Proc. European Conf. Speech Commun., Technol. (Budapest, Hungary), pp. 2539-2542, Sept. 1999

DELAY ESTIMATION FOR TRANSFORM DOMAIN ACOUSTICAL ECHO CANCELLATION

Rabih Abouchakra[*] *Peter Kabal*

Dept. Electrical & Computer Engineering
McGill University, Montreal, Quebec, Canada H3A 2A7

ABSTRACT

Acoustic echo cancellation can be used to remove talker feedback in hands-free systems. Fast convergence and good tracking capabilities cannot be achieved by classical transform domain adaptive filtering algorithms when the reference signal has a variable rank autocorrelation matrix. During the low rank phases of the speech signal, some of the transform-domain tap coefficients become irrelevant to the adaptation process and stop adapting. When the autocorrelation matrix gains full rank, there will be no longer any “frozen” weights. In this paper, we focus on the DCT-LMS algorithm and present a new method using a DCT based delay estimate from other coefficients to move the frozen weights closer to the optimal point and, consequently, reduce the overall re-convergence time.

1 INTRODUCTION

In teleconferencing, acoustic echo appears when the conference room is operating with open microphones and loudspeakers in full duplex mode. The standard approach to eliminate the acoustical echo is to use an adaptive filter. The filter is used to characterize the changing acoustical path between the speaker and the microphone. The synthesized replica is then subtracted from the microphone signal [1].

The most widely used adaptation algorithm is the *least mean squares* or LMS algorithm. Its main disadvantage is its slow convergence. Preprocessing the inputs to the LMS filter with a fixed transformation that decorrelates partially the inputs, will speed convergence. In this paper, we will focus our analysis on the *Discrete Cosine Transform-LMS* algorithm.

The dynamic character of speech, including the variations in the rank of the autocorrelation matrix, slows down the convergence of the DCT-LMS algorithm. During low rank periods, some of the transform domain tap-coefficients stop adapting and effectively “freeze”. It is important to note that in many cases rank deficiency is caused by gaps in the spectrum. When the autocorrelation matrix becomes nonsingular, all the filter weights become relevant and start adapting. However, the weights that have been frozen are “far” from their optimal values. Consequently, a large jump in the MSE is expected and additional convergence time is required for the frozen coefficients to track again.

The purpose of this paper is to reduce this convergence time by moving the frozen weights closer to the optimal point, anticipating a change in the rank of the autocorrelation matrix. The key contribution of this work is to model the changes in the echo path impulse response that result from a change in the spacing between the microphone and loudspeaker as a delay calculated in the DCT domain. This model allows us to synthesize any missing parts of the room response by simply delaying the original response. Accordingly, any filter coefficient that has frozen during the low rank phase, can be updated and brought closer to the actual room response DCT by the same mechanism.

2 DCT-LMS ALGORITHM

The DCT may be defined in several different ways. The DCT-II is the discrete cosine transform first reported by Ahmed *et al.* [2] and is the one that will be used for the DCT-LMS algorithm due to its superior decorrelating capabilities [3].

$$X_{DCT}(m) = \sqrt{\frac{2}{N}} k_m \sum_{n=0}^{N-1} x(n) \cos\left[\frac{(2n+1)m\pi}{2N}\right]$$

$$x(n) = \sqrt{\frac{2}{N}} \sum_{m=0}^{N-1} k_m X_{DCT}(m) \cos\left[\frac{(2n+1)m\pi}{2N}\right]$$

for $m, n = 0, \dots, N-1$, where

$$k_j = \begin{cases} 1 & \text{if } j \neq 0 \text{ and } j \neq N, \\ \dfrac{1}{\sqrt{2}} & \text{if } j = 0 \text{ and } j = N. \end{cases}$$

The DCT-LMS algorithm consists of two stages: the first stage (the sliding DCT) acts as a preprocessor that performs the “pseudo-orthogonalization” of the input vector. For that purpose the DCT uses a sliding window, with the computation being performed for each new input sample. This, in turn, enables the LMS algorithm — the second stage — to operate at the incoming data rate as in its conventional form. A general block diagram of the DCT-LMS algorithm is given in Fig. 1. The vector $\mathbf{x}_n$ (consisting of delayed samples of the input signal $x(n)$) is first transformed into another vector $\mathbf{z}_n$:

$$\mathbf{z}_n = \mathbf{C}\,\mathbf{x}_n$$

where the DCT in matrix form is represented as $\mathbf{C}$. Referring to Fig. 1, $\mathbf{w}_n$ represents the transform domain weight vector, and $d(n)$ the reference signal. The error signal $e(n)$ is

$$e(n) = d(n) - \mathbf{w}_n^T \mathbf{z}_n$$

The weight update equation for each weight $w(n;i)$ is

$$w(n+1;i) = w(n;i) + 2\mu_i\, e(n)\, z(n;i) \qquad (1)$$

where

$$\mu_i = \frac{\mu}{E(z(n;i)^2) + \epsilon}$$

is the adaptive step size for the i^{th} DCT component and μ is a positive constant that governs the rate of convergence.

This work was supported by the Natural Sciences and Engineering Research Council, Canada.

[*] R. Abouchakra is now with Bain & Co., Toronto, Canada

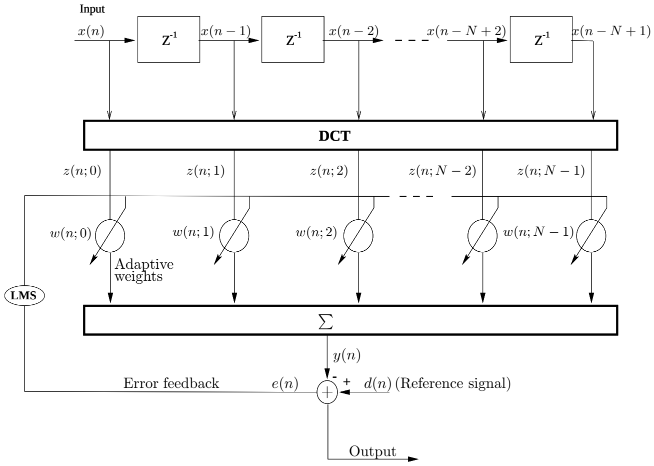

Fig. 1 Block diagram of the DCT-LMS adaptive filter.

The adaptive filter will track the DCT of the room impulse response $\mathbf{h}_n$. If the filter taps match exactly the DCT coefficients of the room impulse response, i.e., if $\mathbf{w}_n = \mathbf{C}\,\mathbf{h}_n$, perfect cancellation will ensue. The reverse implication is not generally true. In the case where some coefficients of $\mathbf{z}_n$ are zero (forming a "gap" in the transform spectrum), the corresponding components of $\mathbf{w}_n$ will not affect the error $e(n)$, and can thus be arbitrary while still achieving perfect cancellation.

If the DCT of the reference signal is null between the frequencies m_o and m_f we will say that the spectrum of the reference signal contains a gap of size M $(M = m_o - m_f)$ starting at m_o. One observes that if $\mathbf{z}_n$ contains a gap, the weights corresponding to the gap will not be updated (see Eq. (1)). In other words, the filter weights at positions m_o through m_f will freeze and stop adapting. On the other hand, $w(n;m)$ continues to be updated correctly for $m \in [0, m_o - 1] \cup [m_f + 1, N - 1]$. In terms of the error surface, the presence of a spectrum gap means that the minimum of the MSE is not unique.

3 SPECTRAL UPDATING

The aim is to estimate the changes in the DCT of the room impulse response during the low rank periods by monitoring the change of the "tracking" coefficients. Later, the frozen weights will be updated to follow the changes in the room impulse response. Since DCT-LMS is used, all the processing should be done in the DCT spectrum domain. The spectral update process is described in Fig. 2.

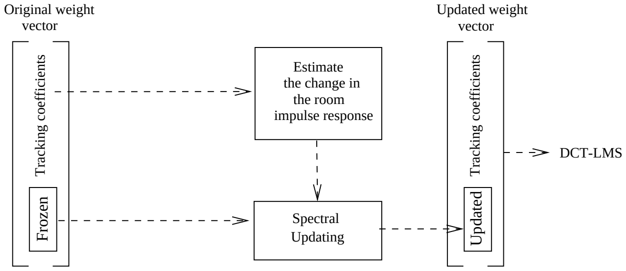

Fig. 2 During low rank phase, the frozen coefficients are updated to track the change of the room impulse response. The change in the room impulse response is estimated by monitoring the changes in the adapting weights.

Consider a teleconferencing scenario where the acoustic echo path is caused by a signal from a loudspeaker being picked up by a microphone. The echo path response (the room response) changes with the movement of the loudspeaker (source) or the microphone (receiver). Further consider a case in which the microphone is attached to a person that moves about the room. The movement of the person (to whom a microphone is attached) is the major variable that will alter the echo path response. Our goal is to quantify the changes in the room impulse response which couple the loudspeaker to the microphone, and come up with a set of parameters based on which the "modified" impulse response can be deduced, given the original response. A radial movement of the source or the receiver with respect to the other, will delay the room impulse response. Therefore, a simple model is to represent all the changes in the room impulse response as a time shift.

The Spectrum Delay Update can be summarized by the following steps:

1. Store $w(n_o - 1; m)$. This vector will serve as a reference for the delay estimation (delay with respect to it). It will be treated as the original room impulse response.
2. Estimate the delay between $w(n;m)$ and $w(n_o - 1; m)$ using only the tracking coefficients ($m \notin [m_o, m_f]$) for $n_o \leq n \leq n_f$.
3. Use the delay estimate to modify the frozen weights. Knowing the original weight vector $\mathbf{w}_{n_o-1}$ and the delay $\delta(n)$, the updated weights $w_{su}(n;m)$ $m \in [m_o, m_f]$ can be calculated.

4 SHIFT PROPERTY OF THE DCT

Consider the k-sample left shifted vector $\mathbf{x}^{k+}$. We wish to derive the relation between its DCT X^{k+}_{DCT} and the transform of the original vector $\mathbf{x}$. We will only consider the case where the last k samples $(x(N) \ldots x(N + k - 1))$ and the first k samples $(x(0) \ldots x(k))$ are zero. This condition is not overly restrictive when $x(n)$ represents a room impulse response with leading samples which are always zero (since the speed of sound is finite). With the above assumption, the general property of the DCT and DST is given by, for $m = 0, 1, \ldots, N - 1$:

$$X^{k+}_{DCT}(m) = \cos(\frac{m\pi k}{N})X_{DCT}(m) + \sin(\frac{m\pi k}{N})X_{DST}(m)$$
$$X^{k+}_{DST}(m) = \cos(\frac{m\pi k}{N})X_{DST}(m) - \sin(\frac{m\pi k}{N})X_{DCT}(m)$$

where the DST X_{DST} (more precisely the DST-II as labeled by [4]) is

$$X_{DST}(m) = \sqrt{\frac{2}{N}}k_{m+1}\sum_{n=0}^{N-1} x(n)\sin\left[\frac{(2n+1)(m+1)\pi}{2N}\right]$$

for $m = 0, \ldots, N - 1$.

The proof of shift property is done by mathematical induction from the one-sample shift relationship [5] and is developed in [6]. The same properties for a *right* shift and its reciprocal shift property are also derived in [6]

5 DELAY ESTIMATION IN THE DCT DOMAIN

We start by considering one fixed frequency m, and try to solve for the delay k

$$X^{k+}_{DCT}(m) = \cos(\frac{m\pi k}{N})X_{DCT}(m) - \sin(\frac{m\pi k}{N})X_{DST}(m)$$

This is an equation of the form

$$a = b\cos(\alpha k) + c\sin(\alpha k)$$

This equation has multiple solutions for the delay k (details in [6]).

If the delay is independent of frequency m, the delay ambiguity can be resolved by examining the solutions at different frequencies. In practice, only frequencies between some m_{min} and m_{max} are used in the estimation process. Our task is to obtain the best delay estimate based on the knowledge of all candidate solutions in the available frequency range. The delay will be the path through the solution candidates with the minimum variance:

1. From each solution candidate at m_{min}, initiate a path (line).
2. Connect the closest points (in terms of the Euclidean Distance), at consecutive frequencies.
3. Associate with each path a metric defined as the sum of the deviations (square of the distance between the point and the line).
4. Compare all the metrics and save the path with the smallest metric (called the *survivor* path).
5. Deduce the delay from the *survivor* path.

Applying the delay estimation algorithm developed here to model a specific receiver movement in various room environment as a delay in the impulse response yields accurate results: the average difference between the estimated delay and the actual delay value ranged between 0.02 sample for “good” rooms and 0.14 sample for “bad” rooms. This is based on estimating the delay each sample. We can reduce computational complexity by estimating the delay less often. The maximum change in delay due to a fast walker for updates every 256 samples (non-overlapped DCT’s) is about 3 samples.

6 EXPERIMENTAL SET UP

To measure the improvements due to the use of the Spectrum Delay Update (SDU), we consider two error measures.

Given a room impulse response DCT vector $\mathbf{H}$ and a filter weights vector $\mathbf{w}$, the Euclidean distance between the weight vectors is

$$d_2(\mathbf{H}, \mathbf{w}) = \sqrt{\sum_{k=1}^{N} |\mathbf{H}_k - \mathbf{w}_k|^2}$$

Based on this distance, we can define the change in *EDMD* (Euclidean Distance Mean Difference) in dB when the SDU modified weights are used.

$$EDMD = E[20\log\{d_2(\mathbf{H}, \mathbf{w}_{su})\} - 20\log\{d_2(\mathbf{H}, \mathbf{w})\}]$$

A second measure is the decrease in the jump in the MSE when the gap vanishes. The average reduction in the MSE will reflect the improvement gained by SDU.

Consider the following experimental set-up:

- Room dimensions are 5 × 4 × 3 m, the receiver is located at position $(1, 1, 1)$ and the source at position $(3, 2, 1)$.
- Signals are sampled at 8 kHz.
- The reference signal is created by filtering a speech segment by the Chebyshev type II bandstop filter. The resulting signal will contain a gap in the DCT spectrum.
- All coefficients of the 256-tap DCT-LMS filter have been initialized to zero.
- The far-end talker’s signal is null (the microphone signal contains only echoes).

In the first phase of the experiment, the receiver moves to $(1, 1, 1 + \Delta x)$, where Δx ranges between 0.05 m and 0.12 m. After an initial convergence period, the tracking coefficients are used to estimate the delay in the impulse response. The frozen coefficients are then updated using SDU. Fig. 3 shows the *EDMD* measure.

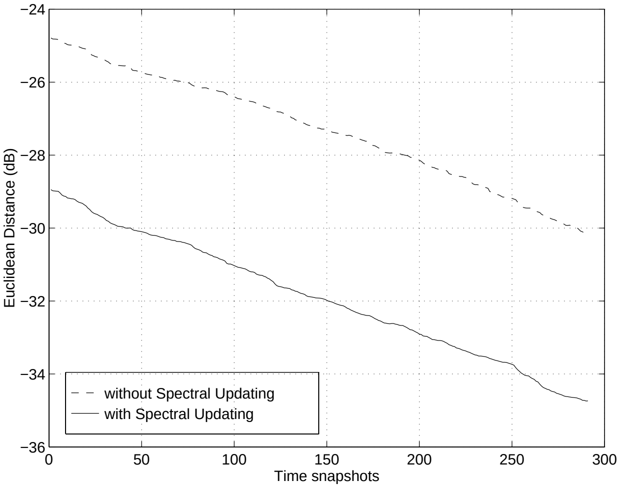

Fig. 3 Euclidean Distance between the filter weights and the room impulse response DCT. Every 10 samples a snapshot of the filter weights is taken. (The reference signal DCT has a gap in positions 90 through 115)

In the second phase of the experiment, the reference signal autocorrelation matrix gains full rank, i.e, its transform no longer contains any gap (the reference signal is not filtered anymore). Since the DCT is computed every sample, it will take 256 samples before the gap vanishes entirely from the spectrum.

The previously frozen coefficients begin adapting from their value at the end of phase one, and a large jump in the MSE is expected. However, if the weight vector was spectrally updated in anticipation of the increase in the rank of $\mathbf{R}$, the resulting MSE is smaller, as shown in Fig. 4. The mean separation between the two MSE curves (in Fig. 4) reflects the improvement gained by SDU; we will refer to it as *MSEMS* (MSE Mean Separation).

7 PERFORMANCE ANALYSIS

Using the experimental set-up outlined in the previous section, with $\Delta x = 0.05$ m, the gap start is fixed to position 95 but the gap end is varied to yield various gap sizes (7 to 34 samples). The two performance measures, *EDMD* and *MSEMS* are calculated for four acoustic environments. The results are shown in Fig. 5 and Fig. 6.

It is obvious from Fig. 5 that SDU brings the filter weights closer to the room impulse response DCT. The acoustic environment, through its effects on the accuracy of the delay based model (representing the change in the room impulse response) and the exactness of the delay estimate, plays an important role in determining the Euclidean Distance gain. It is clear from Fig. 5 that the better the acoustic environment, the bigger the gain. In very reverberant rooms, the change in impulse response

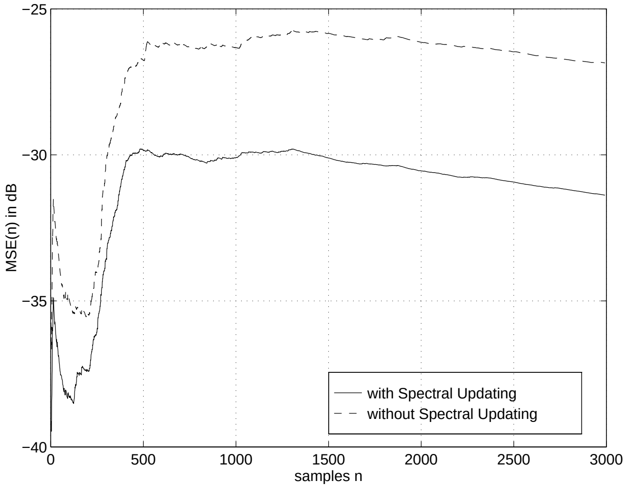

Fig. 4 Evolution of the MSE with time. At $n = 0$, the autocorrelation matrix, $\mathbf{R}$, of the reference signal gains full rank. The dashed curve represents the MSE that results from the coefficients in the gap (positions 90 through 115) being frozen to their value at the end of phase one. The solid curve gives the MSE that results from having spectrally updated the coefficients in the gap in anticipation of the increase in the rank of ($\mathbf{R}$).

due to a movement of the receiver is not well modelled by a time shift. For all environments, *EDMD* increases with the gap size: as the gap gets bigger, more filter coefficients freeze, and more weights are spectrally updated.

The reduction in the MSE that results from spectrally updating the filter coefficients in anticipation of the increase in the rank of $\mathbf{R}$, is shown in Fig. 6. It is clear that *MSEMS* increases with the quality of the acoustic environment and the size of the gap. If more coefficients are frozen in phase one of the experiment (i.e. if the gap is bigger), the benefits of SDU on the MSE become more evident. In the medium room, the reduction in the MSE goes from a fraction of a dB in very small gaps (less than 10 samples) to more than 5 dB for gaps bigger than 30 samples.

Additional analysis has been performed on the performance improvement for a moving receiver. The details can be found in the [6]

8 CONCLUSIONS

The SDU method allows one to update "frozen" DCT coefficients from the "tracking" coefficients. The update is based on an estimate of the change in delay formulated in the DCT domain. Experimental results show that the updated coefficients are closer to the actual values and reduce the convergence time. The amount of improvement depends on the acoustic environment. For rooms with moderate reverberation (for instance, the medium room which is a realistic environment), substantial benefit results from anticipating the change in coefficients due to a change in delay. The results show that SDU can be successfully applied to practical situations.

REFERENCES

[1] O. Muron and J. Sikorav, "Modeling of reverberators and audioconference rooms," *Proc. Int. Conf. Acoust., Speech, Signal Proc.* (Tokyo, Japan), pp. 921–924, 1986.

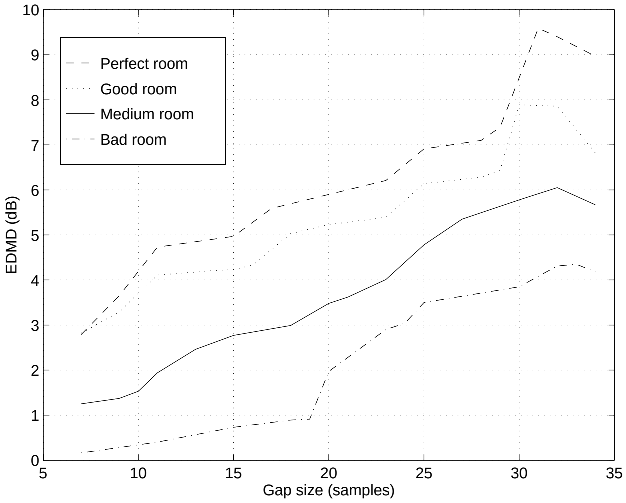

Fig. 5 *EDMD* as a function of the gap size in four different acoustic environments. The gap start is fixed to position 95 and $\Delta x = 0.05$ m.

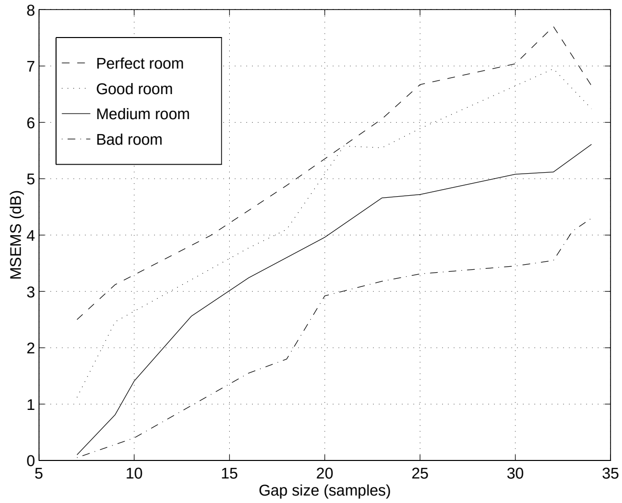

Fig. 6 *MSEMS* with respect to the gap size in four different acoustic environments. The gap start is fixed to position 95 and $\Delta x = 0.05$ m.

[2] N. Ahmed, T. Natarajan, and K. Rao, "Discrete Cosine Transform," *IEEE Trans. Comput.*, vol. 23, pp. 90–93, January 1974.

[3] K. Rao and P. Yip, *Discrete Cosine Transform: Algorithms, Advantages, Applications.* San Diego, Calif.: Academic Press, 1990.

[4] Z. Wang, "Fast algorithms for the Discrete W Transform and for the Discrete Fourier Transform," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. 32, pp. 803–816, August 1984.

[5] P. Yip and K. Rao, "On the shift property of DCT's and DST's," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. 35, pp. 404–406, March 1987.

[6] R. Abouchakra, "Delay estimation for transform domain acoustical echo cancellation," Master's thesis, McGill University, Montreal, Canada, October 1997. (`http://www.tsp.mcgill.ca/thesis`).