A NEW NON-LINEAR PROCESSOR (NLP) FOR BACKGROUND CONTINUITY IN ECHO CONTROL

Oğuz Tanrıkulu

Tellabs Operations
1 Kendall Sq., Bldng 100
Cambridge, MA 02139
USA
Oguz.Tanrikulu@tellabs.com

Kutluyıl Doğançay

School of Electrical and Information Engineering
University of South Australia
Mawson Lakes, SA 5095
Australia
Kutluyil.Dogancay@unisa.edu.au

ABSTRACT

A new method of Nonlinear Processor design is presented for echo control based on the use of spectral modeling and limiting non-linear functions. The objective of the proposed design is to further suppress the residual echo left after the adaptive canceller while preserving the continuity of the coherent background signals such as music.

1. INTRODUCTION

Echo cancellation through adaptive filtering has been studied and used in the real world for over three decades. Two primary application areas are electric echo cancellation and acoustic echo cancellation [1]. While the objective in both areas is essentially the same, the implementation and performance considerations can be vastly different. The echo-path in electric echo cancellation is generally short - generally 9ms to 25ms [2] with some arbitrary bulk-delay preceding the actual impulse response [3]- and usually stationary. The echo-path in acoustic echo cancellation, on the other hand, can easily reach to a couple of hundred milliseconds and the late reflections are time-varying.

In both applications of echo cancellation, much of emphasis in research has been in the design of adaptive algorithms. However, both from theoretical considerations and practical limitations on the computational complexity of the implementation, it is widely known that an adaptive filter almost always falls short of canceling the echo completely. As the last step of processing, a center clipper with dynamic thresholds is used to remove the residual, which is often combined with comfort noise injection [2]. This arrangement, known as the Non-Linear Processor (NLP), has been used for electric echo cancellation for a long time and its performance is adequate for wireline networks where the background noise is usually very low. However, in more complicated call scenarios, e.g. mobile-to-PSTN or mobile-to-mobile (see Fig. 1), the noise in the background can be a high level coherent signal and the removal of the residual echo though a conventional NLP will also remove the background signal. For the listener, the background signal will be modulated creating an annoying artifact.

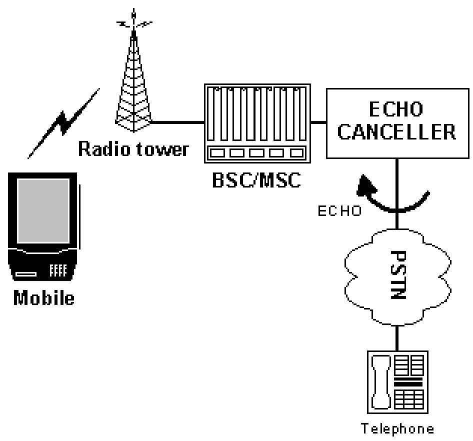

Fig. 1. PSTN-Gateway application for electric echo cancellation.

The objective of this paper is to address the following question: *Is it possible to remove or further attenuate the residual echo while retaining the continuity and the content of the true background signal?* A number of computationally intensive approaches to this problem are not considered due to high density requirements on electric echo cancellers. Another constraint for the design is that the processing delay of echo cancellation should be less than 1msec [2].

The basic structure for echo cancellation is illustrated in Fig. 2, where $x(k)$ is the far-end signal, which is the input of the adaptive filter and $y(k)$ is the output of the adaptive filter. The impulse response of the echo-path is denoted by $\mathbf{w}_*$. The taps of the linear transversal adaptive filter are denoted by $\mathbf{w}(k)$. The near-end signal is composed of a speech signal component, $s(k)$, and an additive noise com-

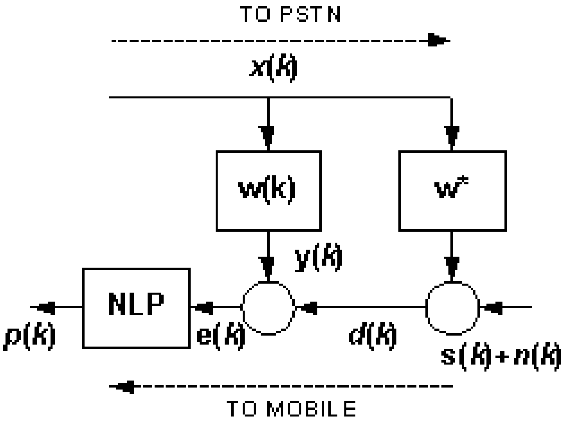

Fig. 2. Echo-path, adaptive filter and the non-linear processor.

ponent, $n(k)$. Based on Fig. 2, we have

$$y(k) = \mathbf{w}^T(k)\mathbf{x}(k) \tag{1}$$

$$d(k) = s(k) + n(k) + \mathbf{w}_*^T(k)\mathbf{x}(k) \tag{2}$$

and therefore

$$e(k) = s(k) + n(k) + [\mathbf{w}_*(k) - \mathbf{w}(k)]^T\mathbf{x}(k) \tag{3}$$

A conventional NLP with comfort noise injection operates as follows:

• In Fig. 2, when both the far-end and the near-end users are active, the echo canceller is in double-talk state. In this case, the adaptation of $\mathbf{w}(k)$ is stopped and the NLP is taken out of the way of $e(k)$ so that $s(k)$ can reach the mobile user without clipping.

• On the other hand, if the far-end user is active alone ($s(k) = 0$), i.e., single-talk state, then (3) becomes

$$e(k) = n(k) + [\mathbf{w}_*(k) - \mathbf{w}(k)]^T\mathbf{x}(k) \tag{4}$$

and the NLP operates on $e(k)$ to further suppress any residual echo.

• When there is no far-end speech, $x(k) = 0$, $e(k) = n(k)$ and the NLP is released. Then we have $p(k) = n(k)$.

When in single-talk state, the NLP will operate on (3) such that

$$p(k) = f(e(k)) + c_n(k) \tag{5}$$

where

$$E\{c_n^2(k)\} = E\{n^2(k)\} \tag{6}$$

and $c_n(k)$ is the comfort noise injected. In other words, the power of the output $p(k)$ will match the power of the background noise, but clearly the injected noise, perceptually, will not be identical to the true background. Techniques such as Spectrally Matched Noise Injection (SMNI) can be used, but the output of an autoregressive filter excited with white noise again does not suffice in terms of matching the true background perceptually. In the next section we introduce a new NLP, which does not use a center clipper with comfort noise injection nor does it use SMNI with white noise input.

2. THE PROPOSED NLP STRUCTURE

The SMNI does not perceptually match a complicated background signal, since the filter that captures the short-term spectral information is excited by an artificially generated stationary noise. The importance of coding the excitation signal is obvious in applications such as Speech Coding/Synthesis. Even if the correct spectral envelope is used, if the correct excitation signal is not applied, the synthesized signal will not sound as required. Therefore, what is really needed in NLP design is to use a processed form of the residual echo—with the background noise signal— and use this signal to synthesize the background [4]. The basic structure for this processing is shown in Fig. 3.

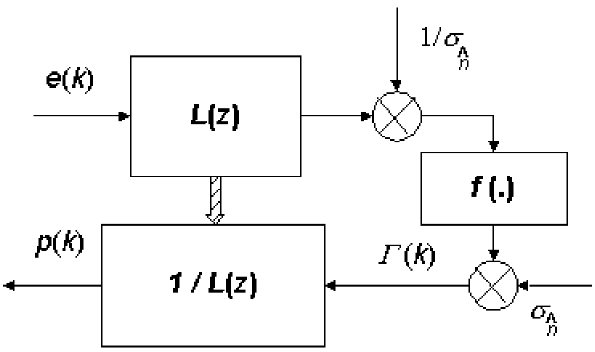

Fig. 3. The proposed NLP structure. The output of $f(.)$ can be mixed with stationary noise.

The operation of the proposed NLP is as follows. During the period where there is only noise in $e(k)$, the impulse response $l(k)$ of $L(z)$ are estimated with the objective of whitening the output of this filter. In this case, assuming perfect whitening, the input signal to $L(z)$ will be relayed to the output of inverse whitening filter $1/L(z)$ without any spectral change because the memoryless nonlinearity $f(.)$ will not change the correlation of the white noise at its input. When the canceller is in double-talk, the structure in Fig. 3 is taken out of loop. Note that, a fixed-order decorrelator is used in practice and, therefore, the output $\hat{n}(k)$ of $L(z)$ cannot be assumed white. Furthermore, if $\hat{n}(k)$ was completely white, exciting the inverse of the filter $l(k)$ with white noise would match the background closely.

When the canceller is in single-talk state, then the parameters of the analysis filter $L(z)$ are copied over to the synthesis filter $1/L(z)$. Clearly, if the output of analysis filter was directly input into the synthesis filter, the design would be moot in that the residual echo in $e(k)$ will be fully synthesized back in $p(k)$. What is needed is further suppression of the residual echo between the analysis and synthesis filters with the added objective of retaining the perception of the background noise signal. In the proposed NLP structure in Fig. 3, these are sought by applying a memoryless saturating non-linear function at the analysis filter output. From Fig. 3 and (3), the output of the canceller is given by

$$p(k) = \Gamma(k) * \tilde{l}(k) \tag{7}$$

where

$$\Gamma(k) = \sigma_{\hat{n}} \mathrm{sgn}[\theta(k)] \tag{8}$$

$$\theta(k) = \frac{\hat{n}(k) + \hat{x}(k)}{\sigma_{\hat{n}}} \tag{9}$$

$$\hat{n}(k) = n(k) * l(k) \tag{10}$$

$$\hat{x}(k) = \tilde{x}(k) * l(k) \tag{11}$$

$$\tilde{x}(k) = [\mathbf{w}_*(k) - \mathbf{w}(k)]^T \mathbf{x}(k) \tag{12}$$

$$l(k) * \tilde{l}(k) = \delta(k) \tag{13}$$

Note that in (7) other saturating non-linear functions can also be used. The reason for using sgn$(.)$ is that while it limits the signal level and therefore attenuates the residual echo, it will also capture, to some degree, the harmonic properties of its input since it retains the zero crossing information.

Now the autocorrelation properties of $p(k)$ are analyzed. Using the results in [5] for the autocorrelation of the output of a sgn$(.)$ function when it is driven by a stationary Gaussian random process $\{\Theta\}$ with variance σ_θ^2, we get

$$R_\Gamma(k) = \frac{2\sigma_{\hat{n}}^2}{\pi} \sin^{-1}\left(\frac{R_\theta(k)}{\sigma_\theta^2}\right) \tag{14}$$

Assuming that $\hat{n}(k)$ and $\hat{x}(k)$ are statistically independent in (9), we have

$$\sigma_\theta^2 = \frac{\sigma_{\hat{n}}^2 + \sigma_{\hat{x}}^2}{\sigma_{\hat{n}}^2} \tag{15}$$

and (14) can be rewritten as

$$R_\Gamma(k) = \frac{2\sigma_{\hat{n}}^2}{\pi} \sin^{-1}\left(\frac{R_{\hat{n}}(k) + R_{\hat{x}}(k)}{\sigma_{\hat{n}}^2 + \sigma_{\hat{x}}^2}\right) \tag{16}$$

By investigating a number of special cases, the behaviour of the proposed NLP can be better understood.

Case 1: ($k = 0$) By using (15), we get

$$R_\Gamma(0) = \sigma_{\hat{n}}^2 \tag{17}$$

and, therefore, the input of the synthesis filter will be driven by an excitation signal with the appropriate power.

Assumption: To analyze further we make the following assumption which is true when the noise in the background is highly coherent:

$$|R_{\hat{n}}(k)| << |R_{\hat{x}}(k)|, k \text{ small} \tag{18}$$

$$|R_{\hat{n}}(k)| >> |R_{\hat{x}}(k)|, k \text{ large} \tag{19}$$

Under this assumption there are two more cases:

Case 2: (Small values of k) Here ignoring the contribution of the noise in the numerator term in (16) we have

$$R_\Gamma(k) \approx \frac{2\sigma_{\hat{n}}^2}{\pi} \sin^{-1}\left(\frac{R_{\hat{x}}(k)}{\sigma_{\hat{n}}^2 + \sigma_{\hat{x}}^2}\right) \tag{20}$$

Since

$$\left|\sin^{-1}\left(\frac{R_{\hat{x}}(k)}{\sigma_{\hat{n}}^2 + \sigma_{\hat{x}}^2}\right)\right| \leq \frac{\pi}{2} \tag{21}$$

we have

$$|R_\Gamma(k)| \leq \sigma_{\hat{n}}^2 \tag{22}$$

as expected, which means that whatever the power of $\hat{x}(k)$ the power of the synthesis filter is limited by $\sigma_{\hat{n}}^2$, hence showing the further residual echo attenuation achieved by the structure.

Case 3: (Large values of k) If the correlation properties of the speech and the noise are stated as above, we can write

$$R_\Gamma(k) \approx \frac{2\sigma_{\hat{n}}^2}{\pi} \sin^{-1}\left(\frac{R_{\hat{n}}(k)}{\sigma_{\hat{n}}^2 + \sigma_{\hat{x}}^2}\right) \tag{23}$$

and also since k is large we can assume $|R_{\hat{n}}(k)| << (\sigma_{\hat{n}}^2 + \sigma_{\hat{x}}^2)$, in which case we can use the approximation approximate sin$^{-1}(x) \approx x$ and write

$$R_\Gamma(k) \approx \frac{2\sigma_{\hat{n}}^2}{\pi(\sigma_{\hat{n}}^2 + \sigma_{\hat{x}}^2)} R_{\hat{n}}(k) \tag{24}$$

which shows that for large k, the correlation of the output of the non-linearity will be proportional to the correlation properties of its input. This aspect of the NLP helps towards maintaining the continuity of the background.

Finally, note that gain control is generally necessary at the output of the proposed NLP to stay within the guidelines of [2].

3. EXPERIMENTS

The performance of the proposed NLP is investigated by using various recordings made between mobile users and PSTN users. The proposed NLP is efficiently implemented following [4]. As an example, the spectrogram of the input $d(k)$ of the echo canceller is shown Fig. 4. The echo canceller is in single-talk. The background noise signal is music and this complicated background can be seen as large number of weak spatio-temporal lines in Fig. 4.

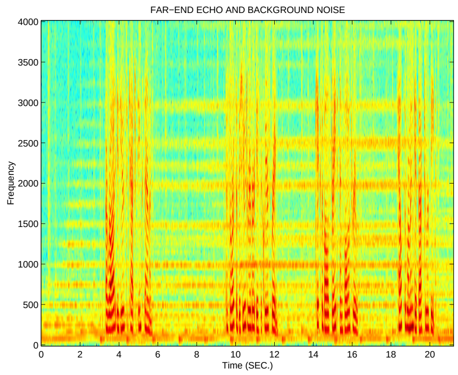

Fig. 4. Spectrogram of the Echo Canceller input $d(k)$.

The objective is to find out how the spectral content of the background changes when we attempt to remove the residual echo signal after the adaptive filter.

The results of a conventional echo cancellation algorithm is shown in Fig. 5. The adaptive part of the echo canceller reduces the echo, but without the NLP the residual echo is still audible. As the second step the NLP of the conventional echo canceller further suppresses the residual echo. The corresponding spectrogram of the echo canceller output is shown in Fig. 5. Here we see that the regions where the echo signal existed are now replaced by comfort noise and the disparity between this signal and the musical background is clear. Most expert listeners would immediately recognize this case as a problem and most naive listeners would describe it as *noise pumping*.

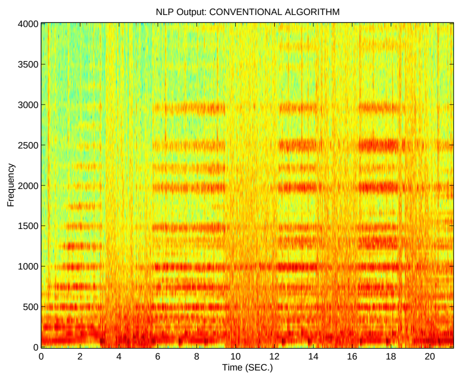

Fig. 5. Spectrogram of NLP output $p(k)$ for conventional algorithm.

The spectrogram of the output of the proposed NLP design in shown in Fig. 6. Note that the same residual echo signal, $e(k)$, is used as in Fig. 5. By examining Fig. 6, we see that in the four regions of echo, the new NLP matches the actual background signal more closely. The tonal nature of the signal is preserved unlike what is seen in Fig. 5. Perceptually, the improvement in the continuity of the background signal is clear. An interesting observation is how the perceptual content of the processed echo changes with the convergence of the adaptive filter. It can be noted that during the first echo burst, where the canceller performs its initial convergence, the output of the NLP has very low level echo. As the echo canceller converges, the performance of the proposed NLP also increases and a high level of background continuity is achieved without any remaining audible echo.

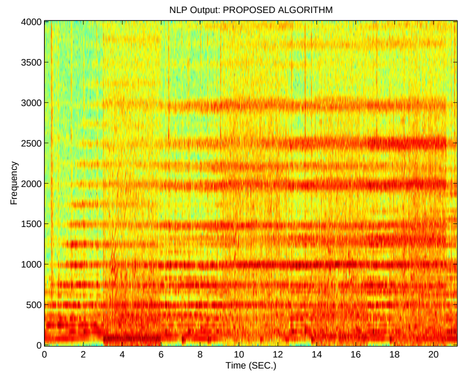

Fig. 6. Spectrogram of NLP output $p(k)$ for the proposed NLP.

4. CONCLUSIONS

A new Non-linear Processor (NLP) is presented where the design criteria was to successfully maintain the continuity of the background signal after echo cancellation. This is a challenging problem since the background and the echo signals are mixed and the background usually has much lower power. The proposed design uses spectral modelling of the background followed by a limiting non-linearity and a synthesis filter. Experimental results with musical background noise show the success of the proposed technique.

5. REFERENCES

[1] J. Benesty, T. Gaensler, D. R. Morgan, M. M. Sondhi and S. L. Gay, *Advances in Network and Acoustic Echo Cancellation*, Springer-Verlag, Berlin, 2001.

[2] ITU-T Recommendation G.168, *Digital Network Echo Cancellers*, April 2000.

[3] P. Ramkumar and O. Tanrıkulu, ”Implementation and evaluation of a real-time line echo bulk-delay estimator,” ICASSP 2000, vol. 6, pp. 3650-3653, Istanbul, Turkey, June 2000.

[4] O. Tanrıkulu and J. O. Dellamorte, *Method and Apparatus for Echo Cancellation Using an Adaptive Lattice Based Non-linear Processor*, US Patent Application, filed on Apr. 11, 2002.

[5] R. F. Baum, “The correlation function of Gaussian noise passed through nonlinear devices,” *IEEE Trans. Inform. Theory*, vol. IT-15, pp. 448–456, July 1969.