OPTIMAL NON-LINEAR PROCESSOR CONTROL FOR RESIDUAL-ECHO SUPPRESSION

Miloš Doroslovački

Department of Electrical and Computer Engineering
The George Washington University
Washington, DC 20052, USA

ABSTRACT

The main goal of this paper is to establish an objective measure of non-linear processor (NLP) control quality for echo suppression after echo cancellation, as a basis for an optimal design of NLP control. As tools for reaching the goal, mathematical modeling and analysis of echo cancellation and suppression are used. Since the optimality assumes the knowledge of parameters of echo cancellation process, effects of uncertain parameters to the objective measure of NLP control quality are studied. Performance of NLP control is investigated under ideal and non-ideal circumstances.

1. NON-LINEAR PROCESSOR CONTROL

Non-linear processors are used for residual echo suppression in telephone networks [1]. They can be seen as devices that block low-power signals and pass high-power signals. The blocking is active or inactive depending on the position of a controllable switch, see Figure 1. When the switch is in the position 0, the blocking is inactive and $e'(n) = e(n)$, where $e(n)$ is the signal obtained after echo cancellation. Note that $e(n)$ is the sum of residual echo and near end signal. When the switch is in the position 1, the blocking is active and $e'(n) = 0$. The position of the switch is controlled by values of the three signals: $P_x(n)$, $P_y(n)$, and $P_e(n)$. These three signals are the average power of far-end signal, average power of hybrid output and average power of signal obtained after echo cancellation, respectively. That is, the NLP control is based on observing the far-end signal $x(n)$, the hybrid output $y(n)$, and the signal obtained after echo cancellation $e(n)$. The controllable switch can be seen as a hard center clipper whose threshold is controlled by $P_x(n)$, $P_y(n)$, and $P_e(n)$. Of course, some other NLPs (e.g. soft center clippers) can be used as well, but they are not considered in this paper.

2. POWER MODEL OF ECHO CANCELLATION IN TELEPHONE NETWORKS

Figure 2 depicts the power model of echo cancellation in telephone networks. The meaning of the quantities in the figure is the following:

$P_x(n)$	–	average power of far-end signal
$P_s(n)$	–	average power of near-end signal
ρ_h	–	hybrid reflection coefficient
$P_y(n)$	–	average power of hybrid output
$P_{\hat{y}}(n)$	–	average power of echo canceller output
$P_e(n)$	–	average power of signal after echo cancellation.

This work was supported by the Telogy Networks, a Texas Instruments Company, under contract AN 19672.

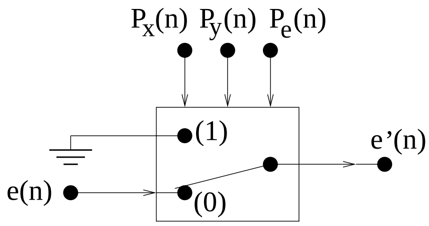

Fig. 1. Non-linear processor.

The instantaneous average power of the hybrid output can be expressed as

$$P_y(n) = \rho_h P_x(n-d) + P_s(n) \quad (1)$$

where d is the pure delay introduced by the hybrid. The equality in (1) is valid in a statistical sense assuming that the far-end signal $x(n)$ and the near-end signal $s(n)$ are zero-mean uncorrelated processes. If they are correlated then

$$P_y(n) = \rho_h P_x(n-d) + P_s(n) + 2\rho\sqrt{P_x(n-d)P_s(n)} \quad (2)$$

where ρ is the correlation coefficient between $x(n-d)$ and $s(n)$. In all further considerations it is assumed that $\rho = 0$.

The (ensemble) average power of signal after echo cancellation is modeled as

$$P_e(n) = P_y(n) - P_{\hat{y}}(n). \quad (3)$$

Taking into consideration that the echo cancellation is an adaptive time-varying process [2], an approximate dynamic model for $P_e(n)$ is

$$P_e(n) = P_s(n) + \rho_0 P_x(n-d) + m(P_y(n) - P_s(n) - \rho_0 P_x(n-d))\mathrm{e}^{-\lambda n} \quad (4)$$

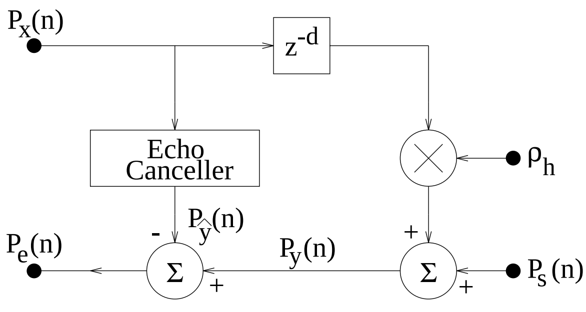

Fig. 2. Power model of echo cancellation.

where the following definitions hold:

ρ_0	–	steady state far-end reflection coefficient after echo cancellation
λ	–	convergence rate parameter
m	–	mode flag ($m = 0$: steady state mode, $m = 1$: transient mode).

The corresponding approximate dynamic model for the echo canceller output is

$$P_{\hat{y}}(n) = (\rho_h - \rho_0)P_x(n-d)(1 - m\mathrm{e}^{-\lambda n}). \tag{5}$$

It is assumed that $\rho_h \geq \rho_0$, i.e. the echo canceller can not make a reflection stronger. Our interest here is to study the NLP control after the adaptive echo canceller has reached the convergence, so we assume $m = 0$ in (4). Also, if not specified differently, $d = 0$ is assumed.

3. OPTIMAL NLP CONTROL

One simple and sound criterion for NLP control can be to maximize the ratio of near-end average power and distorsion/residual echo average power. That is, the goal is to return, after echo cancellation and non-linear processing, a signal whose power is as much as possible close to the power of the near-end signal. When NLP is not used, a measure of echo cancellation gain can be expressed through the ratio of near-end average power and residual echo average power as

$$\eta_{ec} = \frac{\sum\limits_n P_s(n)}{\sum\limits_n (P_e(n) - P_s(n))}. \tag{6}$$

Based on what is mentioned above, the processing gain after NLP can be measured through the ratio of near-end average power and distorsion/residual echo average power as

$$\eta_{nlp} = \frac{\sum\limits_n P_s(n)}{\sum\limits_n |P_{nlp}(n) - P_s(n)|}. \tag{7}$$

Note that in our case, depending on control, $P_{nlp}(n) = 0$ or $P_{nlp}(n) = P_e(n)$. One of these two values should be chosen for each n such that η_{nlp} is maximized. We can also say that the usage of NLP is justified if

$$\eta_{nlp} \geq \eta_{ec}. \tag{8}$$

The maximization of η_{nlp} maximizes also the benefit of using NLP. Based on (6)-(8), the maximization is obtained whenever the control for each n is such that

$$|P_{nlp}(n) - P_s(n)| \leq P_e(n) - P_s(n). \tag{9}$$

The condition reduces to the following control law

$$P_{nlp}(n) = \begin{cases} 0, & P_e(n) > 2P_s(n) \\ P_e(n), & \text{otherwise.} \end{cases} \tag{10}$$

If the steady state far-end reflection coefficient ρ_0 is known, the control law can be written as

$$P_{nlp}(n) = \begin{cases} 0, & 2\rho_0 P_x(n) > P_e(n) \\ P_e(n), & \text{otherwise.} \end{cases} \tag{11}$$

Of course, if we use instead of ρ_0, the estimate $\hat{\rho}_0$, the actual η_{nlp} will be lower than the maximum one. If $\hat{\rho}_0 < \rho_0$, η_{nlp} can be closer to η_{ec}, but not worse than η_{ec}. If $\hat{\rho}_0 > \rho_0$, η_{nlp} can be worse than η_{ec}.

4. OPTIMAL ECHO SUPPRESSION ALGORITHM

The main assumption of the algorithm is that the convergence of adaptive echo canceller has been achieved. The algorithm works by comparing the average power of signal after echo cancellation $P_e(n)$ and the estimated residual echo power $\hat{\rho}_0 P_x(n)$. $P_e(n) \geq 2\hat{\rho}_0 P_x(n)$ defines the inactivity region of blocking. That corresponds to our assessment that the near-end signal component is larger than the residual echo signal component in the signal obtained after echo cancellation. Consequently, $P_e(n) < 2\hat{\rho}_0 P_x(n)$ defines the activity region of blocking that corresponds to our assessment that the near-end signal component is lower than the residual echo signal component. The boundary between the regions (i.e. the decision threshold for activity or inactivity of blocking) is defined as

$$P_e(n) = 2\hat{\rho}_0 P_x(n). \tag{12}$$

The algorithm detects the crossing of a current operating point $(P_e(n), P_x(n))$ through the decision boundary (12). But, the corresponding decision is postponed in order to avoid fast successive switching. This phenomenon can happen because of noise at the far and/or near-end, or because of transitions in average speech power of the far end and/or near end. The delay of the new decision is obtained through so called hangover and operate mechanism [3]. The hangover is activated if a crossing of the decision boundary is caused by a drop of the average power of either far-end signal (when the new decision should be to deactivate blocking) or near-end signal (when the new decision should be to activate blocking). The operate is activated when a crossing of the decision boundary is caused by a jump of the average power of either far-end signal (when the new decision should be to activate blocking) or near-end signal (when the new decision should be to deactivate blocking). The recommended values for hangover and operate delays can be found in [3]. In a few words, the hangover and operate mechanisms allow a transition of NLP state only if conditions for the transition persist.

Figure 3 shows the average output power of NLP when the optimal algorithm for NLP control is used. All hangover and operate delays are set to zero. The steady state far-end reflection coefficient and its estimate is $\rho_0 = \hat{\rho}_0 = -16$ dB. Figure 4 is obtained under the same conditions as Figure 3, but now the hangover delay after a drop of far-end power is 25 ms, while after a drop of near-end power is 40 ms. The operate delay after a jump of far-end power is 45 ms, while after a jump of near-end power is 2.5 ms. The average power of far end in dB is uniformly distributed between -60 and 3 dB, as well as the average power of near end. The periods of constant average power are 250 ms. Average-power changes of near-end signal occur in the middle of periods of constant value for the far-end signal.

As can be seen from Figure 4, the state of NLP changes with delay. State changes do not follow immediately crossings of the decision boundary. That can be a drawback in the case when transitions of average signal power occur abruptly and definitely. When this is not a case, i.e. when we have just short fluctuations in average power, or just impulses we have to ignore, the delays should be used to avoid generation of impulse noise produced by the fast successive switching.

5. PERFORMANCE OF THE OPTIMAL ALGORITHM

For the example in Figure 3, η_{ec} is 12.52 dB and η_{nlp} is 17.50 dB. The gain of using NLP after echo cancellation is 4.98 dB. When

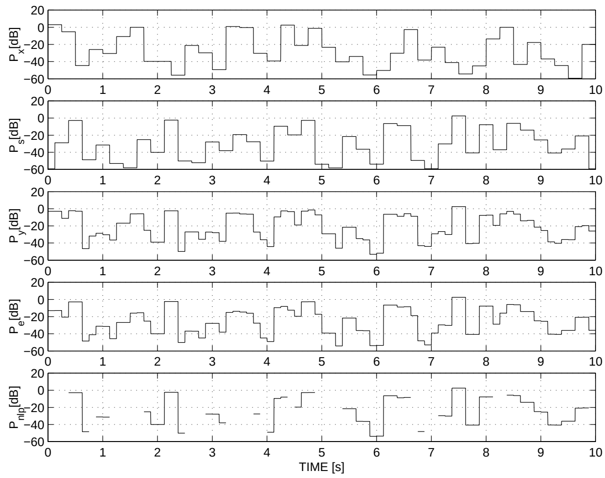

Fig. 3. Optimal NLP without delay.

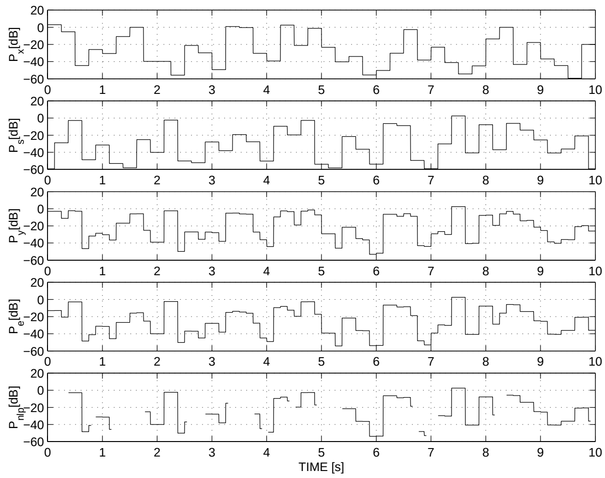

Fig. 4. Optimal NLP with hangover and operate delays.

we use the operate and hangover delays, as in the example in Figure 4, η_{nlp} becomes 16.28 dB, i.e. the gain of using NLP after echo cancellation is 3.76 dB. The gain decreased with respect to the previous one since the NLP reacts optimally but after a delay. A compromise should be found between the immunity to the fast successive switching and η_{nlp} maximization.

Let us assume that the evolution of the average power of far-end and near-end signal can be described by a sequence of i.i.d. random variables. In that case, we can try to find analytically the distorsion/residual echo expected average power and consequently the NLP gain. In general, the expected value of the distorsion/residual echo average power $D = |P_{nlp} - P_s|$ can be found as:

$$E\{D\} = \iint\limits_{(P_x,P_s)\in\mathcal{B}_i} \rho_0 P_x f(P_x) g(P_s) dP_x dP_s + \iint\limits_{(P_x,P_s)\in\mathcal{B}_a} P_s f(P_x) g(P_s) dP_x dP_s, \qquad (13)$$

where $f(P_x)$ and $g(P_s)$ are probability density functions for the average power of far-end and near-end signal, respectively. Also, $\mathcal{B}_i$ and $\mathcal{B}_a$ are the inactivity and activity region in the P_x-P_s plane, respectively. They depend on the support of $f(P_x)$ and $g(P_s)$, as well as on ρ_0 and $\hat{\rho}_0$. In the special case when $f(P_x)$ and $g(P_s)$ are log-uniform distributions (as in the examples corresponding to Figure 3 and 4) we obtain for

$\underline{2\hat{\rho}_0 - \rho_0 \leq L_0/L_1}$:

$$E\{D\} = \rho_0 \frac{L_1 - L_0}{\ln(L_1/L_0)}, \qquad (14)$$

$\underline{L_0/L_1 < 2\hat{\rho}_0 - \rho_0 \leq 1}$:

$$E\{D\} = \frac{1}{(\ln(L_1/L_0))^2}[2\hat{\rho}_0 L_1 - \frac{2\hat{\rho}_0}{2\hat{\rho}_0 - \rho_0} L_0 - (L_0 + \rho_0 L_1)\ln(2\hat{\rho}_0 - \rho_0) - (1+\rho_0) L_0 \ln(L_1/L_0)], \qquad (15)$$

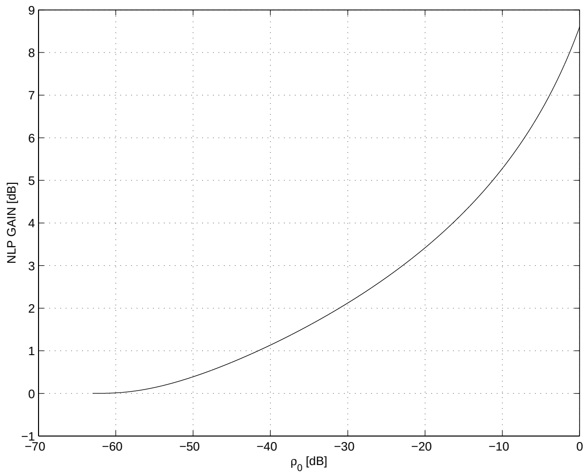

Fig. 5. Gain of using NLP after echo cancellation as a function of steady state far-end reflection coefficient.

$\underline{1 < 2\hat{\rho}_0 - \rho_0 \leq L_1/L_0}$:

$$E\{D\} = \frac{1}{(\ln(L_1/L_0))^2}[-2\hat{\rho}_0 L_0 + \frac{2\hat{\rho}_0}{2\hat{\rho}_0 - \rho_0} L_1 + (L_1 + \rho_0 L_0)\ln(2\hat{\rho}_0 - \rho_0) - (1+\rho_0) L_0 \ln(L_1/L_0)], \qquad (16)$$

$\underline{2\hat{\rho}_0 - \rho_0 > L_1/L_0}$:

$$E\{D\} = \frac{L_1 - L_0}{\ln(L_1/L_0)}, \qquad (17)$$

where L_0 and L_1 are respectively the lower and upper boundary of the support of $f(P_x)$ and $g(P_s)$, and $\hat{\rho}_0 \in [0, 1]$.

Figure 5 shows the dependence of NLP gain on the steady state far-end reflection coefficient ρ_0 and it is obtained by using (14)-(17), where $\rho_0 = \hat{\rho}_0$, $L_0 = 10^{-6}$ and $L_1 = 10^{0.3}$. It is clear that when the reflection coefficient becomes very small there is no benefit of using NLP.

Assuming that the actual steady state far-end reflection coefficient is -30 dB, Figure 6 shows how the NLP gain depends on the

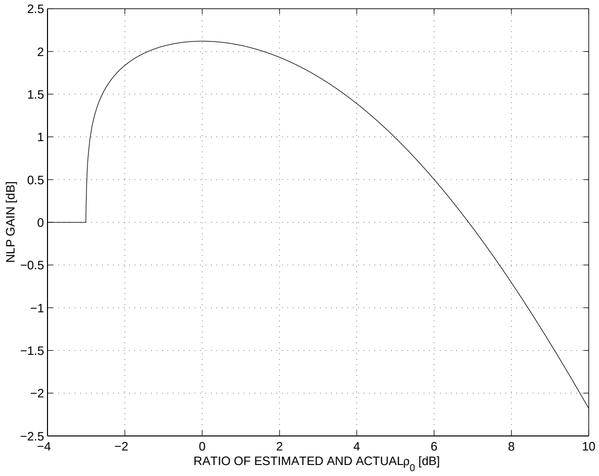

Fig. 6. Gain of using NLP after echo cancellation as a function of ratio of estimated and actual steady state far-end reflection coefficient.

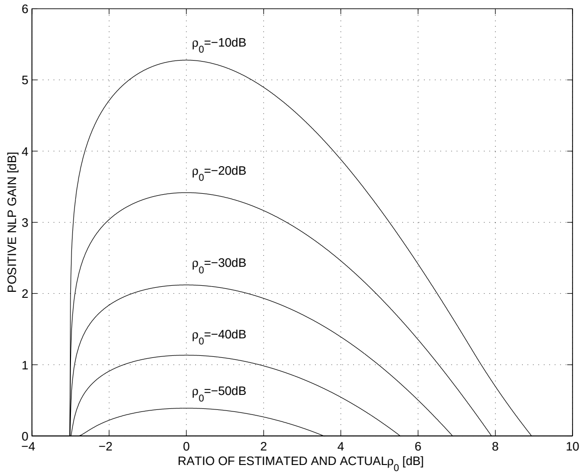

Fig. 7. Positive gain of using NLP after echo cancellation as a function of ratio of estimated and actual steady state far-end reflection coefficient.

ratio between estimated and actual ρ_0. The estimated ρ_0 ($\hat{\rho}_0$) is used in NLP control algorithm. The gain is zero for $\hat{\rho}_0 = \rho_0/2$, and it is maximum for $\hat{\rho}_0 = \rho_0$. It is negative for $\hat{\rho}_0 > 4.89\rho_0$. That is, if $\hat{\rho}_0$ is much higher than ρ_0, there is no gain when using the NLP. But, since the NLP gain curve is flat around the actual value of ρ_0, the NLP is useful in a wide range of $\hat{\rho}_0$. This is especially the case when ρ_0 is higher, as can be seen in Figure 7. The plots in the figure clearly show that the support of the positive NLP gain increases for higher ρ_0. That is, even estimates with a modest accuracy should be beneficial. The results shown in Figure 5, 6 and 7 are consistent with simulation results presented in [4].

Note that the blocking inactivity condition based on $\hat{\rho}_0$ can be written as

$$P_s(n) \geq \rho_0 P_x(n) + 2(\hat{\rho}_0 - \rho_0)P_x(n). \tag{18}$$

When $\hat{\rho}_0 > \rho_0$, for a non-suppressed signal, the near-end component is stronger than the residual far-end component. The weakest near-end signal that can go through and be heard has the power equal to the residual far-end signal, in the case when $\hat{\rho}_0 = \rho_0$. In the case when $\hat{\rho}_0 < \rho_0$, for a non-suppressed signal, the residual far-end signal component can be larger than the near-end component. That is, by underestimating ρ_0, the algorithm is using less often the suppression mechanism. The price of overestimating ρ_0, and using more often the suppression mechanism, is that some near-end signals will not be heard, although they are stronger than the residual far-end signal.

The proposed criterion does not incorporate effects of echo delay to the subjective quality of returned signal. A low-power echo received after a large delay can be unacceptable, although the same one can be tolerable, if received after a short delay. How to quantify echo delay into an objective criterion is an open question. One straightforward idea is to increase the NLP activity region by changing $\hat{\rho}_0$ proportionally to the delay.

6. CONCLUSION

A criterion based on the maximization of an objective measure is proposed for optimal NLP control. An echo suppression algorithm based on the criterion is derived and tested. The essential assumption for application of the optimal NLP control is that the echo canceller has reached the convergence and that an estimate of steady state far-end reflection coefficient is available. In order to obtain the estimate that is reliable, the near-end background noise estimation can be helpful. The advantages of the optimal control are illustrated. Further testing is needed in the case of speech signals. Also, effects of echo delay should be incorporated in the criterion.

7. REFERENCES

[1] S. Gay and J. Benesty, *Acoustic Signal Processing for Telecommunication*, Kluwer, Boston, 2000.

[2] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, Englewood Cliffs, third edition, 1996.

[3] "*ITU-T Recommendation G.168: Digital Network Echo Cancellers*," International Telecommunication Union, Geneva, Switzerland, April 1997.

[4] M. Doroslovački, "Non-linear processor control for echo suppression after echo cancellation," in *Proc. Conf. Inform. Sci. Syst.*, Baltimore, MD, March 2001.