Computers in Biology and Medicine 46 (2014) 29–41

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

A complementary low-cost method for broadband noise reduction in hearing aids for medium to high SNR levels

Márcio Holsbach Costa *

Department of Electrical Engineering, Federal University of Santa Catarina, 88040-900 Florianópolis, SC, Brazil

ARTICLE INFO

Article history:
Received 13 June 2013
Accepted 18 December 2013

Keywords:
Hearing-aids
Noise reduction
Adaptive filter
Speech processing
Noise cancelling

ABSTRACT

This work presents a complementary broadband noise reduction scheme for hearing aid applications. It is designed to attenuate uncorrelated and small-correlation-length acoustic noise with controlled speech distortion. Noisy speech signals are pre-processed by the proposed strategy before being subjected to an existing narrowband noise reduction system. The clean speech signal is estimated by a convex combination of the unprocessed speech signal and the output of a linear predictor. The convex combination coefficient is adjusted to provide noise suppression while avoiding significant unvoiced utterance distortions. The proposed method is optimized to minimize speech mean-square prediction-error. A low-cost adaptive implementation is proposed and compared to the conventional adaptive linear predictor showing an improved performance, as predicted by theory. Four different objective quality measures and subjective assessment performed by normal hearing volunteers indicate that the combined use of the proposed technique with a narrowband noise reduction system consistently improves speech quality for a range of signal to noise ratios. Low-cost digital hearing aids that make use of the conventional adaptive predictor for broadband noise reduction can be easily modified to incorporate the new proposal with a minimum amount of extra computational resources.

1. Introduction

Hearing-aids are essential devices for the social integration of people that suffer from hearing limitations or neurosensorial losses. These conditions affect about 9% of the world's population [1,2].

Hearing-aids are very complex systems consisting of several processing units that perform tasks such as adaptive directionality, noise reduction, dynamic compression and feedback cancellation. These elements interact to improve intelligibility and provide better acoustic comfort for the user. Due to their small size and power consumption requirements, the availability of computational resources for each subsystem is restricted. As a result, each technique should be designed as sparingly as possible.

Despite the great advancements in this area, two main causes of sensorial discomfort – noise amplification and reverberation – still persist [3]. One of the major complaints of hearing aid users is poor speech intelligibility due to background noise. Many studies have demonstrated that hearing impaired people need an SNR-50[1] from 10 to 30 dB higher than that required for the non-impaired [4].

The basic function of a noise reduction system is to lessen the user's perception of environmental acoustic noise, minimizing distortion and masking effects[2]. With noise reduction, acoustic comfort increases while fatigue decreases, which in turn increases the equipment's acceptability[3] [5,6].

Although multi-microphone hearing-aids may have many advantages over single microphone devices[4], some commercial gadgets are still equipped with the latter [7,8].

The most common single microphone noise reduction approaches are [9]: (a) subspace decomposition [10], (b) statistical and parametric modelling [11–13], and (c) Wiener filtering. The first two approaches, although feasible, incur considerable computational complexity (even when look-up tables are used to alleviate the effort) and large time delay due to signal processing [14]. Wiener filtering has also been applied to hearing-aids; however, it tends to generate unpleasant

* Corresponding author. Tel.: +55 48 3721 9506; fax: +55 48 3721 9280.
E-mail address: costa@eel.ufsc.br

[1] SNR-50 is the signal to noise ratio needed for the comprehension of 50% of the speech in a conversation.

[2] Distortion happens when the hearing impaired perceives sound but its intelligibility is compromised due to the lack of high frequency information. The problem is worsened by the existence of background noises, resulting in the complaint that "it is possible to hear, but not understand, speech" [2]. Moreover, masking decreases speech redundancy; as a result, small amounts of noise can lead to significant intelligibility degradation [50].

[3] Although noise reduction systems can provide a substantial improvement in speech quality (acoustical comfort), the same effect does not necessarily occur upon intelligibility. Intelligibility improvements can be obtained by speech enhancement systems [16,51], whose aim is not to reduce noise, but rather highlight the contrast between vowels and consonants [1].

[4] Multi-microphone techniques take advantage of the spatial separation among acoustic sources [1], but they require a considerable distance among microphones. This condition makes them inappropriate for ear-canal devices [52].

0010-4825/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compbiomed.2013.12.009

acoustic artefacts called "musical noise" [3]. Without exception, all techniques present a trade-off between noise reduction and speech distortion [9,15,16].

The limited computational resources available in commercial devices greatly limit the development of new techniques for hearing-aid improvement. In recent years, manufacturers have provided specific hardware to lessen this problem. An example of a very successful commercial architecture for the (software/hardware) implementation of a noise reduction system can be found in Ref. [17]. In this architecture, the digitized acoustic signal is split into different frequency channels; each of them is then subjected to independent attenuation factors before signal reconstruction. Signals from high signal to noise ratio (SNR) channels are preserved, while low SNR channel signals are attenuated. This approach, introduced in Ref. [18], works well only for narrowband noise. For broadband noise, all channels tend to suffer approximately the same attenuation, maintaining the same global SNR.

In Ref. [19], an extensive analysis of background noise databases showed that not all daily-life noises can be referred to as narrowband (low-frequency) background noises. As a result, the authors of Ref. [19] suggest that hearing aids should not only be fine-tuned to the individual audiogram but also to environmental conditions. In fact, in Ref. [19] it was shown that noise in environments such as industry and nature preponderantly present flat spectrum without temporal modulations. In work conditions (industry), hearing-aid users cannot reduce volume due to the possibility of warning sounds [20]. Consequently, many hearing-impaired workers are often forced to endure a certain degree of discomfort. In addition, [21] stated that the most difficult listening situations that commonly face persons with hearing loss feature broadband competition.

Some attempts to overcome the broadband noise problem in hearing-aids can be found in Refs. [22,23]. This work will focus on broadband acoustic noise characterized by small correlation-length[5] (such as low-pass noise sampled at near-Nyquist frequency rates[6]).

The conventional linear adaptive predictor (CLAP) [24–26] is a low-complexity solution that performs well in reducing broadband noise. However, it also tends to cancel uncorrelated speech components, which constitute about 20 to 25% of natural speech in the English language [27]. As a result, it produces muffled speech sounds and musical noise which can severely affect both speech intelligibility and naturalness. Some works have recently addressed the design of practical low distortion broadband noise cancellers based on the CLAP structure. In Ref. [28], a weighted sum of contaminated signal and CLAP output was proposed. This approach aims to enhance the quasi-stationary components of speech (voiced sounds), improving intelligibility and, secondarily, SNR. However, many intelligibility problems can be attributed to poor comprehension of unvoiced sounds. In Ref. [29], a first attempt was made to control the CLAP attenuation of uncorrelated speech components. However, the authors were not successful in accurately determining the optimal control parameter due to the use of very restrictive theoretical assumptions. In Ref. [30], CLAP output and error signals were linearly combined using attenuation factors directly related to instantaneous SNR. This approach provides poor results when unvoiced speech and uncorrelated noise occur simultaneously. Hence, low-cost reduction of uncorrelated noise remains an open issue of great interest for hearing-aids designers.

This work proposes a complementary low-cost technique for broadband noise reduction in hearing-aids for pre-processing of noisy-speech signals before narrowband noise reduction. Clean speech is estimated using a convex combination of the original contaminated signal and the output of a linear predictor. The convex combination weight factor establishes a trade-off between uncorrelated noise reduction and unvoiced speech distortion. An adaptive version of the algorithm is proposed. The proposed system is fitted to hearing-aid applications due to three main requirements: (a) availability of a narrowband noise reduction system to alleviate acoustical discomfort due to CLAP coefficient fluctuations and to reduce narrowband noise; (b) small signal processing time delay (since 6 to 8 ms delays can be undesirably perceived by users, while over 10 ms can be considered annoying [1]); and (c) low extra computational cost (in addition to the processing load previously existent in the hearing aids processing system). The problem is mathematically described in Section 2. Section 3 briefly reviews the prediction of a signal immersed in noise. Section 4 presents the proposed method and its optimization strategy. Section 5 presents a simple adaptive implementation of the proposed technique. Section 6 shows simulation results using synthetic and real speech signals. The results obtained corroborate the theoretical derivations and illustrate the performance of the proposed method. Final conclusions are presented in Section 7.[7] The proposed method would be particularly useful for commercial devices in which it would be used together with an existing narrowband noise reduction system [17,18,31]. Throughout this text, bold uppercase and lowercase letters represent matrices and vectors, respectively, while italics represent scalars.

2. Problem description

The sampled acoustic signal at time instant n is modelled as the sum of a speech signal $x(n)$ and noise $\eta(n)$, resulting in

$$y(n) = x(n) + \eta(n). \tag{1}$$

here, noise $\eta(n)$ is assumed stationary, independent of $x(n)$, zero-mean with power σ_η^2 and with a small correlation-length so that $|E\{\eta(n)\eta(n-k)\}| \le \varepsilon$ for $k \ge K$, where ε is a very small constant and K is a finite integer smaller than the speech correlation-length. Noise is white in the particular case of $K=1$ and $\varepsilon=0$.

Speech signal $x(n)$ has zero mean with power σ_x^2 and is modelled by an autoregressive process with a small correlation-length for unvoiced utterances or a large correlation-length for voiced utterances. The model coefficients are assumed constant in a given time window (about 20 ms).

The mean-square prediction-error (MSPE), resulting from predicting the clean speech $x(n)$ by the unprocessed (contaminated) speech $y(n)$, is given by

$$J_{US} = E\left\{[x(n) - y(n)]^2\right\} = \sigma_\eta^2, \tag{2}$$

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

3. Prediction of a signal immersed in noise

The Wiener filter is a widespread noise reduction technique that presents a known trade-off between speech distortion and noise reduction. Since noise and speech usually share the same frequency range, noise statistics are usually estimated during voice pauses. In practical applications, large time periods between estimations can lead to substantial degradation of the noise reduction process. However, assuming noise has a small correlation-length compared to the speech, a prediction approach can be used to continuously compute pseudo-optimum noise reduction filters without significant performance loss.

[5] The correlation length of a random signal $x(n)$, with exponential decaying autocorrelation function, is defined as $L_X = \sum_{l=0}^{\infty} E\{x(n)x(n-l)\}/E\{x^2(n)\}$ [32]. It measures the signal memory.

[6] This condition can be found in low-cost or severely limited computational systems, as is the case of hearing aid devices.

[7] Preliminary results were published in [53].

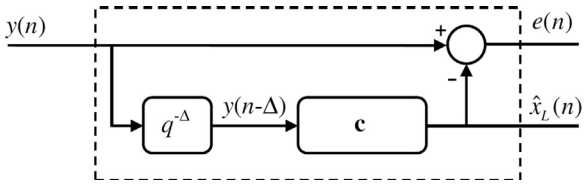

Fig. 1. Forward linear predictor.

Prediction of a desired signal $x(n)$ immersed in noise consists in obtaining an estimate of $x(n)$ from the knowledge of a window of past values of the contaminated signal $y(n)$ [32]. One simple and popular method to achieve this is the forward linear predictor (FLP) [33,34] shown in Fig. 1. Results without significant distortion can be achieved in speech applications if the noise $\eta(n)$ in Eq. (1) has a significantly smaller correlation-length than that of speech signal. For a window of stationarity, FLP output is given by the following inner product

$$\hat{x}_L(n) = \mathbf{c}^T\mathbf{y}(n-\Delta), \tag{3}$$

where the optimal value of Δ for noise reduction lies in the interval between the correlation-length of speech and the correlation-length of noise [33], $\mathbf{y}(n-\Delta)=[\ y(n-\Delta)\ y(n-\Delta-1)\ \ldots\ y(n-\Delta-N+1)]^T$ and $\mathbf{c}=[c_0\ c_1\ \ldots\ c_{N-1}]^T$, where c_i, for $i=0,1,\ldots,N-1$ are the linear prediction coefficients. The minimum MSPE of FLP is [34]

$$J_{LP}(\mathbf{c_o}) = E\{[x(n)-\hat{x}_L(n)]^2\} = \sigma_x^2 - \mathbf{c_o}^T\mathbf{R_{yy}}\mathbf{c_o}, \tag{4}$$

where $\mathbf{R_{yy}}=E\{\mathbf{y}(n)\mathbf{y}^T(n)\}$, and the point of minimum MSPE is [34]

$$\mathbf{c_o} = \mathbf{R_{yy}}^{-1}\mathbf{r}_{x\mathbf{x}_\Delta}, \tag{5}$$

where $\mathbf{r}_{x\mathbf{x}\Delta}=E\{x(n)\mathbf{x}(n-\Delta)\}$, and $\mathbf{x}(n-\Delta)=[x(n-\Delta)\ x(n-\Delta-1)\ \ldots\ x(n-\Delta-N+1)]^T$.

As $x(n)$ is quasi-stationary, the solution in Eq. (5) becomes time-varying and must be tracked through successive windows of quasi-stationarity. The following analyses and derivations assume stationary signals.

3.1. Noise reduction versus speech distortion

From Eq. (5), using $\mathbf{R_{xx}}=E\{\mathbf{x}(n)\mathbf{x}^T(n)\}$, $\mathrm{SNR}=\sigma_x^2/\sigma_\eta^2$ and assuming for (illustrative) convenience that $\mathbf{R_{\eta\eta}}=\sigma_\eta^2\cdot\mathbf{I}$, leads to Ref. [9]

$$\mathbf{c_o} = \left(\frac{\mathbf{I}}{\mathrm{SNR}} + \tilde{\mathbf{R}}_{\mathbf{xx}}\right)^{-1}\tilde{\mathbf{r}}_{x\mathbf{x}_\Delta}, \tag{6}$$

where $\mathbf{R_{yy}}=\mathbf{R_{xx}}+\mathbf{R_{\eta\eta}}$ (independence between speech and noise) was applied. The normalized autocorrelation matrix and normalized cross-correlation vector are respectively defined as

$$\tilde{\mathbf{R}}_{\mathbf{xx}} = \frac{1}{\sigma_x^2}\mathbf{R_{xx}}, \quad \tilde{\mathbf{r}}_{x\mathbf{x}_\Delta} = \frac{1}{\sigma_x^2}\mathbf{r}_{x\mathbf{x}_\Delta}. \tag{7}$$

The two normalized statistics in Eq. (7) are invariant to power changes in $x(n)$ and $\eta(n)$. For extreme SNRs, (6) turns into

$$\begin{aligned} \mathrm{SNR}\to 0 \quad &\Rightarrow \quad \mathbf{c_o}\to\mathbf{0} \\ \mathrm{SNR}\to\infty \quad &\Rightarrow \quad \mathbf{c_o}\to\tilde{\mathbf{R}}_{\mathbf{xx}}^{-1}\tilde{\mathbf{r}}_{x\mathbf{x}_\Delta} \end{aligned}. \tag{8}$$

Eq. (8) clearly shows that there will be no signal at the optimum FLP output in the absence of speech signal ($\mathrm{SNR}\to 0$ and $\mathbf{c_o}\to\mathbf{0}=[0\ 0\ \ldots\ 0\ ]^T$). However, very high SNRs can also lead to signal cancelling at optimum FLP output. For instance, if the correlation-length of a given speech epoch is smaller than Δ, $\mathbf{r}_{x\mathbf{x}\Delta}$ entries will be zeros or very small values. Thus, it is imperative to investigate under which conditions the FLP is suitable for noise reduction.

From Fig. 1, it is possible to verify that $y(n)=e(n)+\hat{x}_L(n)$. Using Eq. (1) in Eq. (3) results in

$$\hat{x}_L(n) = x(n) - x_N(n) + \eta_P(n), \tag{9}$$

where $\eta_P(n)=\mathbf{c}^T\boldsymbol{\eta}(n-\Delta)$ and $x_N(n)=x(n)-\mathbf{c}^T\mathbf{x}(n-\Delta)$ are called the residual noise and the speech distortion, respectively [35]. Using Eq. (9) in Eq. (4), and assuming the optimal setting $\mathbf{c}=\mathbf{c_o}$, FLP MSPE in Eq. (4) can also be given by

$$J_{LP}(\mathbf{c_o}) = \sigma_{x_N}^2 + \sigma_{\eta_P}^2, \tag{10}$$

where

$$\begin{aligned} \sigma_{x_N}^2 &= E\{x_N^2(n)\}|_{\mathbf{c}=\mathbf{c_o}} = E\{[x(n)-\mathbf{c_o}^T\mathbf{x}(n-\Delta)]^2\} \\ &= \sigma_x^2 - \mathbf{c_o}^T\mathbf{R_{xx}}\mathbf{c_o} - 2\mathbf{c_o}^T\mathbf{R_{\eta\eta}}\mathbf{c_o} \end{aligned} \tag{11}$$

is the speech distortion power, and

$$\sigma_{\eta_P}^2 = E\{\eta_P^2(n)\}|_{\mathbf{c}=\mathbf{c_o}} = E\{[\mathbf{c_o}^T\boldsymbol{\eta}(n-\Delta)]^2\} = \mathbf{c_o}^T\mathbf{R_{\eta\eta}}\mathbf{c_o} \tag{12}$$

is the residual noise power.

The prediction improvement obtained by using FLP can be assessed by the ratio G_{LP} of J_{US} in Eq. (2) and $J_{LP}(\mathbf{c_o})$ in Eq. (4)

$$G_{LP} = \frac{J_{US}}{J_{LP}(\mathbf{c_o})} \tag{13}$$

MSPE improvement occurs whenever $G_{LP}>1$. In this situation, using (2) and (10) in Eq. (13) we obtain the following condition

$$\sigma_\eta^2 > \sigma_{x_N}^2 + \sigma_{\eta_P}^2 \tag{14}$$

i.e. the linear predictor is useful (MSPE-wise) only when noise power is bigger than the combined sum of residual noise power and speech distortion inserted by FLP. This condition is usually satisfied during voiced utterances, when the uncorrelated portion of the contaminated speech signal $y(n)$ is basically associated with additive noise. Conversely, the amount of distortion inserted during unvoiced utterances can exceed the advantages obtained by noise reduction, compromising speech naturalness and intelligibility.

4. Proposed method

The proposed architecture, shown in Fig. 2, aims to provide some control in the inserted distortion during unvoiced utterances or noise-only periods and was motivated by Refs. [9] and [29]. Its output signal is given by

$$\hat{x}(n) = y(n) - \alpha e(n) = (1-\alpha)y(n) + \alpha\hat{x}_L(n) \tag{15}$$

This structure estimates clean speech by a convex combination [36] of the unprocessed signal $y(n)$ and the FLP output $\hat{x}_L(n)$. The convex combination parameter α is designed to obtain improvements in both comfort and naturalness (under different SNR conditions) over the conventional FLP solution for speech utterances with distinct statistical characteristics. For such, α should tend to unity during voiced utterances, since these signals can be appropriately estimated by FLP. In the case of unvoiced utterances, the choice of α establishes a trade-off between speech distortion and noise reduction. A first attempt to find the optimum parameter α was presented in Ref. [29]; however, the unrealistic

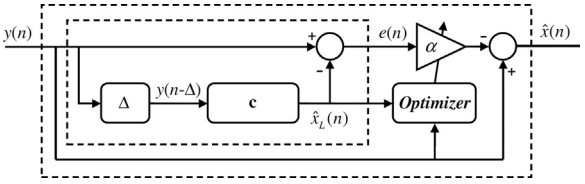

Fig. 2. Proposed structure.

assumption that the CLAP provides perfect predictions of speech ruled out accurate results.

4.1. Cost function

The MSPE for the structure in Fig. 2 is determined by

$$J_{PM}(\mathbf{c},\alpha)=E\{[x(n)-\hat{x}(n)]^2\} \tag{16}$$

Using (1), (3) and (15) in Eq. (16) we obtain

$$J_{PM}(\mathbf{c},\alpha)=E\{[\alpha x(n)-(1-\alpha)\eta(n)-\alpha\mathbf{c}^T\mathbf{y}(n-\Delta)]^2\} \tag{17}$$

Its manipulation yields

$$\begin{aligned}J_{PM}(\mathbf{c},\alpha)=&\alpha^2E\{x^2(n)\}+(1-2\alpha+\alpha^2)E\{\eta^2(n)\}\\&-2\alpha(1-\alpha)E\{x(n)\eta(n)\}-2\alpha^2E\{x(n)\mathbf{x}^T(n-\Delta)\}\mathbf{c}\\&-2\alpha^2E\{x(n)\boldsymbol{\eta}^T(n-\Delta)\}\mathbf{c}+2\alpha(1-\alpha)E\{\eta(n)\mathbf{y}^T(n-\Delta)\}\mathbf{c}\\&+\alpha^2\mathbf{c}^TE\{\mathbf{y}(n-\Delta)\mathbf{y}^T(n-\Delta)\}\mathbf{c}\end{aligned} \tag{18}$$

Considering the statistical independence of $x(n)$ and $\eta(n)$ and assuming Δ large enough so that $E\{y(n-\Delta)\eta(n)\}=0$, yields

$$J_{PM}(\mathbf{c},\alpha)=\alpha^2\sigma_y^2+(1-2\alpha)\sigma_\eta^2-2\alpha^2\mathbf{r}_{\mathbf{x}\mathbf{x}_\Delta}^T\mathbf{c}+\alpha^2\mathbf{c}^T\mathbf{R}_{\mathbf{yy}}\mathbf{c}, \tag{19}$$

where ${\sigma_y}^2=E\{y^2(n)\}={\sigma_x}^2+{\sigma_\eta}^2$. For the optimum FLP coefficient vector $\mathbf{c_o}$, (19) yields

$$J_{PM}(\mathbf{c_o},\alpha)=(\sigma_y^2-\mathbf{c_o}^T\mathbf{R}_{\mathbf{yy}}\mathbf{c_o})\alpha^2-2\sigma_\eta^2\alpha+\sigma_\eta^2. \tag{20}$$

4.2. Optimal convex combination

Eq. (20) is a quadratic function that can be minimized by making its gradient with respect to α equal to zero. The optimal coefficient α is then

$$\alpha_o=\frac{\sigma_\eta^2}{\sigma_y^2-\mathbf{c_o}^T\mathbf{R}_{\mathbf{yy}}\mathbf{c_o}}. \tag{21}$$

In Eq. (21), $\mathbf{c_o}^T\mathbf{R}_{\mathbf{yy}}\mathbf{c_o}$ is the FLP output power for the optimal coefficient vector $\mathbf{c_o}$.

4.3. Minimum mean square prediction error

Using (21) in Eq. (20) and knowing that ${\sigma_y}^2={\sigma_x}^2+{\sigma_\eta}^2$, the minimum MSPE of the proposed technique is given by

$$J_{PM}(\mathbf{c_o},\alpha_o)=\frac{(\sigma_x^2-\mathbf{c_o}^T\mathbf{R}_{\mathbf{yy}}\mathbf{c_o})\sigma_\eta^2}{(\sigma_x^2-\mathbf{c_o}^T\mathbf{R}_{\mathbf{yy}}\mathbf{c_o})+\sigma_\eta^2}. \tag{22}$$

The FLP-MSPE ratio and that of the proposed technique can be obtained dividing (4) by Eq. (22), resulting in

$$G_{PM}=\frac{J_{LP}(\mathbf{c_o})}{J_{PM}(\mathbf{c_o},\alpha_o)}=1+\frac{J_{LP}(\mathbf{c_o})}{J_{US}}=1+\frac{1}{G_{LP}}. \tag{23}$$

where the last equality comes from using (13).

Since G_{LP} is always positive (it is a ratio between mean square values), (23) shows that the structure in Fig. 2 always yields a better MSPE-wise clean speech signal estimate than that of the structure in Fig. 1 (since $1/G_{LP}>0$ for any G_{LP}, resulting in $G_{PM}>1$).

4.4. Interpretation of the optimal convex combination parameter

From Eq. (21) we can come to

$$\alpha_o=\frac{1}{1+(\sigma_x^2-\mathbf{c_o}^T\mathbf{R}_{\mathbf{yy}}\mathbf{c_o}/\sigma_\eta^2)}. \tag{24}$$

Using (2), (4) and (23) in Eq. (24) we obtain

$$\alpha_o=\frac{1}{1+(1/G_{LP})}, \tag{25}$$

where G_{LP} is the FLP noise reduction gain in relation to the original contaminated signal, as defined in Eq. (13). A gain $G_{LP}>1$ indicates that the FLP MSPE is smaller than the MSPE obtained by using the unprocessed signal. The parameter G_{LP} can also be expressed as

$$G_{LP}=\frac{(\sigma_x^2/(\sigma_{x_N}^2+\sigma_{\eta_P}^2))}{(\sigma_x^2/\sigma_\eta^2)}=\frac{\mathrm{SNR}_{\hat{x}_L}}{\mathrm{SNR}_y}, \tag{26}$$

i.e. it is the ratio between the signal-to-prediction-error ratios for the FLP-processed signal and for unprocessed speech. For noiseless speech, $\mathrm{SNR}_y\to\infty$. Then $G_{LP}\to0$ and $\alpha_o\to0$; thus, the output of the proposed structure is $\hat{x}(n)=y(n)$. In case there is only noise ($\mathrm{SNR}_y\to0$), then $G_{LP}\to\infty$, $\alpha_o\to1$ and thus $\hat{x}(n)=\hat{x}_L(n)$. When SNR_y is finite and nonzero:

$$\begin{aligned}&\mathrm{SNR}_{\hat{x}_L}\to\infty \quad\Rightarrow\quad \alpha_o\to1 \quad\Rightarrow\quad \hat{x}(n)\cong\hat{x}_L(n)\\&\mathrm{SNR}_{\hat{x}_L}\to0 \quad\Rightarrow\quad \alpha_o\to0 \quad\Rightarrow\quad \hat{x}(n)\cong y(n)\end{aligned} \tag{27}$$

The first case in Eq. (27) is characteristic of voiced utterances (in which FLP can adequately estimate speech signals), while the second one refers to unvoiced utterances. For unvoiced speech utterances contaminated by uncorrelated noise $\hat{x}(n)=(1/(1+1/\mathrm{SNR}))\times y(n)$ and, as a consequence, the output of the system will be a scaled version of the input (volume control) as a function of the signal to noise ratio.

4.5. Robustness against correlated noise

As pointed out by Ref. [19] there are daily-life situations in which hearing-aid users are preponderantly subjected to broadband noise. However, it is recognized that the majority of situations are characterized by correlated noise. So, it is mandatory to assess the robustness of the proposed method under such condition. Assuming now that the acoustic input signal $y(n)$ is given by a mixture of speech $s(n)$, large correlation-length (e.g. speech-like) noise $v(n)$, and small correlation-length noise $\eta(n)$, we have $y(n)=x(n)+\eta(n)$ where $x(n)=s(n)+v(n)$. Assuming $v(n)$ is independent of $s(n)$ and $\eta(n)$, and that $v(n)$ can be modelled by an AR process with time-varying parameters [37] (in a window of quasi-stationarity), although not exact, $x(n)$ can also be approximated by an AR process [38], whose coefficients can be estimated from the sum of the individual autocorrelation coefficients of $s(n)$ and $v(n)$ [39]. As a result, the FLP coefficient ($\mathbf{c}=[c_0\ c_1\ \dots\ c_{N-1}]^T$) behaviour can still be predicted by its well-established theory [34] by using $\mathbf{R_{xx}}=\mathbf{R_{ss}}+\mathbf{R_{vv}}$. Proceeding in the same way from Eqs. (15) to (25) leads to

$$\alpha_o=\frac{1}{1+((\sigma_{s_N}^2+\sigma_{v_N}^2)/\sigma_\eta^2)}, \tag{28}$$

where ${\sigma_{sN}}^2$ and ${\sigma_{vN}}^2$ are the variances of the unpredictable parts of $s(n)$ (${\sigma_s}^2={\sigma_{sP}}^2+{\sigma_{sN}}^2$) and $v(n)$ (${\sigma_v}^2={\sigma_{vP}}^2+{\sigma_{vN}}^2$), respectively. Eq. (28) generalizes Eq. (25) for ${\sigma_v}^2\neq0$, and, despite being a function of ${\sigma_{vN}}^2$, demonstrates robustness against additive correlated noise, once ${\sigma_{vN}}^2\ll{\sigma_v}^2$. As the power of the innovation of the correlated noise increases (${\sigma_{vN}}^2\to\infty$), α_o decreases ($\alpha_o\to0$), and the algorithm output is biased towards the unprocessed input signal $\hat{x}(n)\to y(n)$ (in this case speech is totally preserved, and the effort to reduce noise must be performed by the hearing-aid correlated-noise-reduction unit).

5. Adaptive implementation

Several adaptive strategies can be used to track the FLP optimum solution presented in Eq. (5) for quasi-stationary signals. As each new sample is made available, the chosen algorithm calculates a new set of predictor coefficients. This mechanism

permits the tracking of statistical characteristics of the involved signals. LMS, NLMS, and Leaky-LMS algorithms are examples of very low computational cost strategies. Variable convergence speed control techniques adapted to the speech signal nature [40] can also be applied to improve robustness and performance.

The conventional linear adaptive predictor is based on the popular LMS algorithm and its update equation is given by

$$\mathbf{c}(n+1)=\mathbf{c}(n)+\mu e(n)\mathbf{y}(n-\Delta), \tag{29}$$

where $e(n)=y(n)-\hat{x}_L(n)=y(n)-\mathbf{c}^T(n)\mathbf{y}(n-\Delta)$ is the prediction error and μ is the convergence step. This choice is computationally simple and facilitates the understanding of the proposed noise reduction scheme. In typical CLAP noise cancellers, delay Δ varies continuously according to the instantaneous pitch of speech. Here, a fixed Δ is considered for simplicity.

Assuming convergence of Eq. (29), it is well known that the LMS steady-state mean-weight vector equals the optimal solution (5). Thus [34],

$$\lim_{n\to\infty} E\{\mathbf{c}(n)\}=\mathbf{R}_{\mathbf{yy}}^{-1}\mathbf{r}_{x\mathbf{x}_\Delta}=\mathbf{c_o} \tag{30}$$

5.1. Practical implementation issues

In order to obtain a real-time dynamic approximation to the optimum parameter α_o, in real conditions, the following estimator can be used (see Eq. (21))

$$\hat{\alpha}_o(n)=\frac{\sigma_\eta^2(n)}{\sigma_y^2(n)-\sigma_{\hat{x}_L}^2(n)}, \tag{31}$$

where ${\sigma_\eta}^2(n)$ is an estimate of the instantaneous additive noise power, which can be estimated when a voice activity detector (VAD) indicates absence of speech, ${\sigma_y}^2(n)$ is an estimate of the input signal power, and $\sigma_{\hat{x}_L}^2(n)$ is an estimate of the FLP output power. These estimates can be obtained in different ways [41]. One very simple strategy can be implemented by using two recursive first order low-pass filters [17], given by

$$\sigma_w^2(n)=\tau_w\sigma_w^2(n-1)+(1-\tau_w)w^2(n) \tag{32}$$

where $w^2(n)$ represents $e^2(n)$ in the absence of speech, and for the whole time, respectively $({\sigma_e}^2={\sigma_y}^2-\sigma_{\hat{x}_L}^2)$. Distinct attack and release time constants can be used. Eq. (32) has unit gain at zero Hertz and a time constant (in s) given by

$$\tau=\frac{T_{samp}}{\ln(\tau_w^{-1})}, \tag{33}$$

where T_{samp} is the sampling period.

5.2. Robustness against correlated noise

Following the same assumptions presented in Section 4.5, using $\mathbf{R_{xx}}=\mathbf{R_{ss}}+\mathbf{R_{vv}}$ in Refs. [33], [42] and [43] permits predictions of steady-state and transient performance of the first- and second-order moments of the adaptive coefficients presented in Eq. (29).

Assuming convergence of the estimators in Eqs. (32) and (31) tends to

$$\hat{\alpha}_o=\frac{1}{1+(\sigma_{s_N}^2/(\sigma_\eta^2+\sigma_{v_N}^2))}. \tag{34}$$

Eq. (34) shows that the proposed α_o estimator presented in Eqs. (31) and (32) is robust to ${\sigma_v}^2$, once ${\sigma_{vN}}^2 \ll {\sigma_v}^2$. Here, differently from Eq. (28), an increasing of ${\sigma_{vN}}^2$ leads to an increasing in $\hat{\alpha}_o$. For ${\sigma_{vN}}^2\to\infty$, then $\hat{x}(n)\to\hat{x}_L(n)$, and the uncorrelated components of both $s(n)$ and $v(n)$ are suppressed. In order to avoid that, estimations of $\sigma_{\hat{x}_L}^2(n)$ during voice absence $(={\sigma_{vP}}^2)$ should be monitored. This information can be used to shut the algorithm down $(\hat{\alpha}_o\to 0)$ under the existence of extremely unfavourable conditions $(\sigma_{v_p}^2(n)>\kappa$, $\kappa\in\Re^+)$.

A step by step description of the implementation of the proposed algorithm can be found in Table 1 (in which τ_a and τ_r are respectively the attack and release time constants for the denominator of Eq. (31); τ_η is the time constant for the numerator of Eq. (31); τ_v is the time constant for $\hat{x}_L^2(n)$; and a limiter avoids large estimation errors). Compared to CLAP the extra computational cost of the proposed method has only 10 multiplications, 5 sums and 1 division per iteration.

6. Results

This section presents simulations and application examples to illustrate the performance of the proposed algorithm and corroborate the theoretical results obtained in the previous sections.

Table 1
Proposed algorithm: implementation steps and computational complexity.

Equation	Comment	Complexity
$\mathbf{y}(n-\Delta)=[\,y(n-\Delta)\ \cdots\ y(n-\Delta-N+1)\,]^T$	Input vector	None
$\hat{x}_L(n)=\mathbf{c}^T(n)\mathbf{y}(n-\Delta)$	CLAP output	N MUL, (N-1) SUM
$e(n)=y(n-\Delta)-\hat{x}_L(n)$	CLAP error	1 SUM
$\mathbf{c}(n+1)=\mathbf{c}(n)+\mu e(n)\mathbf{y}(n-\Delta)$	CLAP update equation	(N+1) MUL, N SUM
if (VAD = 0)	Noise power	3 MUL, 1 SUM
$\sigma_\eta^2(n)=\tau_\eta\sigma_\eta^2(n-1)+(1-\tau_\eta)e^2(n)$		
$\sigma_{\hat{x}_L}^2(n)=\tau_v\sigma_{\hat{x}_L}^2(n-1)+(1-\tau_v)\hat{x}_L^2(n)$		
		3 MUL, 1 SUM
if $(z(n)>\sigma_z^2(n))$	Denominator of Eq. (31)	2 MUL, 1 SUM
then $\sigma_z^2(n)=\tau_a\sigma_z^2(n-1)+(1-\tau_a)e^2(n)$		
else $\sigma_z^2(n)=\tau_r\sigma_z^2(n-1)+(1-\tau_r)e^2(n)$		
if $(\sigma_z^2(n)>\sigma_\eta^2(n))$ then $\sigma_z^2(n)=\sigma_\eta^2(n)$	Limiter	None
if $(\sigma_{\hat{x}_L}^2(n)<\kappa)$	Eq. (31)	1 DIV
then $\hat{\alpha}_o(n)=\sigma_\eta^2(n)/\sigma_z^2(n)$		
else $\hat{\alpha}_o(n)=0$		
$\hat{x}(n)=(1-\hat{\alpha}_o(n))y(n)+\hat{\alpha}_o(n)\hat{x}_L(n)$	Eq. (16)	2 MUL, 2 SUM

6.1. Simulated signals

Initially, the proposed method and CLAP were compared under completely known conditions to quantify the convex combination coefficient α and SNR influences. Two artificial input signals were used: (1) a simulated unvoiced utterance modelled by a 22 order autoregressive (AR) process [44], obtained by applying Burg's method to a 22 millisecond real speech epoch of the phoneme /s/; (2) a simulated voiced sound obtained in the same way as described before, but originated from the phoneme /a/. Both phonemes were produced by a male speaker. The sampling frequency was 15.625 kHz, and an ideal VAD was used to signal the beginning and ending of each utterance. Fig. 3 shows that the correlation-lengths are significantly different for these signals. The number of adaptive coefficients was $N=10$, the convergence step $\mu=10^{-6}$, delay $\Delta=1$, additive noise was white Gaussian and the SNRs used were (iv) 0.4, (iii) 3, (ii) 10.4, and (i) 20 dB. The optimal convex combination parameter α_o was calculated from the available theoretical statistical parameters. Figs. 4 and 5 show the results obtained. The plotted curves show the following MSPEs: (a) $E\{[x(n)-y(n)]^2\}=\sigma_\eta^2$ (dotted), (b) $E\{[x(n)-\hat{x}_L(n)]^2\}$ (dashed), and (c) $E\{[x(n)-\hat{x}(n)]^2\}$ (continuous). The CLAP output (dashed line) results in a higher MSPE than that obtained from the unprocessed signal for high SNRs (20 and 10.4 dB). This situation is reversed for low SNRs (0.4 and 3 dB). This results from the fact that CLAP not only reduces contamination noise but also uncorrelated components (unvoiced sounds) of speech. For high SNRs, the noise reduction savings obtained by CLAP are overcome by the inserted amount of speech distortion. The optimal coefficient α_o, calculated from Eq. (21), is shown as an asterisk and clearly coincides with the point of minimum of the proposed method curve for all SNRs. In addition, α_o results in an MSPE smaller than or equal to that obtained by CLAP or unprocessed signal. It can also be seen that the value of α_o increases as SNR decreases. For SNR > 20 dB then $\alpha_o \cong 0$, indicating that the best speech signal estimate is basically obtained from the unprocessed signal. On the other hand, $\alpha_o \cong 1$ for SNR < 0 dB, indicating that the best estimate is obtained from the CLAP output. In the 3 dB < SNR < 10 dB range, there is a wide set of values of α around α_o that results in smaller MSPEs than those produced by CLAP or the unprocessed signal. These results indicate the robustness to errors of the proposed method in estimating α_o.

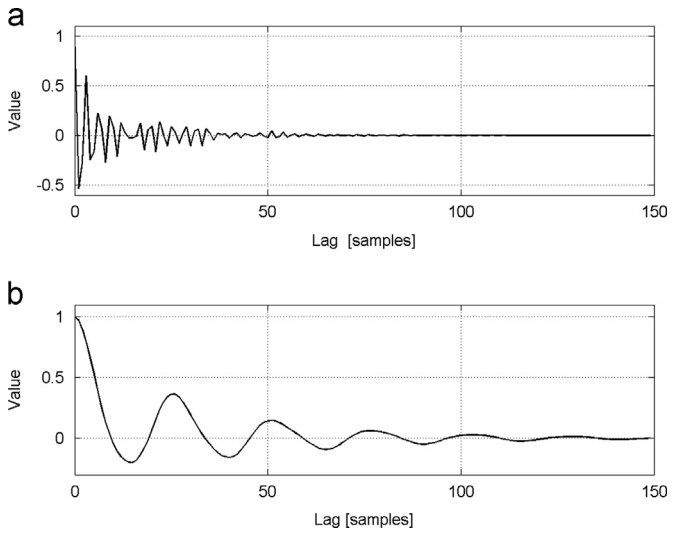

Fig. 3. Autocorrelation functions of the simulated speech sounds. (a) unvoiced sound, and (b) voiced sound.

6.2. Synthetic signals

The second example made use of a synthetic speech input-signal designed for telephonometry applications [45]. Despite its stationary behaviour, it has spectral characteristics similar to those found in natural speech. Additive artificial noise was white Gaussian and SNR = −3, 0, 3, 10, and 20 dB (evaluated only during speech occurrence). The parameters used were the same as those in the first example, except sampling frequency, which was set to 16 kHz. An ideal VAD was used. The optimal convex combination parameter was calculated previously, using the whole signals, and

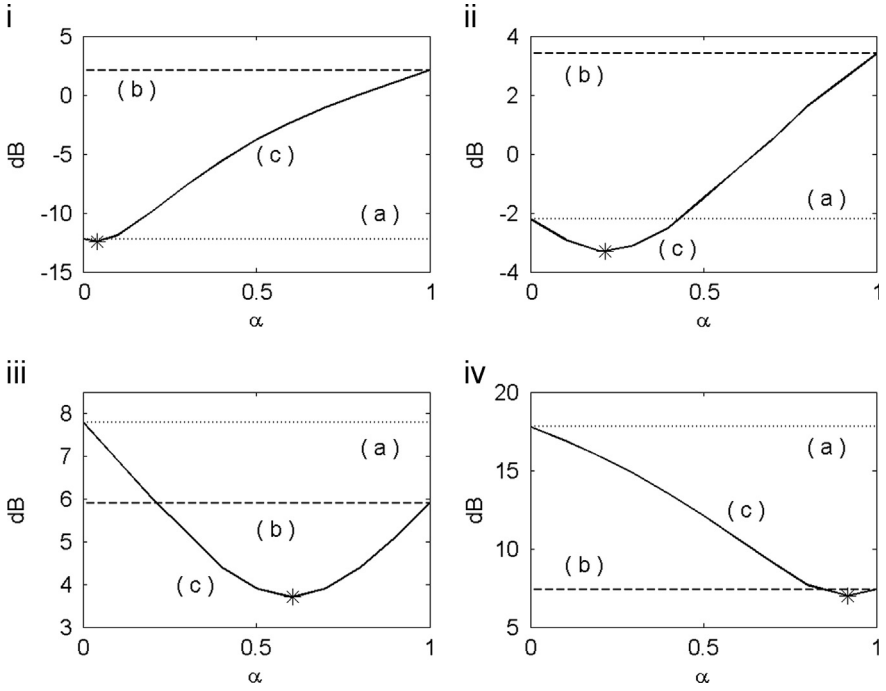

Fig. 4. MSPE for simulated unvoiced sound. (i) SNR=20, (ii) 10.4, (iii) 3, and (iv) 0.4 dB. (a) $E\{[x(n)-y(n)]^2\}=\sigma_\eta^2$ (dotted), (b) $E\{[x(n)-\hat{x}_L(n)]^2\}$ (dashed), (c) $E\{[x(n)-\hat{x}(n)]^2\}$ (continuous). The asterisk shows α_o (see Eq. (21)).

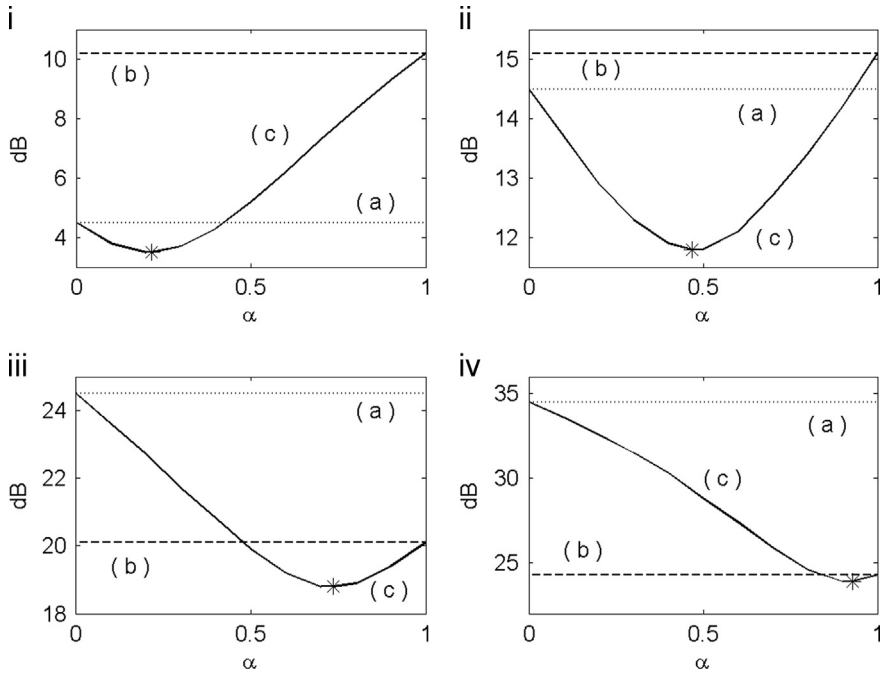

Fig. 5. MSPE for simulated voiced sound. (i) SNR=20, (ii) 10.4, (iii) 3, and (iv) 0.4 dB. (a) $E\{[x(n)-y(n)]^2\}=\sigma_\eta^2$ (dotted), (b) $E\{[x(n)-\hat{x}_L(n)]^2\}$ (dashed), (c) $E\{[x(n)-\hat{x}(n)]^2\}$ (continuous). The asterisk shows α_o (see Eq. (21)).

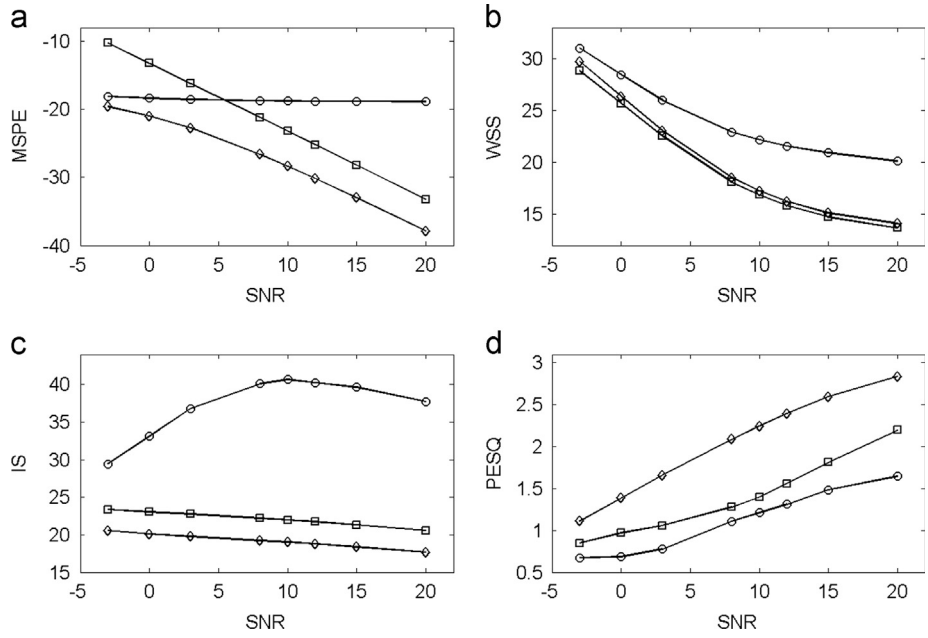

Fig. 6. Average score according to (a) MSPE (dB), (b) WSS, (c) IS, and (d) PESQ for the: (□) unprocessed signal, (○) signal processed by CLAP, and (◇) signal processed by the proposed method. Synthetic (telephonometry) signal and additive white Gaussian noise.

was kept fixed during all signal processing. Four objective quality criteria were used to quantify speech estimate performance, namely: mean-square prediction-error (MSPE), weighted spectral slope (WSS), Itakura–Saito measure (IS) and broadband (16 kHz) perceptual evaluation of speech quality measure (PESQ) [16]. MSPE, WSS and IS present smaller indices for higher quality signals while the PESQ relates a higher index to a higher quality. As expected, Fig. 6 consistently shows a smaller MSPE for the proposed method. CLAP MSPE is practically constant, suggesting that the whole amount of uncorrelated components is filtered out independently of its nature (speech or noise), again as expected. As a result, CLAP presents good performance only for small SNRs, whereas SNRs bigger than 5 dB result in MSPEs higher than those obtained from unprocessed signals. This occurs due to the attenuation of the uncorrelated part of the speech. Such being the case, quality improvements due to noise reduction are overcome by the degradation that results from speech distortion. Fig. 6b shows that the proposed method presents WSS indexes comparable to those provided by unprocessed speech. These values are significantly smaller than those obtained from CLAP. This can be explained by the fact that WSS penalizes large distances in the spectral peak locations (formants), minimizing tilt and overall level differences. From this, it can be inferred that, due to a nonzero convergence step, the adaptation and tracking processes lead to spectral peak distortions, degrading speech quality. The

proposed method minimizes this effect, decreasing the coefficient fluctuation contribution. Fig. 6c shows that, according to the IS index, the proposed method results in significantly better clean speech estimates. IS penalizes spectral global level differences between the analyzed and the desired signals. Although this is not a very interesting feature in psychoacoustics terms (studies have shown that differences in spectral level have little effect on subjective speech quality), the results obtained clearly indicate a decrease in contamination/distortion levels. The very high speech quality improvement indicated by the IS index for signals processed by the proposed method is probably related to the high uncorrelated signal content of the synthetic signal used in telephonometric applications. With real speech signals, quality improvements are still expected to be significant in relation to the CLAP and the results obtained from unprocessed signals, but moderate as compared to those obtained with the synthetic signal. CLAP presented the highest IS indexes (poor quality) for all tested SNRs. Fig. 6d shows comparative PESQ results. The PESQ criterion measures distortions commonly found in telecommunication nets (loss of packets, delays, CODEC distortions) and presents a high correlation with the subjective Mean Opinion Test. The results obtained show a consistent quality degradation of the speech processed by CLAP when compared to unprocessed speech (the same phenomenon observed in WSS and IS criteria), and a significant quality increase for signals processed by the proposed method.

6.3. Speech signals

The next two examples (Figs. 7 and 8) made use of real (nonstationary) speech signals recorded by six (three male and three female) speakers with different voice characteristics. Both were analyzed under different SNRs (from −6 dB – extremely annoying subjective condition to 36 dB – low contamination level, in steps of 3 dB). Results from the well-known Ephraim–Malah (EM) technique [13] are also presented. Comparisons with the EM method were provided so as to permit a general idea on the relative performance of the proposed method in relation to other more complex techniques. More expensive computational techniques, such as those based in super-Gaussian methods [46,47], provide higher levels of noise reduction and speech quality but at higher computational cost and a larger global propagation delay (usually not tolerated by hearing-aid users). Figs. 7 and 8 present the same symbol convention as Fig. 6, but with the addition of EM results represented by asterisks connected by a continuous line. Curves in Figs. 7 and 8 result from the average of 50 and 15 runs (with different epochs of the same noise signals), respectively, in order to obtain smooth curves.

In the first experiment (Fig. 7), the contamination noise was white Gaussian noise, the delay was $\Delta=1$, $\mu=0.1$, $N=100$ coefficients, $\tau_\eta=0.999$, $\tau_a=0.99$, and $\tau_r=0.997$. A real VAD [48] was used to show the viability of the proposed algorithm in practical applications. Careful visual inspection permitted the author to conclude that this VAD produced accurate estimates of speech activity for SNR > 6 dB. Fig. 7 refers to a male speaker and the results obtained corroborate the same basic tendencies observed in Fig. 6 for MSPE, WSS and PESQ. Clearly, and as expected due to its higher complexity, the Ephraim–Malah technique presented the best quality results according to PESQ (for SNR > 30 dB the unprocessed signal presents a very good quality and, as a result, any processing will degrade it). However, for medium to high SNR the EM intrinsic nonlinear processing produced higher MSPE, WSS and IS indexes when compared to the unprocessed signal. CLAP inferior performance can be attributed to coefficient fluctuation and musical noise effect (characterized by spurious peaks in the spectrum and excessive attenuation of uncorrelated signals). Compared to the CLAP and EM, the proposed technique presented a superior MSPE performance for SNR > 6 dB. Compared to the unprocessed signal and CLAP, the presented example suggests that the proposed technique can improve PESQ scores in the range of 6 dB < SNR < 30 dB. For SNR > 30 dB the proposed method presented the same PESQ indexes as those of the EM and CLAP techniques. Despite some variability (due to speech content and individual pitch), similar results were obtained for all six speakers. Fig. 7d suggests that the performance of the proposed method could only be comparable to that of EM in high SNR conditions.

Fig. 8 presents the results of the experiment in which a real acoustic (nonstationary) wind-noise, sampled at near-Nyquist frequency rate, with a small correlation-length (whose normalized autocorrelation decays below ± 0.05 after 18 lags and below ± 0.03 after 65 lags) was used to additively contaminate the (real) speech input signal. Due to the non-white noise statistical characteristic, a delay of $\Delta=20$ was used. Roughly, the same basic

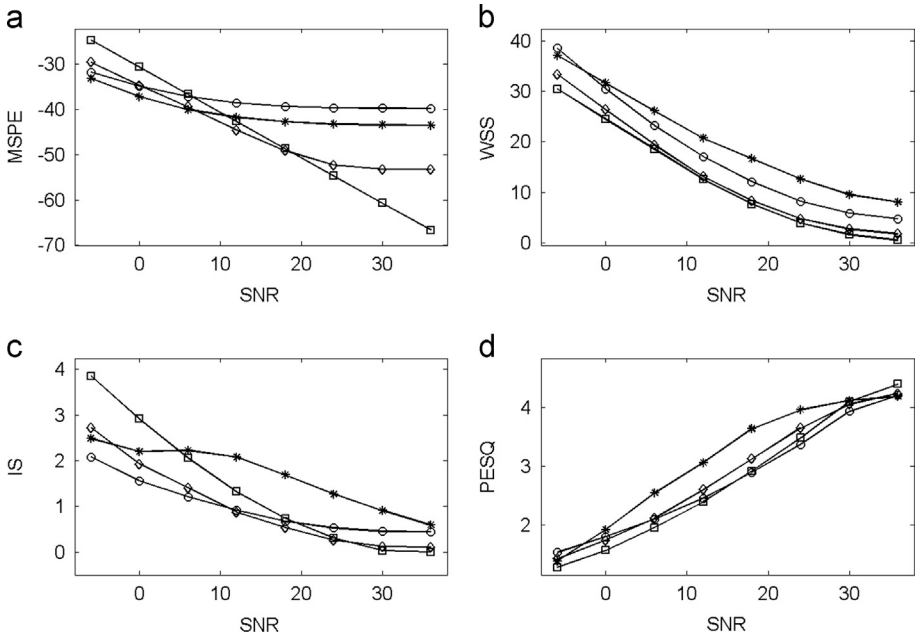

Fig. 7. Average score according to (a) MSPE (dB), (b) WSS, (c) IS, and (d) PESQ for the: (▫) unprocessed signal, (○) signal processed by CLAP, (◇) signal processed by the proposed method, and (∗) signal processed by the EM technique [13]. Real speech signal (male speaker) and white Gaussian noise.

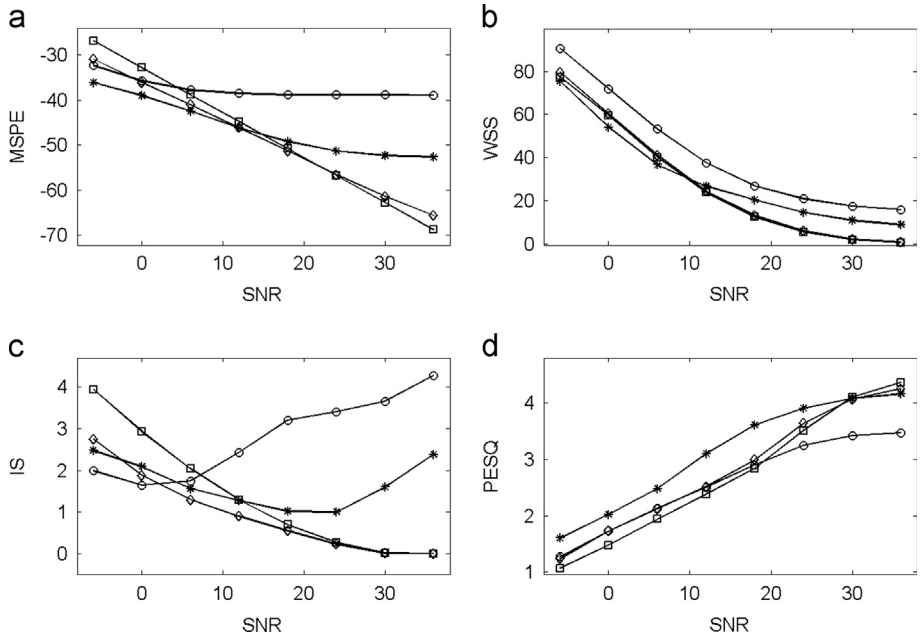

Fig. 8. Average score according to (a) MSPE (dB), (b) WSS, (c) IS, and (d) PESQ for the: (□) unprocessed signal, (○) signal processed by CLAP, (◇) signal processed by the proposed method, and (∗) signal processed by the EM technique [13]. Real speech signal (female speaker) and small correlation-length (nonstationary) real wind-noise.

results showed in Figs. 6 and 7 were obtained. It is worth pointing out that for SNR < 30 dB the proposed method resulted in PESQ indexes higher than those obtained by both CLAP and unprocessed signal. The PESQ results also suggest that the proposed method can not only alleviate speech distortions produced by CLAP but also improve the speech quality in medium to high SNR range (18 dB < SNR < 27 dB). For SNR > 30 dB speech distortion introduced by all methods (EM, CLAP and proposed one) resulted in lower PESQ than that of the unprocessed signal. Experiments showed that, as the delay Δ was increased, there was a decrease in the processed signal quality due to an increased loss of some low-correlated intrinsic speech components. The delay choice is directly associated with a trade-off between the desired noise reduction and the maintenance of the speech quality obtained by the prediction process. Despite the superiority of the EM technique, its implementation requires computational resources and processing time delays, sometimes unavailable in some hearing-aids. Thus, the proposed method aims to obtain a performance improvement in adaptive prediction-based systems with a minimum increase in complexity. Despite using a real noise signal, this example dealt with a mild nonstationary condition; severe nonstationarities, such as abrupt changes in noise power level, may affect the tracking of the optimum convex combination parameter, resulting in undesirable power fluctuations at the output, and requiring more elaborated strategies.

6.4. Subjective analysis

A preliminary Degradation Category Rating Test with eight volunteers was performed to assess the subjective performance of the proposed method. In this experiment two different sentences (in Brazilian Portuguese: “os pesquisadores acreditam nessa teoria” and “ela saía discretamente” meaning, respectively, “the researchers believe in this theory” and “she went out quietly”) uttered by two individuals (one male and one female) in a 125 m³ (free volume) semi-anechoic chamber were acquired by a microphone Le Son ML-70S and a hand-held professional digital recorder (Microtrack-II model 2496 from M-Audio) with a sampling frequency of 16 kHz. Beginning and end of speech segments were manually annotated emulating an ideal VAD. Speech signals were contaminated by a nonstationary small-correlation-length artificial noise (whose normalized autocorrelation decays below ± 0.05 after 6 to 17 samples), resulting in twelve contaminated speech signals (6 male and 6 female) with SNRs equal to 18, 21, 24, 27, 30 and 33 dB. Such contamination levels can be distinguished by untrained listeners and were chosen in ord er to represent conditions found in daily-life, no intelligibility problems were reported. The used parameters were a delay of $\Delta = 20$ and $N = 100$ coefficients. For each signal, the following sound files were generated: (a) clean speech; (b) contaminated speech (DIRTY); (c) contaminated signal processed by a narrowband noise reduction system (NNRS); (d) contaminated signal processed by CLAP followed by a NNRS processing (CLAPN); and (d) contaminated signal processed by the proposed method followed by NNRS processing (NEWN). The narrowband noise reduction system is a software-based proprietary noise-reduction algorithm (Acústica Amplivox Company Ltda), especially designed for hearing-aid devices, consisting of a filter-bank architecture similar to the one described in Ref. [17]. The purpose of including such a system is to permit the assessment of the noise reduction system global performance in real hearing-aid applications, since narrowband and broadband noise reduction must be both performed in order to obtain adequate sound quality and acceptability by the user. NNRS reduces small amounts of side-effect speech distortions caused by the adaptive predictor (due to coefficient fluctuations and musical noise) which could result in unsatisfactory performance. Eight volunteers (four male and four female) without hearing complaints were selected for subjective evaluation of the sound files. Each volunteer was instructed to comparatively quantify the subjective quality of each set (out of twelve) of five speech files (related to the same sentence and SNR) in a continuous scale from -5 (worst) to 5 (best) where the midscale (zero) was associated with the clean (uncontaminated) speech file. The results obtained are shown in box and whisker diagrams where the set of sample values comprised between the lower and upper quartiles (denominated by q_1 and q_3, respectively) is represented by a rectangle whose median is indicated by a bar. The sample values are considered outliers when greater than $q_3 + \varpi \cdot (q_3 - q_1)$ or less than $q_3 - \varpi \cdot (q_3 - q_1)$, whereas q_1 and q_3 are defined as the percentage values of 25 and 75%, respectively. Variable ϖ is defined as the default value of 1.5 [49] and represents the upper and lower extremes, which are not considered outliers. The vertical axis indicates subjective satisfaction and the horizontal one presents

the associated type of processing. The plus signal (+) indicates the presence of an outlier.

Fig. 9 shows the results obtained for SNR=24, 27 and 30 dB. The horizontal dotted line indicates the clean speech score (fixed in zero). The median of contaminated speech was negative (worse quality than clean speech) for all tested SNR. Despite having the highest level of noise reduction, CLAPN presented score results close to the ones of the contaminated signal. This can be explained due to the undesired cancelling of unvoiced speech components and the associated quality degradation (see Section 6.1). Despite intelligibility maintenance, speech processed by CLAPN sounds muffled, displeasing normal hearing people due to its lack of naturalness. Apparently, normal hearing volunteers overvalued naturalness compared to noise reduction. This aspect must be taken into consideration and results of this work must be analyzed with care, since they could not be completely applied to hearing impaired people, due to significant differences in the sensitivity of the hearing system. For severe hearing-impaired people, an increase in CLAPN scores (as compared to contaminated speech) is expected [50]. Such characteristic seems to justify its use in hearing-aid applications.

The NEWN processing resulted in a consistent increase of speech quality when compared to CLAPN for all tested SNRs. This result agrees with those presented in Section 6.1. Comparisons with single NNRS processing showed that the proposed algorithm produces a better subjective quality for SNR $\geq$ 24 dB (for SNR=33 dB both NNRS and NEWN presented similar performances since the noise contamination level is very low). Still in the SNR $\geq$ 24 dB range, the NEWN frequently resulted in higher scores than those associated with clean signals. This occurred due to the reduction of audible microphone electric noise associated with speech recordings, especially during voiced sound periods. For SNR < 24 dB the NEWN median score was lower than the median of the contaminated speech. This probably happened because of predictor coefficient fluctuations (due to the tracking process) associated with high contamination noise levels (see Fig. 6b and text explanation in Section 6.2). Interviews with volunteers after the experiments confirmed that the main complaint on the quality of the speech processed by NEWN, that originated the low scores given by some volunteers (especially for SNR < 24 dB), refers to the (side-effect) subjective sensation of low-frequency amplitude modulation of speech along the sentences. This probably results from large variance estimates of the instantaneous additive noise power, input signal power, and FLP output power, which lead to significant $\alpha_o(n)$ fluctuations. One possible strategy to alleviate this problem would be low-pass filter Eq. (31).

Fig. 10 shows spectrograms of one experiment in which a male speech was contaminated with SNR=27 dB. Hot colours mean higher magnitudes while cold colours mean lower ones. Clean speech is presented in Fig. 10a. Contaminated speech is presented in Fig. 10b. Comparisons between Fig. 10a and Fig. 10b show an increase in the noise floor after contamination especially noticeable during pauses (the original blue levels in Fig. 10a have changed to green in Fig. 10b). Similarity between noise floors in Fig. 10b and Fig. 10c indicates the lack of capability of NNRS to reduce broadband noise. CLAPN result is shown in Fig. 10d. It restored the original noise floor during pauses but also excessively reduced high-frequency speech components during unvoiced utterances, originating a muffling sensation. NEWN, shown in Fig. 10e, apparently restored the original background noise floor, especially during pauses, with a small noticeable distortion of the original spectrum in relation to the

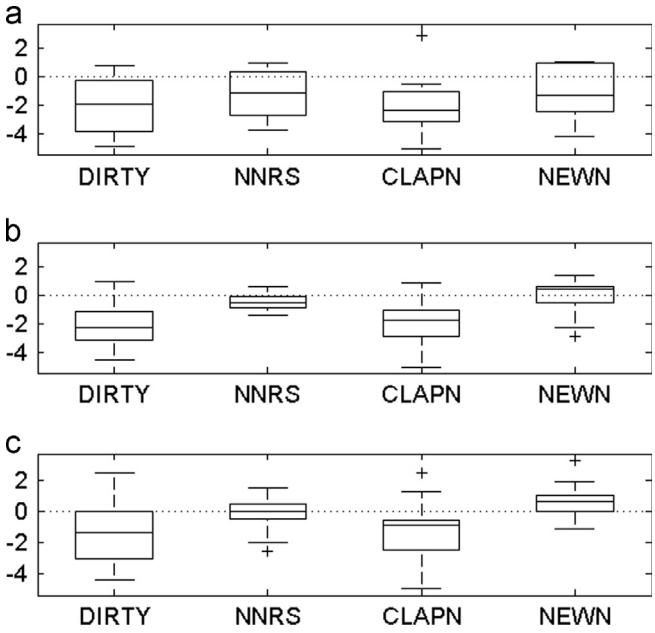

Fig. 9. Subjective evaluation of speech signals contaminated by a small correlation-length nonstationary artificial noise. SNR: (a) 24; (b) 27, and (c) 30 dB. Contaminated speech (DIRTY); contaminated speech processed by the narrowband noise reduction system (NNRS); contaminated speech processed by CLAP followed by NNRS (CLAPN); and contaminated speech processed by the proposed method followed by NNRS (NEWN). The dotted line represents the clean speech score.

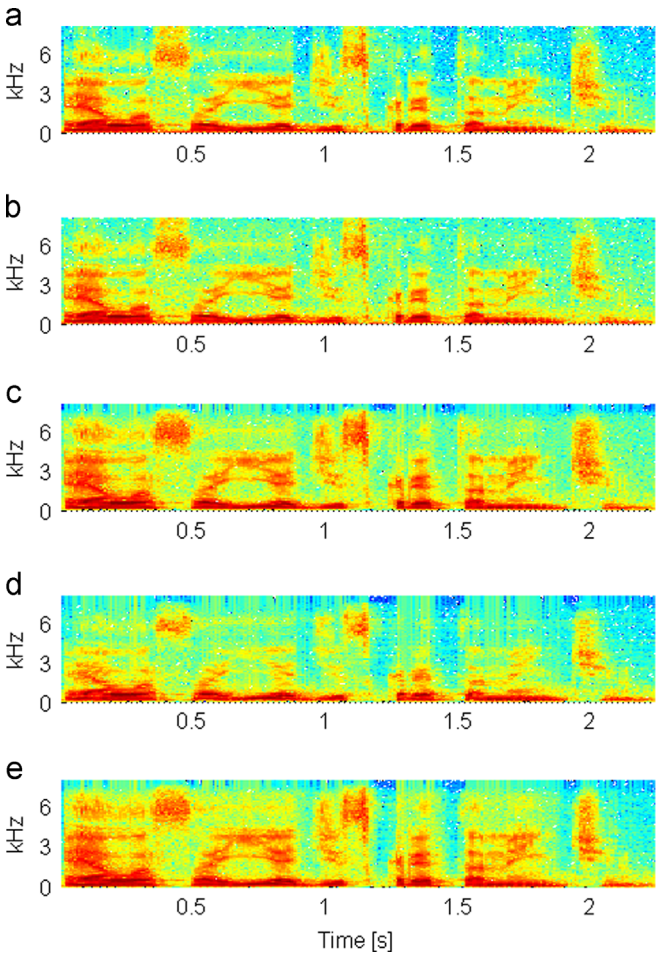

Fig. 10. Spectrograms of a real speech signal (male speaker) contaminated by a small correlation-length nonstationary artificial noise for SNR=27 dB. (a) clean speech; (b) contaminated speech; (c) contaminated speech processed by the narrowband noise reduction system (NNRS); (d) contaminated speech processed by CLAP followed by NNRS (CLAPN); (e) contaminated speech processed by the proposed method followed by NNRS (NEWN). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

NNRS. Speech spectrum variations resulting from NEWN processing, as compared to uncontaminated speech, can be attributed to the intrinsic trade-off between noise reduction and speech distortion.

6.5. Robustness to correlated noise

Finally, simulations were performed to demonstrate the behaviour of the adaptive algorithm from a holistic viewpoint. Mixtures of speech $s(n)$, small correlated-length noise $\eta(n)$ and speech-like noise $v(n)$ were processed by the proposed adaptive algorithm, followed by the narrowband noise reduction system described in the previous section. Here, results for a male speaker, a speech-like noise generated from the AR model presented in Section 6.1 (phoneme /a/), and artificial white noise are shown. The algorithm used the same parameters as described in the first experiment of Section 6.3. Each plot in Fig. 11 refers to a different $\mathrm{SNR_{UNC}}$ ($=-10$, 0, 10, 20, 30, ∞) that defines the signal to noise ratio between speech and white noise ($\mathrm{SNR_{UNC}}=10\cdot\log_{10}(\sigma_s^2/\sigma_\eta^2)$). The abscissa of each plot refers to the ratio between speech and speech-like noise powers ($\mathrm{SNR_{COR}}=10\cdot\log_{10}(\sigma_s^2/\sigma_v^2)$), while the ordinate refers to PESQ quality of the output signal. Speech power is kept fixed in all simulations. Results from the Ephraim–Malah technique [13], CLAP output processed by the NNRS, and contaminated speech are also presented. Analysis of Fig. 11 clearly

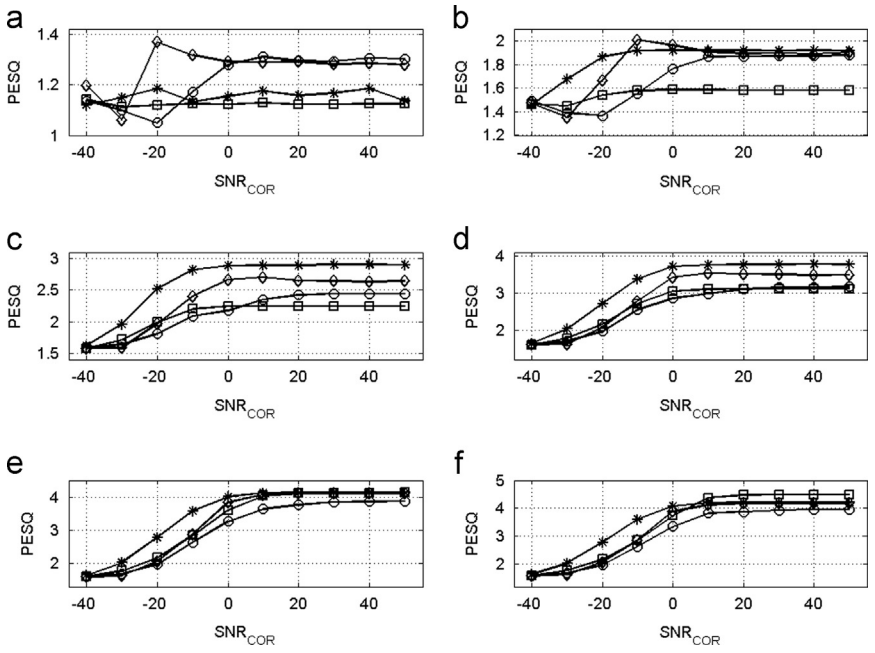

Fig. 11. PESQ for a mixture of speech, small-correlated and speech-like noise. (a) $\mathrm{SNR_{UNC}}=-10$; (b) $\mathrm{SNR_{UNC}}=0$; (c) $\mathrm{SNR_{UNC}}=10$; (d) $\mathrm{SNR_{UNC}}=20$; (e) $\mathrm{SNR_{UNC}}=30$ dB; (f) $\mathrm{SNR_{UNC}}=\infty$ dB. Ephraim–Malah technique [13] (∗), CLAP followed by NNRS (o), contaminated speech (□), proposed method followed by NNRS (◇).

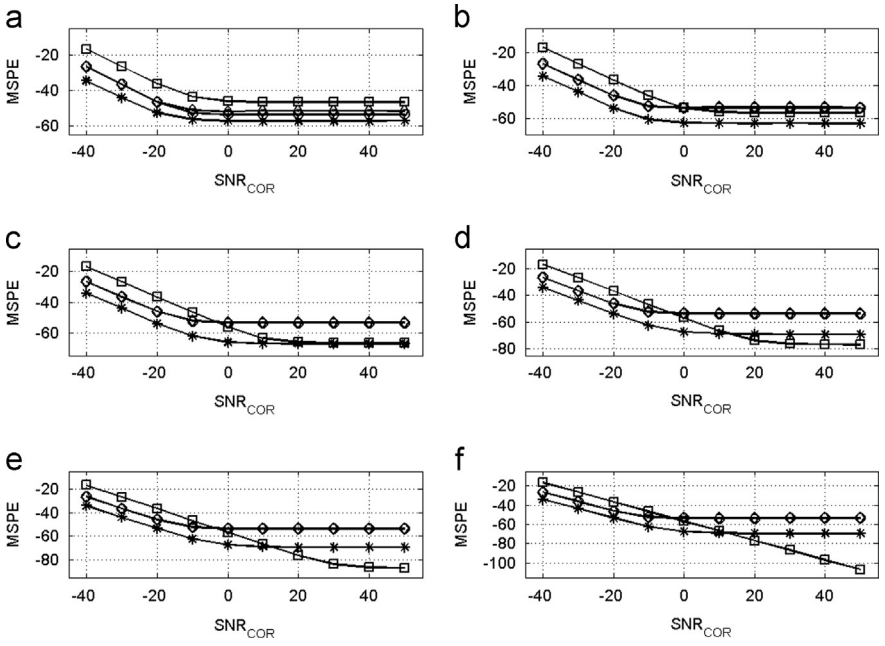

Fig. 12. MSPE for a mixture of speech, small-correlated and speech-like noise. (a) $\mathrm{SNR_{UNC}}=-10$; (b) $\mathrm{SNR_{UNC}}=0$; (c) $\mathrm{SNR_{UNC}}=10$; (d) $\mathrm{SNR_{UNC}}=20$; (e) $\mathrm{SNR_{UNC}}=30$ dB; (f) $\mathrm{SNR_{UNC}}=\infty$ dB. Ephraim–Malah technique [13] (∗), CLAP followed by NNRS (o), contaminated speech (□), proposed method followed by NNRS (◇).

indicates that the performance of the proposed algorithm does not collapse under correlated noise influence. The proposed technique showed to be robust to correlated noise for $SNR_{COR} > 0$, and presented a better performance than the CLAP for $SNR_{COR} \geq -20$ in all SNR_{UNC} scenarios.

Fig. 12 refers to the MSPE behaviour for the same signals presented in Fig. 11. It shows that the proposed technique basically presents the same MSPE of the CLAP for $SNR_{UNC} \geq 0$ dB.

The results reported here indicate the potentiality and practical applicability of the proposed method and its adaptive low-cost implementation, as well as the usefulness of the developed theoretical equations as guidelines for its optimum design. The results also suggest that linear prediction methods (like CLAP) could be used as complementary broadband noise reduction in hearing-aid applications once some strategies are introduced to alleviate the excessive cancelling of small-correlated portions of the speech. Such method, however, must be accompanied by a narrowband noise reduction system to filter out tracking effects (coefficient fluctuation) in adaptive implementations. The use of more complex adaptive methods and robust power estimators is expected to widen the SNR range in which speech quality improvement was obtained.

7. Conclusions

This work presented a complementary broadband noise reduction scheme for hearing-aids. The optimum setting for maximum performance was theoretically obtained, resulting in a smaller mean-square prediction-error as compared to the conventional linear predictor. Experiments with simulated and real signals corroborate the analytical results for a low-cost adaptive implementation of the proposed method. Four different objective quality measures indicate speech quality improvement when compared with the conventional adaptive predictor results. Preliminary Degradation Category Rating experiments corroborate the expected results when the proposed algorithm is followed by a narrowband noise reduction strategy in order to filter out the effects of undesirable coefficient fluctuations. Low-cost digital hearing-aids that make use of the conventional adaptive predictor for broadband noise reduction can be easily modified to incorporate the new proposal with a minimum amount of extra computational resources.

Conflict of interest statement

None declared.

Acknowledgements

The author would like to thank Alexandre Ferreira and Acústica Amplivox Company for bringing attention to this subject and providing their narrowband noise reduction routine, Prof. Pedro Fickel, Prof. José Carlos Bermudez, and the anonymous reviewers for their insightful comments and suggestions during the preparation of this manuscript. This work was supported by the Brazilian Ministry of Science and Technology (CNPq) under grants 559418/2008-6, and 303803/2009-6.

References

[1] K. Chung, Challenges and recent developments in hearing aids—Part I: Speech understanding in noise, microphone technologies and noise reduction algorithms, Trends Amplif. 8 (3) (2004) 83–124.
[2] H.H. Kim, D.M. Barrs, Hearing aids: a review of what's new, Otolaryngol. Head Neck Surg. 134 (2006) 1043–1050.
[3] T. Fillon, J. Prado, Evaluation of an ERB frequency scale noise reduction for hearing aids: a comparative study, Speech Commun. 39 (2003) 23–32.
[4] N.A. Whitmal, J.C. Rutledge, J. Cohen, Reducing correlated noise in digital hearing aids: a wavelet-based method for extracting speech from background noise, IEEE Eng. Med. Biol. (1996) 88–96.
[5] D.J. Schum, Noise Reduction in Hearing Aids: What Works and Why, News from Oticon (2003) 1–20.
[6] H.G. Mueller, J. Weber, B.W.Y. Hornsby, The effects of digital noise reduction on the acceptance of background noise, Trends Amplif. 10 (2) (2006) 83–94.
[7] N.A. Whitmal, J.C. Rutledget, Noise reduction in hearing aids: a case for wavelet-based methods, in: Proceedings of the International Conference of the IEEE Engineering in Medicine and Biology Society, 20, 3, 1998, pp. 1130-1135.
[8] J. Benesty, J. Chen, Y.A. Huang, Noise reduction algorithms in a generalized transform domain, IEEE Trans. Audio Speech Lang. Process. 17 (6) (2009) 1109–1123.
[9] J. Chen, J. Benesty, Y. Huang, S. Doclo, New insights into the noise reduction Wiener filter, IEEE Trans. Audio Speech Lang. Process. 14 (4) (2006) 1218–1234.
[10] J. Chen, J. Benesty, Y. Huang, Study of the noise reduction problem in the Karhunen–Loève expansion domain, IEEE Trans. Audio Speech Lang. Process. 17 (4) (2009) 787–802.
[11] J. Li, S. Sakamoto, S. Hongo, M. Akagi, Y. Suzuki, Adaptive b-order generalized spectral subtraction for speech enhancement, Signal Process. 88 (2008) 2674–2776.
[12] Y. Ephraim, D. Malah, Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator, IEEE Trans. Acoust. Speech Signal Process. 32 (6) (1984) 1109–1121.
[13] Y. Ephraim, D. Malah, Speech enhancement using a minimum mean-square error log-spectral amplitude estimator, IEEE Trans. Acoust. Speech Signal Process. 33 (2) (1985) 443–445.
[14] M. Marzinzik, B. Kollmeier, Speech pause detection for noise spectrum estimation by tracking power envelope dynamics, IEEE Trans. Speech Audio Process. 10 (2) (2002) 109–118.
[15] R. Bentler, L.-K. Chiou, Digital noise reduction: an overview, Trends Amplif. 10 (2) (2006) 67–82.
[16] P.C. Loizou, Speech Enhancement: Theory and Practice, CRC, 2007.
[17] Voyager, Time Domain Filter Bank, User's Manual, Information Note, Gennum (2005) 1–24.
[18] X. Fang, M.J. Nilsson, Noise Reduction Apparatus and Method, US Patent 6.757.395, 2004.
[19] J. Koopman, B.A.M. Franck, W.A. Dreschler, Toward a representative set of real-life noises, Audiology 40 (2001) 78–91.
[20] J.-H. Lin, P.-C. Li, S.-T. Tang, P.-T. Liu, S.-T. Young, Industrial wideband noise reduction for hearing aids using a headset with adaptive-feedback active noise cancellation, Med. Biol. Eng. Comput. 43 (2005) 739–745.
[21] D.J. Schum, Noise-reduction circuitry in hearing aids: (2) goals and current strategies, Hear. J. 56 (6) (2003) 32–40.
[22] M. Li, H.G. McAllister, N.D. Black, T.A. Pérez, Perceptual time–frequency subtraction algorithm for noise reduction in hearing aids, IEEE Trans. Biomed. Eng. 48 (9) (2001) 979–988.
[23] E. Verteletskaya, B. Simak, Noise reduction based on modified spectral subtraction method, Int. J. Comput. Sci. 38 (2011) 1–7.
[24] R. Sambur, Adaptive noise cancelling for speech signals, IEEE Trans. Acoust. Speech Signal Process. 26 (5) (1978) 419–423.
[25] I. Nakanishi, Y. Itoh, Y. Fukui, Noise reduction system based on frequency domain adaptive filter using modified DFT pair, IEEE Int. Symp. Circuits Syst. 11 (2000) 737–740.
[26] A. Kawamura, K. Fujii, Y. Itoh, Y. Fukui, A new noise reduction method using linear prediction error filter and adaptive digital filter, IEEE Int. Symp. Circuits Syst. 3 (2002) 488–491.
[27] G. Hu, D. Wang, Segregation of unvoiced speech from nonspeech interference, J. Acoust. Soc. Am. 124 (2) (2008) 1306–1319.
[28] G.P. Eatwell, Noise Reduction Filter, US Patent 5,742,694, 1998.
[29] A. Kawamura, Y. Iiguni, Y. Itoh, Noise reduction method based on linear prediction with variable step-size, IEICE Trans. Fundam. E88–A (4) (2005) 855–861.
[30] D. Giesbrecht, P. Hetherington, Periodic Signal Enhancement System, US Patent 7,680,652 B2, 2010.
[31] N. Westerlund, M. Dahl, I. Claesson, Speech enhancement for personal communication using an adaptive gain equalizer, Signal Process. 85 (2005) 1089–1101.
[32] D.G. Manolakis, V.K. Ingle, S.M. Kogon, Statistical and Adaptive Signal Processing, Spectral Estimation, Signal Modeling, Adaptive Filtering and Array Estimation, McGraw-Hill, 2000.
[33] J. Zeidler, Performance analysis of LMS adaptive prediction filters, Proc. IEEE 78 (12) (1990) 1781–1806.
[34] S. Haykin, Adaptive filter theory, Prentice-Hall, 2002.
[35] Y. Ephraim, H.L. van Trees, A signal subspace approach for speech enhancement, IEEE Trans. Speech Audio Process. 3 (4) (1995) 251–266.
[36] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press (2000) 24.
[37] M. Kuropatwinski, W.B. Kleijn, Estimation of the excitation variances of speech and noise AR-models for enhanced speech coding, in: IEEE International Conference on Acoustics, Speech, and Signal Processing, 2001, pp. 669–672.
[38] J.L. Noga, Bayesian State-space Modeling of Spatio-temporal Non-Gaussian Radar Returns, Ph.D. Thesis, University of Cambridge, 1998.
[39] H. Sameti, H. Sheikhzadeh, D. Li, R.L. Brennan, HMM-based strategies for enhancement of speech signals embedded in nonstationary noise, IEEE Trans. Speech Audio Process. 6 (5) (1998) 445–455.

[40] J.E. Greenberg, Modified LMS algorithms for speech processing with an adaptive noise canceller, IEEE Trans. Speech Audio Process. 6 (4) (1998) 338–351.
[41] J. Taghia, J. Taghia, N. Mohammadiha, S. Jinqiu, V. Bouse, R. Martin, An evaluation of noise power spectral density estimation algorithms in adverse acoustic environments, in: IEEE International Conference on Acoustics, Speech and Signal Processing, 2011, pp. 4640–4643.
[42] C.M. Anderson, E.H. Satorius, J.R. Zeidler, Adaptive enhancement of finite bandwidth signals in white Gaussian noise, IEEE Trans. Acoust. Speech Signal Process. 31 (1) (1983) 17–28.
[43] M. Latos, M. Pawelcyk, Adaptive algorithms for enhancement of speech subject to a high-level noise, Arch. Acoust. 35 (2) (2010) 202–212.
[44] N.D. Gaubitch, D.B. Ward, P.A. Naylor, Statistical analysis of the autoregressive modeling of reverberant speech, J. Acoust. Soc. Am. 120 (6) (2006) 4031–4039.
[45] ITU-T, Test Signals for Use in Telephonometry, International Telecommunication Union, Dec 1, 2009.
[46] T. Lotter, P. Vary, Speech enhancement by MAP spectral amplitude estimation using super-Gaussian speech model, EURASIP J. Appl. Signal Process. 7 (2005) 1110–1126.
[47] J. Jensen, R. Heusdens, Improved subspace-based single-channel speech enhancement using generalized super-Gaussian priors, IEEE Trans. Audio Speech Lang. Process. 15 (3) (2007) 862–872.
[48] J. Sohn, N.S. Kim, W. Sung, A statistical model-based voice activity detection, IEEE Signal Process Lett. 6 (1) (1999) 1–3.
[49] D. Shannon, Box-and-whisker plots with the SAS, Pharm. Stat. 2 (2003) 291–295.
[50] H. Levitt, Noise reduction in hearing aids: a review, J. Rehabil. Res. Dev. 38 (1) (2001) 111–121.
[51] J. Benesty, S. Makino, J. Chen, Speech Enhancement, Springer, 2005.
[52] V. Hamacher, J. Chalupper, J. Eggers, E. Fischer, U. Kornagel, H. Puder, U. Rass, Signal processing in high-end hearing aids: state of the art, challenges, and future trends, EURASIP J. Appl. Signal Process. 18 (2005) 2915–2929.
[53] M.H. Costa, An improved adaptive predictor for uncorrelated noise reduction in hearing aids, in: European Signal Processing Conference, 2011, pp. 476-480.

Márcio Holsbach Costa received the B.E.E. degree from Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Brazil, the M.Sc. degree in biomedical engineering from Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil, and the Doctoral degree in electrical engineering from Universidade Federal de Santa Catarina (UFSC), Florianópolis, Brazil, in 1991, 1994, and 2001, respectively. From 1994 to 2004, he was with the Biomedical Engineering Group at Universidade Católica de Pelotas. Since 2004 he is with the Department of Electrical Engineering at Universidade Federal de Santa Catarina. Currently, he is a visiting researcher at the Speech and Audio Processing Group, Imperial College—London. His present research interests focus on biomedical signal processing, linear and nonlinear adaptive filters, hearing aids and active noise and vibration control.