AN IMPROVED ADAPTIVE PREDICTOR FOR UNCORRELATED NOISE REDUCTION IN HEARING AIDS

Márcio Holsbach Costa

Departamento de Engenharia Elétrica, Universidade Federal de Santa Catarina
Centro Tecnológico, Cidade Universitária, 88040-900, Florianópolis-SC, Brazil
phone: + (55) 48 3721-9506, fax: + (55) 48 3721-9280, email: costa@eel.ufsc.br

ABSTRACT

This work presents a new low-cost technique to reduce uncorrelated acoustical noise in digital hearing aids. The clean speech is estimated by a convex combination of the conventional adaptive linear predictor output and the contaminated input signal. The convex combination coefficient is adjusted to control the attenuation of the uncorrelated noise while avoiding significant unvoiced utterance distortion. The proposed adaptive architecture is presented and designed to minimize the mean square prediction error. Comparisons with the conventional adaptive predictor indicate a superior performance of the new solution. The proposed technique can be easily implemented in existing hearing aids systems that incorporate the adaptive predictor, with a minimum amount of extra computational resources.

1. INTRODUCTION

Hearing aids are essential devices for social integration of people that suffer from hearing limitations or neurosensorial losses. One of the most important complaints of hearing aids users is poor speech intelligibility due to background noise. Although there have been many advantages in multi-microphone techniques, most commercial equipments are still equipped with only one microphone, which limits the applicable strategies [1]. The most common noise reduction techniques are: subspace decomposition methods, statistic or parametric modelling techniques, Wiener filtering, and spectral subtraction techniques [2]. Without exception, all techniques present a trade-off between noise reduction and speech distortion [2]-[4].

Digital hearing aids are very complex systems compounded of a set of subsystems performing tasks such as noise reduction, dynamic compression, feedback cancelling, and sound classification [5]. These elements interact with each other to improve intelligibility and propitiate better acoustical comfort to the user.

The limited computational resources of the available commercial devices strongly limit the development of new techniques for hearing aids improvement. Lately, manufacturers have provided dedicated hardware to overcome this problem. One example of a very successful commercial architecture for implementation of a noise reduction system can be found in [6]-[7]. In this architecture, the digitized acoustical signal is split in different frequency channels. Then, each channel is subjected to independent attenuation factors before signal reconstruction. Signal components pertaining to channels with higher signal to noise ratio (SNR) are enhanced, while those belonging to low SNR channels are attenuated, resulting in a better global SNR. However, this approach presents a good performance only for narrowband noise. In the case of broadband noise, all channels tend to suffer approximately the same attenuation, maintaining the same global SNR. As a result, complementary broadband noise reduction techniques are necessary in order to effectively obtain a high quality speech signal. The most important type of broadband noise in hearing aids is the uncorrelated noise, and many attempts to overcome this problem can be found in the main scientific databases. An interesting low computational cost alternative for uncorrelated noise reduction is the linear adaptive predictor [8]-[9]. Despite the advantages of the adaptive predictor as an uncorrelated noise reduction technique, it presents the significant disadvantage of cancelling the speech uncorrelated components, which constitute about 20 to 25% of the natural speech in the English language [10]. Some works have recently addressed this problem, attempting to obtain practical and low distortion noise cancellers. In [11], a technique that weights the sum of the contaminated signal with the adaptive predictor output was proposed. This approach aims at enhancing the quasi-stationary components of the speech (voiced sounds), improving intelligibility and, secondarily, the SNR. However, many intelligibility problems can be referred to losses of comprehension of unvoiced sounds. As a result, this strategy has severe limitations for noise reduction applications.

In [12], the output and error signals of the adaptive predictor are linearly combined using attenuation factors directly related to the instantaneous SNR. This approach does not provide good results when unvoiced speech and uncorrelated noise occur simultaneously. Nowadays, low cost uncorrelated noise reduction techniques are of great commercial interest, and constitute an open scientific and technical subject.

This work presents a new uncorrelated noise reduction technique that estimates the clean speech by using a convex combination of the contaminated original signal and the output of the conventional linear adaptive predictor. The convex combination weight factor establishes a trade-off between uncorrelated noise reduction and unvoiced speech distortion. The proposed algorithm outperforms the conventional adaptive predictor performance, increasing the speech quality.

 - ISSN 2076-1465

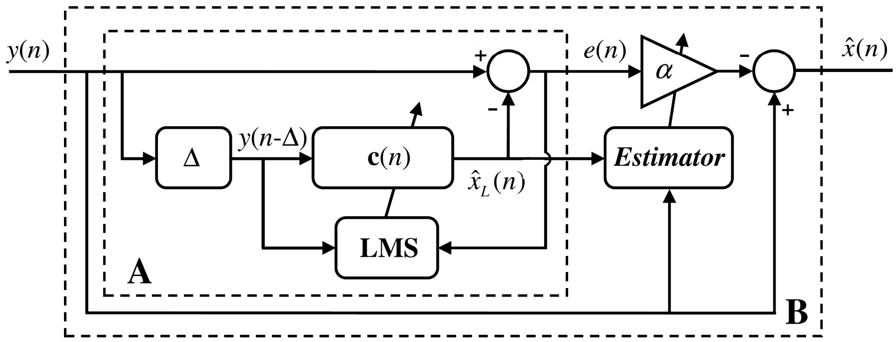

Figure 1 – Proposed architecture.

2. NOMENCLATURE AND NOISE CONTAMINATION MODEL

Along this text, bold uppercase and lowercase letters represent matrices and vectors, respectively, while italic letters represent scalars.

The sampled acoustic signal at time instant n is defined as the sum of the speech signal $x(n)$ and the noise $\eta(n)$, resulting in $y(n) = x(n)+\eta(n)$. The noise $\eta(n)$ is assumed stationary, independent of $x(n)$, zero-mean, with power σ_η^2, and uncorrelated ($E\{\eta(n)\eta(n-k)\} = 0$ for $k \geq 1$). The speech signal $x(n)$ has zero mean with power σ_x^2 and is modelled by an autoregressive process with small correlation length for unvoiced utterances or large correlation length for voiced utterances. The model coefficients are constant in a given time window.

3. LINEAR ADAPTIVE PREDICTOR

The conventional linear adaptive predictor is shown in Fig. 1, internally to the dashed square denoted by letter A. The N-coefficient least mean square (LMS) vector update equation is given by

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

where $y(n)$ is the input signal, Δ is a delay of sample units, $\mathbf{y}(n\text{-}\Delta) = [\ y(n\text{-}\Delta)\ \ y(n\text{-}\Delta\text{-}1)\ \ \ldots\ \ y(n\text{-}\Delta\text{-}N\text{+}1)\ ]^T$, $\mathbf{c}(n) = [\ c_0(n)\ c_1(n)\ \ldots\ c_{N\text{-}1}(n)\ ]^T$, $\hat{x}_L(n)$ is the predictor's output, $e(n) = y(n)\text{-}\hat{x}_L(n) = y(n)\text{ - }\mathbf{c}^T(n)\mathbf{y}(n\text{-}\Delta)$ is the prediction error and μ is the convergence step. It is well known that, assuming convergence of the coefficients, the LMS leads to an unbiased steady-state solution given by

$$\lim_{n\to\infty} E\{\mathbf{c}(n)\} = \mathbf{R}_{\mathbf{yy}}^{-1}\mathbf{r}_{x\mathbf{x}_\Delta} = \mathbf{c_o}, \tag{2}$$

where, $\mathbf{R}_{\mathbf{yy}} = E\{\mathbf{y}(n)\mathbf{y}^T(n)\}$ is the $(N{\times}N)$ input signal correlation matrix and $\mathbf{r}_{x\mathbf{x}_\Delta} = E\{x(n)\mathbf{x}(n\text{-}\Delta)\}$. Using (2) and $\mathbf{R}_{\eta\eta} = \sigma_\eta^2\mathbf{I}$, after some mathematical manipulation, it is possible to come to [2]

$$\mathbf{c_o} = \left(\frac{\mathbf{I}}{SNR} + \tilde{\mathbf{R}}_{\mathbf{xx}}\right)^{-1}\tilde{\mathbf{r}}_{x\mathbf{x}_\Delta}, \tag{3}$$

where

$$\tilde{\mathbf{R}}_{\mathbf{xx}} = \frac{1}{\sigma_x^2}\mathbf{R}_{\mathbf{xx}} \quad , \quad \tilde{\mathbf{r}}_{x\mathbf{x}_\Delta} = \frac{1}{\sigma_x^2}\mathbf{r}_{x\mathbf{x}_\Delta}, \tag{4}$$

$\text{SNR} = \sigma_x^2/\sigma_\eta^2$ is the signal to noise ratio, and $\mathbf{R}_{\mathbf{xx}}$ is the $(N{\times}N)$ clean speech correlation matrix. For extreme values of SNR we obtain:

$$\begin{cases} SNR \to 0 & \Rightarrow \quad \mathbf{c_o} \to \mathbf{0} \\ SNR \to \infty & \Rightarrow \quad \mathbf{c_o} \to \tilde{\mathbf{R}}_{\mathbf{xx}}^{-1}\tilde{\mathbf{r}}_{x\mathbf{x}_\Delta} \end{cases}. \tag{5}$$

Equation (5) permits to verify the SNR effect over the conventional predictor performance when it is used as an uncorrelated noise reduction technique. Two interesting characteristics can be directly observed: (a) in the absence of speech, the predictor output tends to zero, increasing the acoustical comfort of the hearing aids user, and (b) in case of a large SNR there is signal distortion (because of the delay Δ), since in this case $\mathbf{c_o} \to [\ 1\ 0\ \ldots\ 0\ ]^T$ would be the desired solution. As a result, the conventional adaptive predictor is expected to degrade the quality of speech signals in high SNR conditions. A similar problem, but with big consequences, occurs for intermediate SNR conditions and unvoiced speech (situation characteristic of usual conversation). In such a situation $\mathbf{r}_{x\mathbf{x}_\Delta} \cong \mathbf{0}$ and thus $\mathbf{c_o} \to \mathbf{0} = [\ 0\ 0\ \ldots\ 0\ ]^T$. As a result, the adaptive predictor significantly attenuates unvoiced utterances, decreasing speech intelligibility and naturalness.

3.1 Minimum Mean Square Prediction Error

The mean square prediction error (MSPE) of the conventional adaptive predictor is given by

$$J_L(\mathbf{c}) = E\left\{\left[x(n) - \hat{x}_L(n)\right]^2\right\}. \tag{6}$$

and the minimum MSPE is given by

$$J_L(\mathbf{c_o}) = \sigma_x^2 - \mathbf{c_o}^T\mathbf{R}_{\mathbf{yy}}\mathbf{c_o}. \tag{7}$$

Equation (7) quantifies the performance of the adaptive predictor in estimating the clean speech. Later it will be compared with the steady-state MSPE of the proposed algorithm.

4. PROPOSED SYSTEM

The proposed system seeks to minimize the uncorrelated noise while avoiding significant attenuation of unvoiced utterances. Its architecture is shown in Fig 1 internally to the dashed block labelled B. Note that the conventional predictor is part of the whole structure. The output of the proposed system is given by

$$\hat{x}(n) = y(n) - \alpha e(n) = (1-\alpha)\, y(n) + \alpha \hat{x}_L(n). \qquad (8)$$

The convex combination parameter α permits to balance the contributions of the contaminated input signal and the output of the conventional adaptive predictor. It ranges from 0 to 1 and should be dynamically designed to minimize distortions of the output signal in case of occurrence of unvoiced utterances or large SNR.

4.1 Performance Surface

The mean square error between the clean speech and the output signal of the proposed system is a function of the predictor's coefficients and of the parameter α. It is given by

$$J(\mathbf{c},\alpha) = E\left\{\left[x(n) - \hat{x}(n)\right]^2\right\}. \qquad (9)$$

Using (8) in (9) we obtain

$$\begin{aligned} J(\mathbf{c},\alpha) &= \alpha^2 E\left\{y^2(n)\right\} - 2\alpha E\left\{y(n)\eta(n)\right\} \\ &+ E\left\{\eta^2(n)\right\} - 2\alpha^2 E\left\{y(n)\mathbf{y}^T(n-\Delta)\right\}\mathbf{c} \\ &+ 2\alpha E\left\{\eta(n)\mathbf{y}^T(n-\Delta)\right\}\mathbf{c} \\ &+ \alpha^2 \mathbf{c}^T E\left\{\mathbf{y}(n-\Delta)\mathbf{y}^T(n-\Delta)\right\}\mathbf{c} \end{aligned}. \qquad (10)$$

Knowing that $E\{\mathbf{y}(n\text{-}\Delta)\mathbf{y}^T(n\text{-}\Delta)\} = \mathbf{R_{yy}}$, $E\{y^2(n))\} = \sigma_y^2$, $E\{\eta(n)\mathbf{y}(n\text{-}\Delta)\} = E\{\eta(n)\boldsymbol{\eta}(n\text{-}\Delta)\} = \mathbf{r}_{\eta\boldsymbol{\eta}_\Delta}$, $E\{y(n)\eta(n)\} = \sigma_\eta^2$, and $E\{y(n)\mathbf{y}(n\text{-}\Delta)\} = \mathbf{r}_{x\mathbf{x}_\Delta} + \mathbf{r}_{\eta\boldsymbol{\eta}_\Delta}$ then (10) turns to

$$\begin{aligned} J_x(\mathbf{c},\alpha) &= \alpha^2\sigma_y^2 + (1-2\alpha)\sigma_\eta^2 - 2\alpha^2\mathbf{r}_{x\mathbf{x}_\Delta}^T\mathbf{c} \\ &+ \alpha^2\mathbf{c}^T\mathbf{R_{yy}}\mathbf{c} - 2\alpha(\alpha-1)\mathbf{r}_{\eta\boldsymbol{\eta}_\Delta}^T\mathbf{c} \end{aligned}. \qquad (11)$$

Assuming that $\Delta > 0$, the minimum of the performance surface (11) ($\mathbf{c} = \mathbf{c_o}$ and $\mathbf{r}_{x\mathbf{x}_\Delta} = \mathbf{R_{yy}c_o}$) is given by

$$J(\mathbf{c_o},\alpha) = \left(\sigma_y^2 - \mathbf{c_o}^T\mathbf{R_{yy}}\mathbf{c_o}\right)\alpha^2 - 2\sigma_\eta^2\alpha + \sigma_\eta^2. \qquad (12)$$

4.2 Optimal Setting

Equation (12) is a quadratic function, whose point of minimum can be obtained differentiating it with respect to α and equating to zero, in a way that

$$\frac{\partial}{\partial\alpha} J(\mathbf{c_o},\alpha) = 2\left(\sigma_y^2 - \mathbf{c_o}^T\mathbf{R_{yy}}\mathbf{c_o}\right)\alpha - 2\sigma_\eta^2 = 0. \qquad (13)$$

From (13) the optimum solution is

$$\alpha_o = \frac{\sigma_\eta^2}{\sigma_y^2 - \mathbf{c_o}^T\mathbf{R_{yy}}\mathbf{c_o}}, \qquad (14)$$

where σ_η^2 is the noise power, σ_y^2 is the input signal power, and $\mathbf{c_o}^T\mathbf{R_{yy}c_o}$ is the optimal predictor output power.

4.3 Minimum Mean Square Prediction Error

Using (14) in (12) we obtain the minimum MSPE of the proposed technique, given by

$$J(\mathbf{c_o},\alpha_o) = \frac{\left(\sigma_x^2 - \mathbf{c_o}^T\mathbf{R_{yy}}\mathbf{c_o}\right)\sigma_\eta^2}{\left(\sigma_x^2 - \mathbf{c_o}^T\mathbf{R_{yy}}\mathbf{c_o}\right) + \sigma_\eta^2}. \qquad (15)$$

A ratio between both MSPEs, predictor and proposed system, can be obtained dividing (7) by (15)

$$G_J = \frac{J_L(\mathbf{c_o})}{J(\mathbf{c_o},\alpha_o)} = 1 + \frac{J_L(\mathbf{c_o})}{\sigma_\eta^2}. \qquad (16)$$

Since all terms in equation (16) are positive, then $G_J \geq 1$ and, therefore, it is possible to conclude that the proposed solution leads to an output signal that better represents the speech signal, in the mean square sense. This occurs due to a better trade-off between noise reduction and distortion of the uncorrelated part of the speech. The minimum $J_L(\mathbf{c_o})$ encompasses both speech distortion and residual noise.

5. PRACTICAL IMPLEMENTATION ISSUES

In order to obtain a real-time dynamic approximation to the optimum parameter α_o, in real nonstationary conditions, the following estimator of (14) can be used

$$\hat{\alpha}_o(n) = \frac{\sigma_\eta^2(n)}{\sigma_y^2(n) - \sigma_{\hat{x}_L}^2(n)}, \qquad (17)$$

where $\sigma_\eta^2(n)$ is an estimate of the instantaneous additive noise power, which can be obtained 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 predictor output power. These estimates can be obtained using three recursive first order low-pass filters, given by

$$\sigma_w^2(n) = \tau_w\sigma_w^2(n-1) + (1-\tau_w)\, w^2(n) \qquad (18)$$

where $w(n)$ and $\sigma_z^2(n)$ represent variables $\eta(n)$, $y(n)$, $\hat{x}_L(n)$ and their respective powers for each filter.

6. SIMULATION RESULTS

This section presents simulation results to illustrate the performance of the proposed algorithm. Due to space limitations, only two examples are presented. The first compares the minimum MSPE of the conventional adaptive linear predictor with that of the proposed algorithm under completely known conditions. This is done to show the effects of both the SNR and the value of parameter α in their performance. The input signal was a simulated unvoiced utterance modelled by a 22 order autoregressive model [13], obtained from a 20 milliseconds male speaker epoch of the /s/ phoneme. The sampling frequency was 15.625 kHz, $N = 10$, $\mu = 10^{-6}$, $\Delta = 1$, additive Gaussian white noise, $\sigma_x^2 = 6$, and SNR = 0.4, 3, 10.4, and 20 dB. Fig. 2 shows the results obtained for: $E\{[x(n)\text{-}y(n)]^2\} = \sigma_\eta^2$ (dotted), $E\{[x(n)\text{-}\hat{x}_L(n)]^2\}$ (dashed), and $E\{[x(n)\text{-}\hat{x}(n)]^2\}$ (continuous line). It can be verified that for high SNRs the conventional adaptive predictor provides a higher MSPE when compared to the original contaminated signal, this condition is reversed for low SNRs. This occurs because the conventional predictor attenuates not only the uncorrelated contamination noise but the unvoiced components of the speech as well. The optimum parameter α_o, evaluated by equation (14), is shown in Fig. 2 as an asterisk and clearly coincides with the minimum

MSPE obtainable with the proposed algorithm in all SNR conditions. The parameter α_o is seen to lead always to the minimum MSPE. Additionally, the value of α_o increases as the SNR decreases, *e.g.*, for SNR > 20dB we obtain $\alpha_o \cong 0$, suggesting the best possible speech estimate is basically the unprocessed (contaminated) speech. For SNR < 0 dB we have $\alpha_o \cong 1$, indicating that the best achievable estimate is the conventional adaptive predictor output. For the 3 dB < SNR < 10 dB range, there exists a wide range of α values around α_o that results in smaller MSPEs than the conventional predictor and the (original) contaminated input signal. This result demonstrates the robustness of the proposed algorithm to α_o estimation errors.

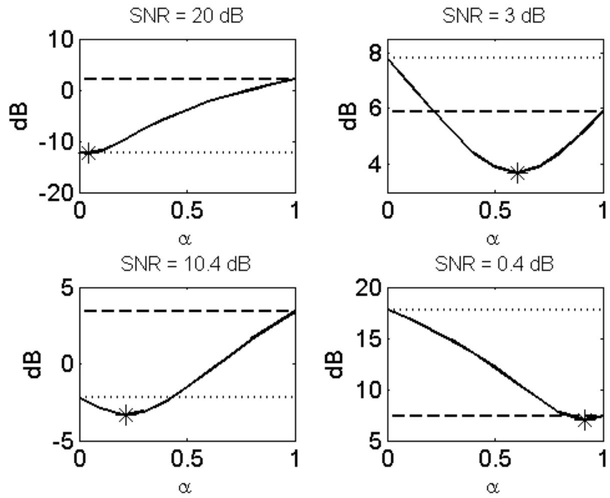

Figure 2 – Simulated unvoiced speech MSPEs obtained from: the contaminated signal (dotted), the conventional predictor (dashed), and the proposed algorithm (continuous line). The asterisk represents the MSPE of the proposed algorithm obtained with α_o.

The second example makes use of a real speech input signal from a male speaker. The artificial noise is white Gaussian with different SNRs (-3, 0, 3, 8, 10, 12, 15, and 20 dB). The parameters used were $\Delta = 1$, $\mu = 0.01$, $N = 10$. Figs. 3 to 6 show the results obtained for the MSPE and three objective speech quality measures: weighted-spectral slope metric (WSS), Itakura-Saito measure (ISS), and perceptual evaluation of speech quality (PESQ) [4].

The MSPE presented in Fig. 3 indicates the proposed algorithm always results in better mean square estimates when compared to the conventional adaptive predictor. For SNR < -10 dB then $\alpha \to 1$ and both algorithms present approximately the same behaviour. For SNR > 20dB there are basically no differences between the MSPE of the contaminated input signal and the output signal of the proposed algorithm, suggesting no speech distortions. In Fig. 4, the WSS criterion suggests the performance of the new algorithm results in estimates with approximately the same quality of the contaminated signal and both are better than the conventional predictor output. The WSS criterion penalizes large distances at the locations of the spectral peaks, minimizing tilt and overall level differences. As a result, this criterion is not able to demonstrate quality gains for the proposed algorithm since additive noise does not change spectral peaks locations. However, WSS indicates that the new strategy reduces the effects of coefficient fluctuations that occur due to a small but nonzero convergence step, resulting in lower WSS indexes for all tested SNRs.

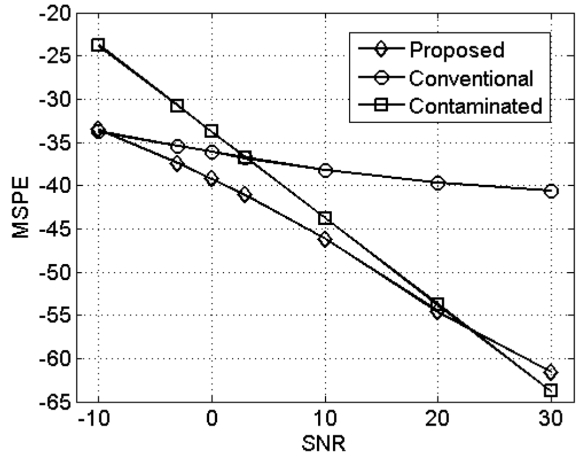

Figure 3 – MSPE for real speech and white noise.

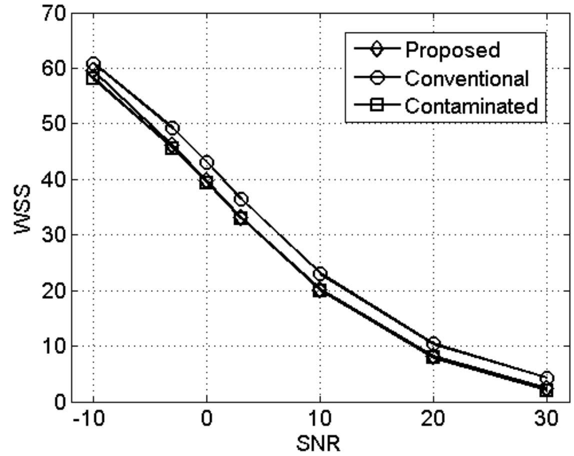

Figure 4 – WSS for real speech and white noise.

Fig. 5 shows that the IS results resemble the MSPE ones in Fig. 3. The IS indexes suggest the new algorithm, again, results in better estimates of the clean speech. The IS penalizes differences at the spectral global levels between the desired and target signals, indicating significant speech distortion associated with the conventional predictor's output.

In Fig. 6, PESQ results can be visualized. This criterion measures distortions commonly found in telecommunication systems and presents a high correlation with the Mean Opinion Score test. The obtained results indicate that the conventional predictor does not result in significant subjective quality improvement when compared to the contaminated speech, and signal degradation occurs for SNR > 3 dB. The proposed algorithm presents a significant quality improvement in the -10dB < SNR < 10 dB range. A preliminary analysis indicates that variations of the conventional adaptive linear predictor such as [8]-[9], [11], [12] could benefit from using the proposed strategy. Assuming the availability of a VAD, the proposed algorithm requires only 5 extra sums, 10 multiplications and 1 division (corresponding to 16 multiplications [14]) compared to the conventional predictor.

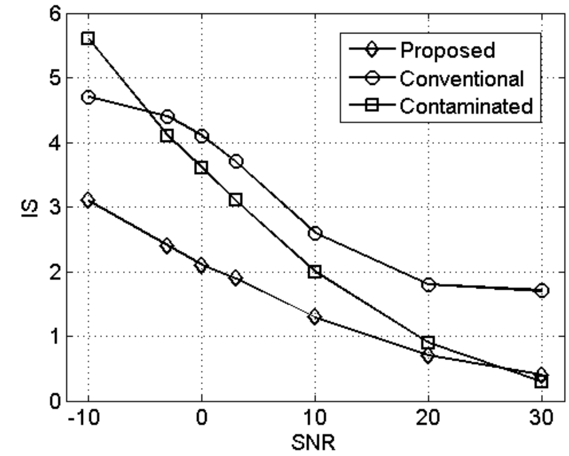

Figure 5 – IS for real speech and white noise.

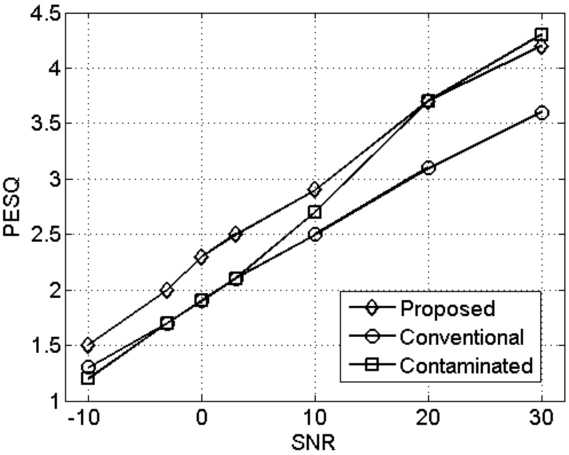

Figure 6 – PESQ for real speech and white noise.

7. CONCLUSIONS

This work presented the derivation of a new low-cost solution for uncorrelated noise reduction in hearing aids. The optimum setting for maximum performance was theoretically obtained, resulting in a smaller mean square prediction error as compared with the conventional adaptive linear predictor. Simulations using artificial and real speech signals corroborate the theoretical results. Three different objective quality measures indicate speech quality improvement when compared with the conventional adaptive predictor. Uncorrelated noise reduction systems based on the conventional adaptive linear predictor and its variations can be easily modified to incorporate the new proposal with a minimum amount of extra computational resources.

ACKNOWLEDGMENTS

Thanks to Alexandre Ferreira and Acústica Amplivox Company and for bringing my attention to this subject and to Prof. José Carlos Bermudez for his insightful comments. This work was supported by the Brazilian Ministry of Science and Technology (CNPq) under grants 559418/2008-6, and 303803/2009-6.

REFERENCES

[1] J. Benesty, J. Chen, and Y.A. Huang, "Noise reduction algorithms in a generalized transform domain," *IEEE Transactions on Audio, Speech, and Language Processing* vol. 17, no. 6, pp. 1109-1123, 2009.

[2] J. Chen, J. Benesty, Y. Huang, and S. Doclo, "New insights into the noise reduction Wiener filter," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 4, pp. 1218-1234, 2006.

[3] R. Bentler, and L.-K. Chiou, "Digital noise reduction: an overview," *Trends in Amplification*, vol. 10, no. 2, pp. 67-82, 2006.

[4] P.C. Loizou, *Speech enhancement: theory and practice*, CRC, 2007.

[5] K. Chung, "Challenges and recent developments in hearing aids – part I: speech understanding in noise, microphone technologies and noise reduction algorithms," *Trends in Amplification*, vol. 8, no. 3, pp. 83-124, 2004.

[6] X. Fang, and M. J. Nilsson, *Noise Reduction Apparatus and Method*, US Patent 6.757.395, 2004.

[7] Voyager, *Time domain filter bank, user's manual*, Information note, Gennum, 2005, pp. 1-24.

[8] R. Sambur, "Adaptive noise canceling for speech signals," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-26, no. 5, pp. 419-423, 1978.

[9] A. Kawamura, K. Fujii, Y. Itoh, and Y. Fukui, "A new noise reduction method using linear prediction error filter and adaptive digital filter," in *Proc*. ISCAS 2002, Scottsdale, USA, May 26-29. 2002, pp. 488-491.

[10] G. Hu, and D. Wang, "Segregation of unvoiced speech from nonspeech interference," *Journal of Acoustical Society of America*, vol. 124, no. 2, pp. 1306-1319, 2008.

[11] G.P. Eatwell, *Noise reduction filter*, US Patent 5,742,694, 1998.

[12] D. Giesbrecht, and P. Hetherington, *Periodic signal enhancement system*, US Patent 7,680,652 B2, 2010.

[13] N.D. Gaubitch, D.B. Ward, and P.A. Naylor, "Statistical analysis of the autoregressive modeling of reverberant speech," *Journal of Acoustical Society of America*, vol. 120, no. 6, pp. 4031-4039, 2006.

[14] Y. Lu, R. Fowler, W. Tian, L. Thompson, "Enhancing echo cancellation via estimation of delay", *IEEE Transactions on Signal Processing*, vol. 53, no. 11, pp. 4159-4168, 2005.