansion order larger than two has been hardly adopted, due to the complications in adapting such systems and controlling learning rates.

7.2 NONLINEARITY EFFECT ON THE PERFORMANCE OF AN AEC

Before introducing some nonlinear models it is convenient to investigate and analyse the consequence of nonlinearities from a performance perspective. We analyse a loudspeaker model using a real loudspeaker as a case study. Then we show how a nonlinearity of such loudspeaker deteriorates the performance of a conventional AEC.

7.2.1 Nonlinearities in the echo transmission chain

In studying the effects of distortion caused by a loudspeaker, many authors adopted a simplified circuit model of the electro-mechanical-acoustic transducer [70, 19, 56, 136, 137]. An approximation of the loudspeaker model can be justified analysing the kind of nonlinearities involved in the transmission chain, depicted in Fig. 7.1, as done in [138].

The main source of nonlinearities is found in part B (see Fig. 7.1), since the loudspeaker and the power amplifier are operated at the highest signal level of the transmission chain. This part of the system is assumed to be weakly time-variant, e.g. due to temperature drift. The acoustic echo path C is known to be linear and time-variant, while the microphone and the amplifier C can be modeled as linear shift-invariant (LSI) systems (see Paragraph 3.1.1) because of their low signal amplitudes. Also the nonlinear quantization of the A/D and D/A converters can be neglected in this context. If nonlinear distortions are

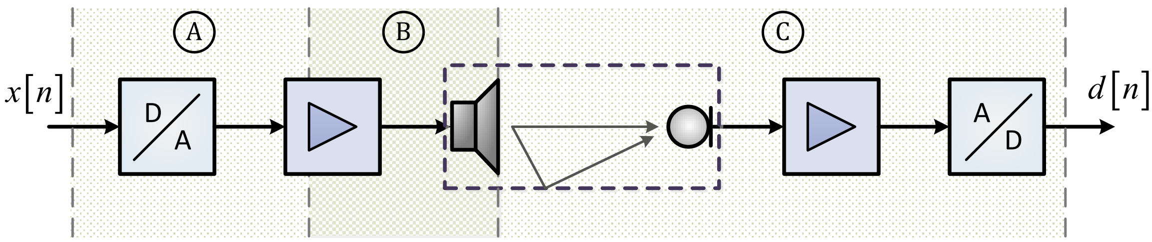


Fig. 7.1: *Echo transmission chain.*

mainly caused by an overdriven amplifier, they are approximately memoryless and can be modeled by a saturation curve [136, 96]. In particular, in [96], parts A and C of Fig. 7.1 are modelled with adaptive FIR filters and part B is realized by a saturation curve with one adaptive parameter. However, the adaptation of the whole system results computationally very demanding. On the other side, in [136], a system with non-adaptive nonlinearity models part A in Fig. 7.1 as a delay, part B by a 7-th order polynomial, and part C as a classical NLMS adaptive filter. With negligible additional effort an ERLE improvement is obtained, without affecting convergence properties of the adaptive filter. However, experiments in [138] show, that both systems obtain their good results only if the major cause of nonlinearities is a clipping amplifier. In many non-portable applications, like smartphones, the power amplifier is not necessarily overdriven, but it is still desiderata to operate a small, cheap speaker at its power limit. With such an echo path the systems in [136] and [96] do not achieve remarkable ERLE improvements. This shows the need to develop another kind of nonlinear echo canceller which is appropriate for systems with loudspeaker nonlinearities.

This kind of nonlinearity is caused by the loudspeaker [49], especially when it is operated at its power limit. Due to the long time constants of the electro-mechanical system, the memory of this nonlinear behaviour cannot be neglected, as confirmed in [138]. To combat this type of nonlinearity, adaptive systems with memory are required. A *time-delay neural network*, being such a system, is proposed in [19]. With a cascade of a time-delay neural network and an adaptive FIR filter, considerable improvement of nonlinear echo reduction is achieved. A disadvantage is the need for a second reference microphone to provide an error signal for the adaptive neural network. In [143], adaptive VFs have been proposed for line echo cancelling. However, due to their high numerical complexity they have not been used in practical systems yet. In [138], an acoustic echo canceller with a second order adaptive Volterra filter has been developed and a method that keeps the computational complexity modest is proposed. From then on, other works have been proposed using Volterra models, as previously said in Section 7.1.

7.2.2 Loudspeaker identification by means of a neural network

In order to prove to evaluate nonlinear models in presence of distortions caused by a loudspeaker, we exploit the generalization capabilities of an artificial neural networks (ANN) [60] to obtain a functional model of a loudspeaker. In order to obtain adequate examples for the training of an ANN, we use data collected in a thesis work [112]. Data consists of 11 signals with linearly increasing amplitude including sinusoidal sweeps with frequency rate from 10 to 500 Hz in 16 bit wave form with a sample rate of 48 kHz. These signals

Electrical resistance $[\Omega]$	11.06
Mechanical compliance of driven suspension	0.14E-0.3
Loudspeaker resonance frequency $[\mathrm{Hz}]$	77.19
Equivalent acoustic volume	64.9E-03
Mechanical stiffness of driver suspension $[\mathrm{N/m}]$	4.08
Force factor $[\mathrm{N/A}]$	14.8
Electric Q factor	0.73
Sound Pressure Level	99.649
Total Q factor	0.62
Efficiency	3.95%
Equivalent inductance $[\mathrm{mH}]$	1.15
Equivalent piston area $\left[\mathrm{m}^2\right]$	56.8E-03
Nominal impedance $[\Omega]$	16
Mechanical mass of the driven diaphragm $[\mathrm{g}]$	29.7

Table 7.1: *Technical description of the loudspeaker model APW300, S.I.P.E. S.P.A. Electroacoustics.*

are used to excite a commercial loudspeaker, model APW300, produced by S.I.P.E. S.P.A. Electroacustics in Chiaravalle (AN), Italy, whose technical data are reposted in Table 7.1 and whose frequency response is depicted in Fig. 7.2. Measurements are conducted in an anechoic room in order to avoid any reverberations; all the data are finally decimate at 2 kHz.

We use a dynamic ANN with 20 inputs (10 MA and 10 AR), with 12 spline

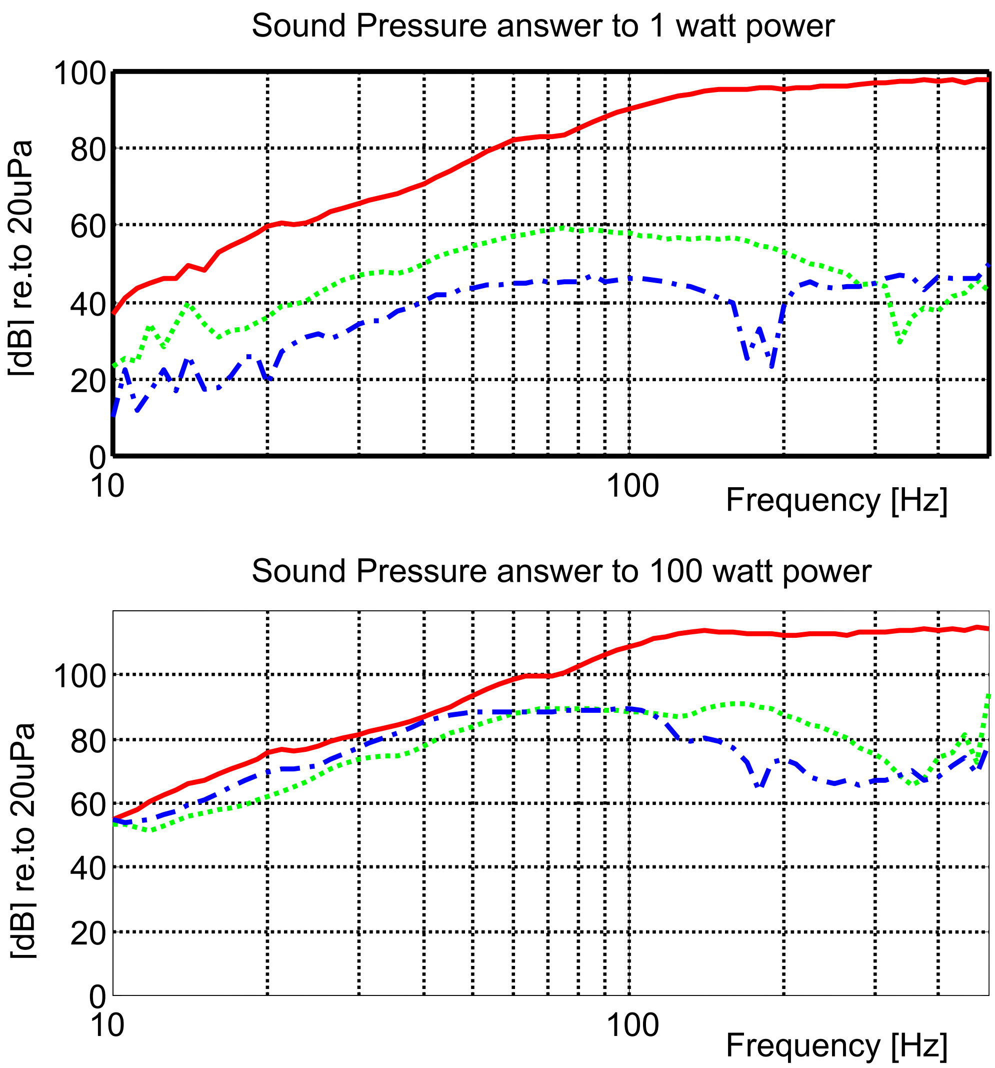
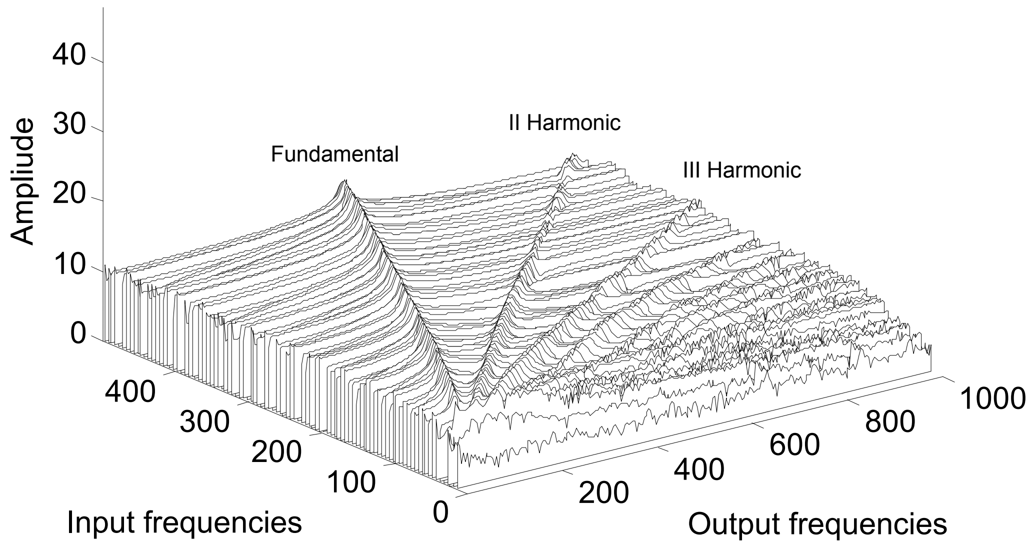
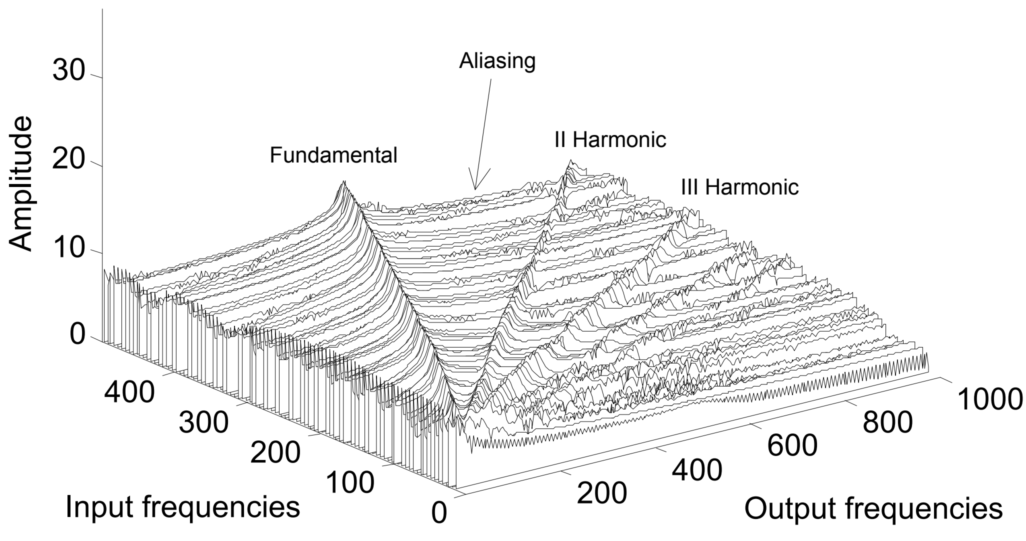


Fig. 7.2: *Frequency response of the loudspeaker APW300 at 1 W and 100 W. The red line is the fundamental harmonic, the green line is the second harmonic and the blue line is the third harmonic.*

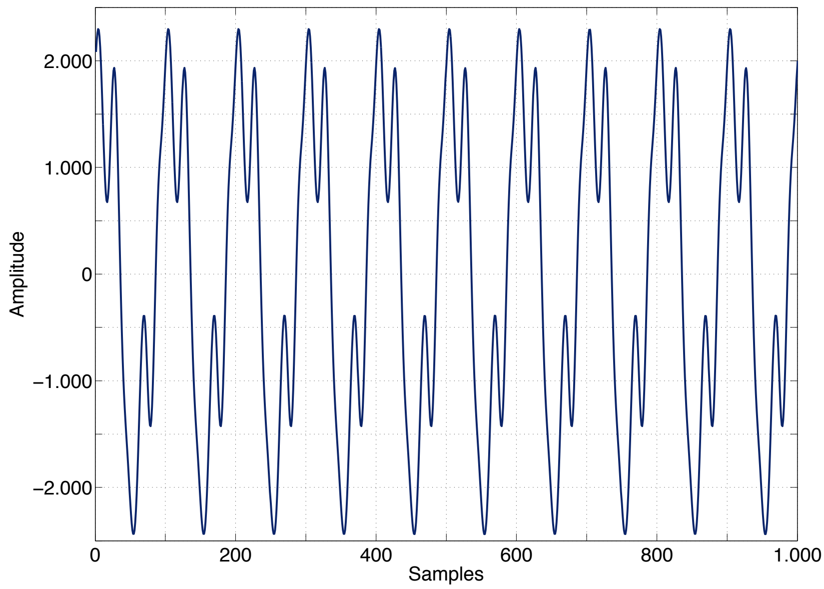


(a) Harmonic distortion of loudspeaker APW300.

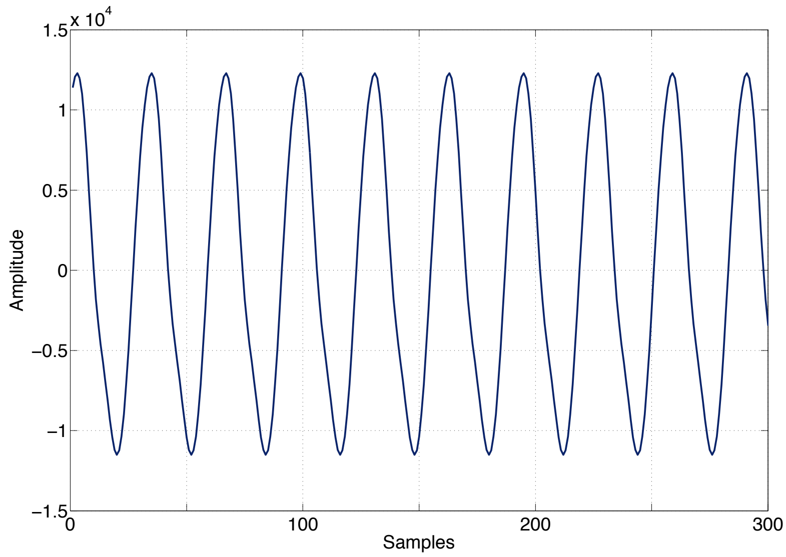


(b) Harmonic distortion of neural model.

Fig. 7.3: *Comparison between the harmonic distortion of the loudspeaker APW300 (a) and the neural loudspeaker model (b). The input signal is a sweep.*



(a) Sine at 80 Hz.



(b) Sine at 250 Hz.

Fig. 7.4: *Distortion effect of the neural loudspeaker model (a) on a sine at 80 Hz and (b) on a sine at 250 Hz. Being a professional loudspeaker the distortion is more evident at low frequencies.*

neurons [55, 113] with 28 points and fixed step $\Delta x = 0.5$. A *backpropagation algorithm* is used as learning rule; however, the learning rate is normalized with a quantity proportional to the input signal energy. In order to evaluate the distortion produced by this loudspeaker and the identification capability of the adopted ANN it is possible to compare Fig. 7.3 (a) and Fig. 7.3 (b).

The distortion effect of the neural loudspeaker model is depicted in Fig. 7.4 where it is clear that, being the APW300 a professional loudspeaker, nonlinearities affect a signal at low frequencies, so that a sine at 80 Hz results more distorted than a sine at 250 Hz. However, this is sufficient to produce a worsening in the performance of an AEC.

7.2.3 Performance worsening in an AEC process

In order to evaluate the loss of quality caused by loudspeaker distortions in an AEC process, we compare AEC performance using both an ideal purely linear model and the neural loudspeaker model previously described. We use a common hands-free scenario of a typical office room with a reverberation time of $T_{60} \approx 130$ ms, thus resulting the AIR depicted in Fig. 6.1 (b). We

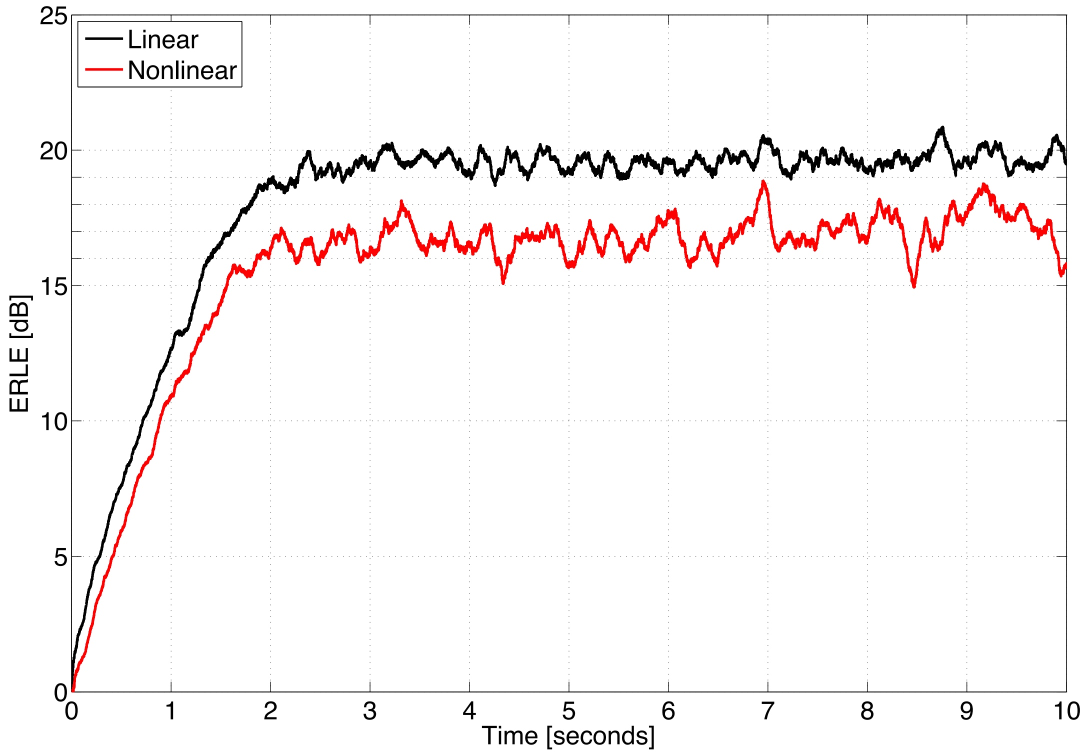


Fig. 7.5: *Loss of quality in terms of ERLE caused by loudspeaker distortions when the far-end input is a white Gaussian noise.*

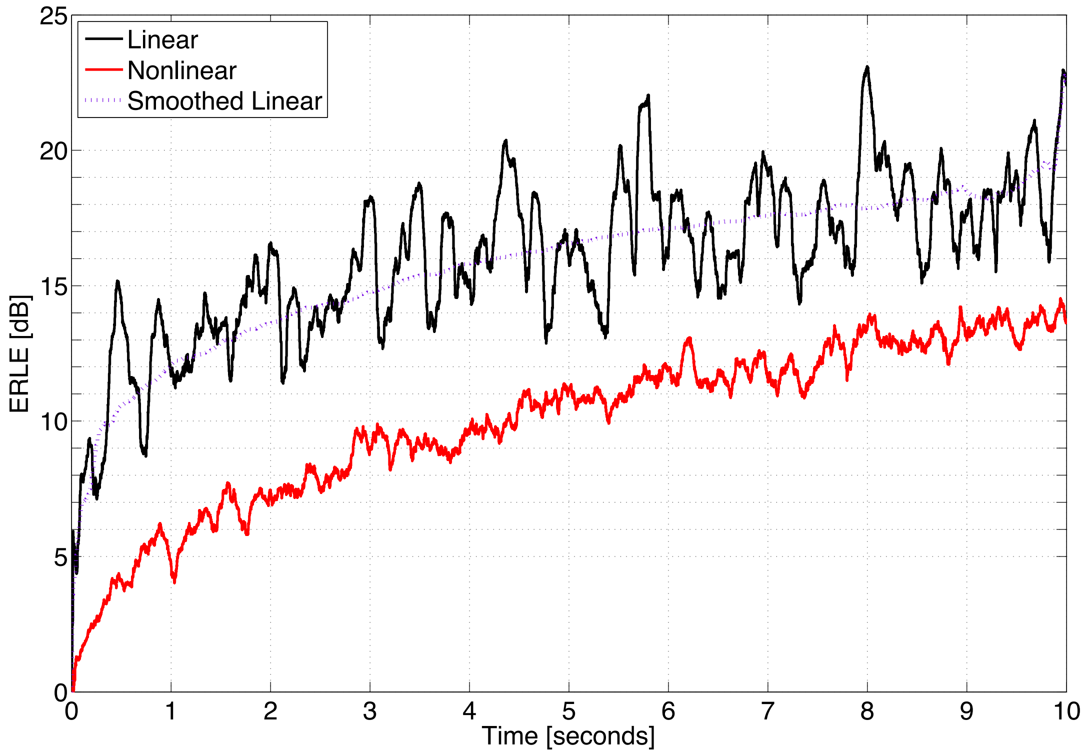


Fig. 7.6: *Loss of quality in terms of ERLE caused by loudspeaker distortions when the far-end input is a coloured noise. The dotted line represents the average performance in the linear case which clarifies the difference from the nonlinear performance.*

evaluate performance in terms of ERLE in three cases: when the far-end signal is a white Gaussian noise with zero mean and unitary variance, when the far-end signal is a coloured noise obtained through an autoregressive process of the white Gaussian noise signal, and eventually when the far-end input is a female speech signal. In all the cases an additive white Gaussian noise is added providing 20 dB of SNR in order to simulate some near-end background noise.

In Fig. 7.5, AEC performance in terms of ERLE is represented when the far-end signal is white Gaussian noise. The black line represents the ERLE performance in absence of distortions while the red line denotes the ERLE performance in presence of loudspeaker distortions. It is quite evident from this graph that the presence of distortions in the echo signal causes a loss

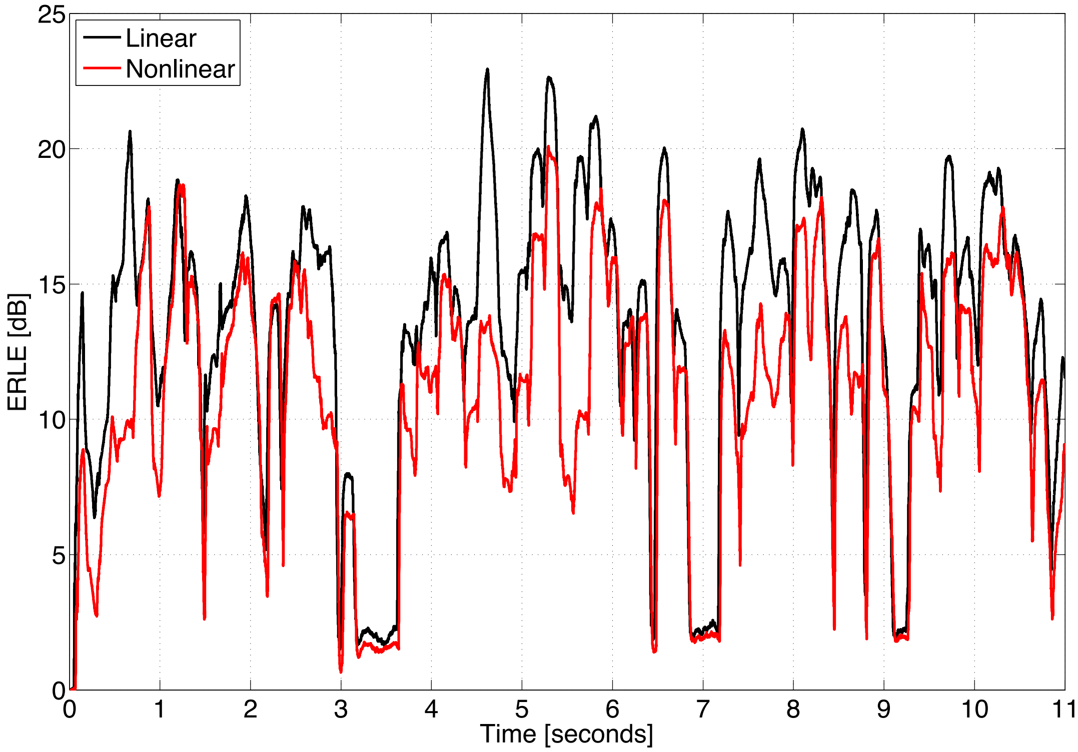


Fig. 7.7: *Loss of quality in terms of ERLE caused by loudspeaker distortions when the far-end input is a female speech signal.*

of quality of about 3 dB. The gap between performance in absence and in presence of distortions is more evident when the far-end signal is a coloured signal, i.e. a speech-like signal, as it is possible to see in Fig. 7.6, when the loss of quality is comprised on average within the range from about 3 to 7 dB. A confirmation of this trend is achieved when the far-end signal is a speech signal, as depicted in Fig. 7.7, where the loss of quality is larger than 7 dB in some peak of the signal. These results show that in hands-free acoustic applications, even with a professional loudspeaker, an important loss of quality can be obtained in presence of nonlinearities. Let us note that the performance in the linear ideal case represents the maximum achievable quality. Therefore an NAEC may improve the performance of an AEC in presence of distortions and may reach at most the achievable performance, thus we may expect to plug the performance gap as much as possible.