A CLASS OF DOUBLE-TALK DETECTORS BASED ON THE HOLDER INEQUALITY

Constantin Paleologu† Jacob Benesty⋆ Tomas Gaensler‡ Silviu Ciochină‡

† University Politehnica of Bucharest, Romania, e-mail: {pale, silviu}@comm.pub.ro
⋆ INRS-EMT, University of Quebec, Montreal, Canada, e-mail: benesty@emt.inrs.ca
‡ mh Acoustics, Summit, NJ, USA, e-mail: tfg@mhacoustics.com

ABSTRACT

Most of the echo cancellers are equipped with a double-talk detector (DTD) in order to control the behavior of the adaptive filter during double-talk situations. In this paper, we propose a class of DTDs based on the Holder inequality. These DTDs are simple to implement, have low computational complexity, and perform well even for low echo-to-noise ratios. As a particular case, it is shown that the well-known Geigel algorithm can be obtained from this approach.

Index Terms— Echo cancellation, double-talk detector (DTD), Holder inequality, Geigel algorithm.

1. INTRODUCTION

In both network and acoustic echo cancellation an adaptive filter is used to identify the echo path, providing at its output a replica of the echo [1]. This is further subtracted from the reference signal in order to cancel the echo, recovering the near-end signal from the error signal of the adaptive filter. Consequently, the adaptive filter works in a “system identification” scenario [2], since the main goal is to model an unknown system (i.e., the echo path). Nevertheless, the scheme can be also interpreted as an “interference cancelling” configuration [2], because we aim to cancel an undesired signal (i.e., the echo) that corrupts the near-end signal. However, even if the formulation of the problem is straightforward, there are several factors that can bias the echo canceller behavior. Among these, the double-talk situation is probably the most challenging.

In general, in the absence of the near-end speech (i.e., the single-talk case), the adaptive filter works quite well with, of course, an appropriate choice of parameters. The situation becomes far more complicated when the talkers on both sides speak simultaneously, i.e., the double-talk scenario. In this case, the near-end speech acts like a large level of uncorrelated disturbance to the adaptive filter, and it may cause it to divergence. For this reason, the echo canceller is usually equipped with a double-talk detector (DTD) [1] in order to control the behavior of the adaptive filter during these periods. The standard procedure is to slow down or completely halt the adaptation process during double talk.

Many very interesting DTDs have been proposed in the literature. Maybe the simplest one is the well-known Geigel algorithm [3], which provides an efficient and low-complexity solution especially for network echo cancellation. Other more complex algorithms, e.g., based on the coherence and cross-correlation methods [4], [5], have proven to give better results but with an increased computational complexity. In [6], a noise-robust DTD based on normalized cross-correlation and a noise offset was developed. In this paper, we propose a class of DTDs based on the Holder inequality.

2. SOME PRELIMINARIES AND THE HOLDER INEQUALITY

Let us consider the real-valued vector

\[ \mathbf{a} = [a_0 \quad a_1 \quad \cdots \quad a_{L-1}]^T \]

of length \( L \), where superscript \( T \) denotes transposition. The \( \ell_1, \ell_2, \) and \( \ell_\infty \) (maximum) norms [7] of the vector \( \mathbf{a} \) are defined as, respectively,

\[ \|\mathbf{a}\|_1 = \sum_{l=0}^{L-1} |a_l|, \]

\[ \|\mathbf{a}\|_2 = \sqrt{\sum_{l=0}^{L-1} a_l^2} = \sqrt{\mathbf{a}^T \mathbf{a}}, \]

and

\[ \|\mathbf{a}\|_\infty = \max_{0 \leq l \leq L-1} |a_l|. \]

It can be shown that [7]

\[ 1 \leq \frac{\|\mathbf{a}\|_1}{\|\mathbf{a}\|_2} \leq \sqrt{L}, \]

\[ 1 \leq \frac{\|\mathbf{a}\|_1}{\|\mathbf{a}\|_\infty} \leq L, \]

\[ 1 \leq \frac{\|\mathbf{a}\|_2}{\|\mathbf{a}\|_\infty} \leq \sqrt{L}. \]
These inequalities are very important since the ratios of different vector norms are lower and upper bounded by values independent of the characteristic of the vector.

Let \( \mathbf{a} \) and \( \mathbf{b} \) be two vectors of length \( L \), the Holder inequality [7] states that

\[
| \mathbf{a}^T \mathbf{b} | \leq \| \mathbf{a} \|_p \| \mathbf{b} \|_q, \quad \frac{1}{p} + \frac{1}{q} = 1.
\]

(8)

In particular,

\[
| \mathbf{a}^T \mathbf{b} | \leq \| \mathbf{a} \|_\infty \| \mathbf{b} \|_1, \quad (9)
\]

\[
| \mathbf{a}^T \mathbf{b} | \leq \| \mathbf{a} \|_2 \| \mathbf{b} \|_2. \quad (10)
\]

3. DOUBLE-TALK DETECTION ALGORITHMS BASED ON THE HOLDER INEQUALITY

Let us consider the single-talk echo cancellation scenario. In this case, the microphone signal (i.e., the reference or desired signal) is

\[
y(n) = \mathbf{h}^T \mathbf{x}(n) + \nu(n),
\]

(11)

where \( n \) is the discrete-time index,

\[
\mathbf{h} = [ h_0 \ h_1 \ \cdots \ h_{L-1} ]^T
\]

(12)
is the impulse response (of length \( L \)) of the system that we need to identify (i.e., the echo path),

\[
\mathbf{x}(n) = [ x(n) \ x(n-1) \ \cdots \ x(n-L+1) ]^T
\]

(13)
is a vector containing the most recent \( L \) samples of the far-end zero-mean signal \( x(n) \) (i.e., the input signal of the system), and \( \nu(n) \) is a zero-mean additive noise signal (i.e., the background noise at the near-end), which is independent of \( x(n) \).

From (9) and (11), we get

\[
|y(n)| \leq | \mathbf{h}^T \mathbf{x}(n) | + | \nu(n) | \\
\leq \| \mathbf{h} \|_\infty \| \mathbf{x}(n) \|_1 + | \nu(n) |.
\]

(14)

Now, from (14), we can deduce a first detection statistic as

\[
\xi_1 = T_\infty \| \mathbf{x}(n) \|_1 + \sigma_\nu,
\]

(15)

where \( T_\infty \) is a threshold that obviously depends on \( \| \mathbf{h} \|_\infty \) and \( \sigma_\nu^2 = E[\nu^2(n)] \) is the variance of the noise with \( E[\cdot] \) denoting mathematical expectation. Consequently, if \( \xi_1 \geq |y(n)| \), we can state that there is no double talk but if \( \xi_1 < |y(n)| \), we can declare double talk.

Also, we can use (9) differently to obtain

\[
|y(n)| \leq | \mathbf{h}^T \mathbf{x}(n) | + | \nu(n) | \\
\leq \| \mathbf{h} \|_1 \| \mathbf{x}(n) \|_\infty + | \nu(n) |.
\]

(16)

Therefore, based on (16), a second detection statistic can be deduced as

\[
\xi_2 = T_1 \| \mathbf{x}(n) \|_\infty + \sigma_\nu,
\]

(17)

where \( T_1 \) is an approximation of \( \| \mathbf{h} \|_1 \). Thus, if \( \xi_2 < |y(n)| \), double talk is declared but for \( \xi_2 \geq |y(n)| \), we have no near-end speech. This algorithm can be seen as a generalization of the Geigel algorithm [3] since the noise is taken into account. It is known that the detection statistic of the Geigel DTD is defined as

\[
\xi_G = T_G \| \mathbf{x}(n) \|_\infty
\]

(18)

and the double talk is declared when \( \xi_G < |y(n)| \). As we can see from (18), the existence of the system noise is not taken into account. Consequently, the Geigel DTD may perform poorly when the level of the background noise is high, interpreting this situation as double talk.

Finally, using (10), we get

\[
|y(n)| \leq | \mathbf{h}^T \mathbf{x}(n) | + | \nu(n) | \\
\leq \| \mathbf{h} \|_2 \| \mathbf{x}(n) \|_2 + | \nu(n) |.
\]

(19)

Based on (19), a third detection statistic can be defined as

\[
\xi_3 = T_2 \| \mathbf{x}(n) \|_2 + \sigma_\nu,
\]

(20)

where the threshold \( T_2 \) depends on \( \| \mathbf{h} \|_2 \). Again here \( \xi_3 \) is compared to \( |y(n)| \). Condition \( \xi_3 < |y(n)| \) implies double talk, otherwise there is no double talk.

As we can see, all the previous developed DTDs are based on the Holder inequality. The derived detection statistics [see (15), (17), and (20)] take into account the existence of the system noise, in terms of its variance. In practice, this parameter can be easily estimated during silences. The computational complexity of the proposed DTDs are similar to the Geigel algorithm. Regarding the computational complexity of (15) and (20), the required input signal norms \( \| \mathbf{x}(n) \|_1 \) and \( \| \mathbf{x}(n) \|_2 \) can be efficiently computed in a recursive way.

The main problem is how to choose the thresholds \( T_\infty \), \( T_1 \), and \( T_2 \). These parameters depend on \( \| \mathbf{h} \|_\infty \), \( \| \mathbf{h} \|_1 \), and \( \| \mathbf{h} \|_2 \), respectively, which are unavailable in practice. However, let us remember that the threshold \( T_1 \) is similar to the Geigel threshold \( T_G \), which is chosen assuming a minimum echo path attenuation. Also, we know the following inequalities [see (5) and (6)]:

\[
\| \mathbf{h} \|_1 \leq \sqrt{L} \| \mathbf{h} \|_2,
\]

(21)

\[
\| \mathbf{h} \|_1 \leq L \| \mathbf{h} \|_\infty.
\]

(22)

Consequently, from (21) and (22), we get

\[
T_2 \geq \frac{T_1}{\sqrt{L}},
\]

(23)

\[
T_\infty \geq \frac{T_1}{L}.
\]

(24)
Therefore, after we set the threshold $T_1 = T_G$ (similar to the Geigel DTD), the other thresholds can be chosen based on (23) and (24). Here, it could be useful to know or estimate the sparseness degree of the echo path, i.e., the number of “active” or non-zero coefficients (denoted by $M$) [8], because it makes more sense to use $M$ instead of $L$ in (23) and (24).

4. SIMULATION RESULTS

Simulations were performed in the context of network echo cancellation. The echo path has 512 coefficients (the sampling frequency is 8 kHz); it is the first impulse response from ITU-T G168 Recommendation [9], padded with zeros [Fig. 1(a)]. The same length is used for the adaptive filter. The far-end signal is a white Gaussian noise. The output of the echo path (i.e., the echo signal) is corrupted by an independent white Gaussian noise with different values of the echo-to-noise ratio (ENR). The near-end speech appears between times 2 and 4 [Fig. 1(b)]. The performance measure is the normalized misalignment (in dB), defined as $20 \log_{10} \frac{\| \mathbf{h} - \mathbf{h}(n) \|_2}{\| \mathbf{h} \|_2}$, where $\mathbf{h}(n)$ denotes the adaptive filter at time index $n$.

The well-known normalized least-mean-square (NLMS) algorithm [2] is used for adaptation, i.e.,

$$\mathbf{h}(n) = \mathbf{h}(n - 1) + \frac{\mu \mathbf{x}(n)e(n)}{\delta + \mathbf{x}^T(n)\mathbf{x}(n)},$$

where $e(n) = y(n) - \mathbf{x}^T(n)\mathbf{h}(n - 1)$ is the error signal, $\mu$ is the normalized step-size, and $\delta$ is the regularization constant. In our simulations, we set $\mu = 0.2$ and $\delta = 20\sigma_x^2$ (where $\sigma_x^2$ is the input signal variance). We assume that the power of the background noise, $\sigma_n^2$, is known. As we have mentioned in the previous section, this parameter can be easily estimated during silences.

The proposed DTDs based on the Holder inequality are compared to the Geigel DTD. The hangover time for all the DTDs is set to 240 samples. Following the discussion from the previous section, the threshold of the Geigel algorithm is set equal to the threshold of the second Holder DTD [see (17)], i.e., $T_G = T_1 = 0.35$. The other thresholds are set based on the inequalities (23) and (24). Let us assume that the number of active coefficients of the echo path is $M = 64$. This implies that $T_\infty \geq 0.35/64 \approx 0.0055$ and $T_2 \geq 0.35/8 \approx 0.0437$. The experimental results show that the last threshold matches very well the previous estimated value, so that we set $T_2 = 0.0437$. However, the threshold $T_\infty$ needs to be chosen lower; in our simulations we set $T_\infty = 0.0025$.

In the first experiment, a low level background noise is considered at the near-end such that ENR = 20 dB. The re-
results are presented in Fig. 2. As expected, the DTDs based on the Holder inequality perform similarly to the Geigel DTD, since the noise level is low. In Fig. 3, the detection statistics of the DTDs are plotted, as compared to the detection threshold |y(n)|. Again, the detection statistics of the Holder-based DTDs behave similarly to the Geigel DTD.

The advantage of the proposed DTDs become more apparent for a high level of the background noise. The next experiment is performed using ENR = 0 dB. The results are provided in Fig. 4. It can be noticed that the Geigel DTD does not perform well in this case. Due to the high level of the noise, the Geigel algorithm confuses this situation with double talk and inhibits the adaptive algorithm, resulting in a slow convergence rate. On the other hand, the proposed DTDs take into account the system noise (in terms of its power estimate), being robust in this case. The evolution of the detection statistics (Fig. 5) also support these aspects.

In practice, it is not always easy to estimate the sparseness character of the echo path, i.e., the number of active coefficients. Consequently, the thresholds $T_\infty$ and $T_2$ are not very easy to set. However, the threshold $T_1$ is chosen in a simpler way, similarly to the choice of the threshold $T_G$ from the Geigel algorithm. Moreover, using $T_1 = T_G$, this Holder-based DTD is more robust as compared to the Geigel DTD, especially for high level background noise. Therefore, among the three proposed DTDs based on the Holder inequality, the second algorithm [defined in (17)] can be considered as the most practical one.

5. CONCLUSIONS

A class of DTDs based on the Holder inequality has been proposed in this paper. These DTDs have low computational complexity, are simple to implement, and perform well even for a high level of the background noise. The well-known Geigel algorithm can be also derived as a particular case of this approach; however, it is outperformed by the proposed algorithms. From a practical point of view, the Holder-based DTD given in (17) can be considered as the most suitable one for real-world echo cancellation applications.

6. REFERENCES


