Enhancement of Noisy Speech Using Sliding Discrete Cosine Transform

Vitaly Kober

Department of Computer Sciences, Division of Applied Physics
CICESE, Ensenada, B.C. 22860, Mexico
vkober@cicese.mx

Abstract. Denoising of speech signals using a sliding discrete cosine transforms (DCT) is proposed. A minimum mean-square error (MMSE) estimator in the domain of a sliding DCT is derived. In order to provide speech processing in real time, a fast recursive algorithm for computing the sliding DCT is presented. The algorithm is based on a recursive relationship between three subsequent local DCT spectra. Extensive testing has shown that background noise in actual environment such as the helicopter cockpit can be made imperceptible by proper choice of suppression parameters.

1 Introduction

Processing of speech degraded due to additive background noise is of interest in a variety of tasks. For example, many speech transmission and coding systems, whose design is predicated on a relatively noise-free environment, degrade quickly in quality and performance in the presence of background noise. Thus, there is a considerable interest in and application for the development of such systems, which compensate for the presence of noise. In many cases, intelligibility is affected by background noise so that a principal objective of a speech processing system may be to improve intelligibility. Numerous systems have been proposed to remove or reduce background noise [1-8]. These systems provide an apparent improvement in signal-to-noise ratio, but intelligibility is in fact reduced. In this paper, an approach to speech denoising on the base of a sliding DCT is used.

In many filtering and spectral analysis applications, the signals such as speech have inherently infinite length. Moreover, since the signal properties (amplitudes, frequencies, and phases) usually change with time, a single orthogonal transform is not sufficient to describe such signals. As a result, the concept of short-time signal processing with filtering in the domain of an orthogonal transform can be used [9]. The short-time orthogonal transform of a signal x_k is defined as

$$X_s^k = \sum_{n=-\infty}^{\infty} x_{k+n} w_n \psi(n,s), \tag{1}$$

where w_n is a window sequence, $\psi(n,s)$ represents the basis functions of an orthogonal transform. Equation (1) can be interpreted as the orthogonal transform of x_{k+n} as

A. Sanfeliu and J. Ruiz-Shulcloper (Eds.): CIARP 2003, LNCS 2905, pp. 229–235, 2003.

viewed through the window w_n. X_s^k displays the orthogonal transform characteristics of the signal around time k. Note that while increased window length and resolution are typically beneficial in the spectral analysis of stationary data, for time-varying data it is preferable to keep the window length sufficiently short so that the signal is approximately stationary over the window duration.

We assume that the window has finite length around n=0, and it is unity for all $n \in [-N_1, N_2]$. Here N_1 and N_2 are integer values. This leads to signal processing in a sliding window [10]. In other words, local filters in the domain of an orthogonal transform at each position of a moving window modify the orthogonal transform coefficients of a signal to obtain only an estimate of the pixel x_k of the window. The choice of orthogonal transform for sliding signal processing depends on many factors. The DCT is one the most appropriate transform with respect to the accuracy of power spectrum estimation from the observed data that is required for local filtering, the filter design, and computational complexity of the filter implementation. For example, linear filtering in the domain of DCT followed by inverse transforming is superior to that of the discrete Fourier transform (DFT) because a DCT can be considered as the DFT of a signal evenly extended outside its edges. This consequently attenuates boundary (temporal aliasing) effects caused by circular convolution that are typical for linear filtering in the domain of DFT. In the case of DFT, speech frames are usually windowed to avoid temporal aliasing and to ensure a smooth transition of filters in successive frames. For the filtering in the domain of DCT, the windowing operation can be skipped. In such a manner the computational complexity can be further reduced.

The presentation is organized as follows. In Section 2, we present computationally efficient algorithm for computing the sliding DCTs. In Section 3, an explicit filter formula minimizing the MMSE defined in the domain of the sliding DCT is derived. We also test the filter performance in actual environment such as the helicopter cockpit. Section 4 summarizes our conclusions.

2 Fast Algorithm for Computing the Sliding DCT

The discrete cosine transform is widely used in many signal processing applications such as adaptive filtering, video signal processing, feature extraction, and data compression. This is because the DCT performs close to the optimum Karhunen-Loeve transform for the first-order Markov stationary data, when the correlation coefficient is near 0.9 [11]. Four types of DCTs were classified [12]. The DCT discussed in the paper is referred to the type-II. The kernel of the DCT is defined for the order N as

$$DCT = \left\{ k_s \cos\left(\pi \frac{s(n+1/2)}{N} \right) \right\}, \tag{2}$$

where n, $s=0, 1,\ldots, N-1$; $k_s = \begin{cases} 1/\sqrt{2} & \text{if } s = 0, \\ 1 & \text{otherwise.} \end{cases}$. For clarity, the normalization factor $\sqrt{2/N}$ for the forward transform is neglected until the inverse transform. The sliding cosine transform (SCT) is defined as

$$X_s^k = \sum_{n=-N_1}^{N_2} x_{k+n} \cos\left(\pi \frac{(n+N_1+1/2)s}{N}\right), \tag{3}$$

where $N=N_1+N_2+1$, $\{ X_s^k ; s=0, 1,\ldots, N-1\}$ are the transform coefficients around time k. The coefficients of the DCT can be obtained as $\{ C_0^k = X_0^k / \sqrt{2} ; C_s^k = X_s^k , s=1,\ldots, N-1\}$. We now derive fast algorithm for the SCT on the base of a recursive relationship between three subsequent local DCT spectra [13]. The local DCT spectra at the window positions $k-1$ and $k+1$ are given by

$$X_s^{k-1} = \sum_{n=-N_1-1}^{N_2-1} x_{k+n} \cos\left(\pi \frac{(n+N_1+1/2)s}{N} + \frac{\pi s}{N}\right), \tag{4}$$

$$X_s^{k+1} = \sum_{n=-N_1+1}^{N_2+1} x_{k+n} \cos\left(\pi \frac{(n+N_1+1/2)s}{N} - \frac{\pi s}{N}\right). \tag{5}$$

Using properties of the cosine function and equations (4) and (5), we can write

$$X_s^{k+1} = 2X_s^k \cos\left(\frac{\pi s}{N}\right) - X_s^{k-1} + \cos\left(\frac{\pi s}{2N}\right) \times \left(x_{k-N_1-1} - x_{k-N_1} + (-1)^s \left(x_{k+N_2+1} - x_{k+N_2}\right)\right). \tag{6}$$

We see that the computation of the DCT at the window position $k+1$ involves values of the input sequence x_k as well as the DCT coefficients computed in two previous positions of the moving window. The number of arithmetic operations required for computing the sliding discrete cosine transform at a given window position is evaluated as follows: the SCT for the order N with $N=N_1+N_2+1$ requires $2(N-1)$ multiplication operations and $2N+5$ addition operations; the DCT requires one extra operation of multiplication. Table 1 lists numerical results of computational complexity for the proposed algorithm and known fast DCT algorithms. Note that fast DCT algorithms require the length of a moving window to be of a power of 2, $N=2^M$. In contrast, the length of a moving window for the proposed algorithm is an arbitrary integer value determined by the characteristics of the signal to be processed.

We see that the proposed algorithm yields essentially better results when the length of the window increases.

Table 1. Number of multiplications and additions for computing the sliding DCT

M	Fast DCT[14, 15]		Proposed algorithm	
	Mult.	Add.	Mult.	Add.
16	33	81	30	37
32	81	209	62	69
64	193	513	126	133
128	449	1217	254	261
256	1025	2817	510	517

The inverse algorithms for the sliding DCT can be written as follows.

$$x_k = \frac{1}{N}\left(2\sum_{s=1}^{N-1} X_s^k \cos\left(\pi \frac{(N_1 + 1/2)s}{N} \right) + X_0^k \right), \tag{7}$$

where $N=N_1+N_2+1$. The computational complexity is N multiplication operations and N addition operations. If x_k is the central pixel of the window, that is, $N_1=N_2$ and $N=2N_1+1$, then the inverse transform is simplified to

$$x_k = \frac{1}{N}\left(2\sum_{s=1}^{N_1} (-1)^s X_{2s}^k + X_0^k \right). \tag{8}$$

We note that in the computation only the spectral coefficients with even indices are involved. The computation requires one multiplication operation and N_1+1 addition operations.

3 Denoising of Speech Signals in the Sliding DCT Domain

The objective of this section is to develop a noise suppression technique on the base of the sliding DCT, and to test the algorithm performance in actual noise environment. We design locally adaptive filters to enhance noisy speech. Assume that a clean speech signal $\{a_k\}$ is degraded by zero-mean additive noise $\{v_k\}$

$$x_k = a_k + v_k, \tag{9}$$

where $\{x_k\}$ is a noisy speech sequence.

Let $\{ X_s^k, A_s^k, V_s^k, \hat{A}_s^k ;\ s=0, 1,..., N-1\}$ be the DCT transform coefficients around time k of noisy speech, clean speech, noise, and filtered signal, respectively. Here $N=2N_1+1$ is the length of the DCT. Note that N_1 is an arbitrary integer value, which is determined by pitch period of speech. One can be chosen to be approximately as the maximum expected pitch period for adequate frequency resolution.

Various criteria can be exploited for the filter design. In the following analysis we use the criterion of the MMSE around time k which is defined in the domain of DCT, taking into account (8), as follows:

$$MMSE_k = E\left\langle \sum_{t=0}^{N_1} \alpha_t^2 \left[A_{2t}^k - \hat{A}_{2t}^k \right]^2 \right\rangle, \tag{10}$$

where $E\langle . \rangle$ denotes the expected value.

As we mentioned above, the length of the window is chosen in such a way that noise can be considered as stationary in the window. Let $P_t^k = E\left\langle \left| V_t^k \right|^2 \right\rangle$ denote the power spectrum of noise in the domain of DCT. Suppose that $\hat{A}_t^k = X_t^k H_t^k$, here H_t^k is the filter to be designed around time k. By minimizing $MMSE_k$ with respect to H_t^k, we arrive to a version of the Wiener filter in the domain of DCT:

$$H_t^k = \frac{\left| X_t^k \right|^2 - P_t^k}{\left| X_t^k \right|^2} = 1 - \frac{P_t^k}{\left| X_t^k \right|^2}. \tag{11}$$

The MMSE estimation of the processed speech in the domain of the sliding DCT is given by

$$\hat{A}_t^k = \begin{cases} \left(1 - \dfrac{P_t^k}{\left| X_t^k \right|^2} \right) X_t^k, \text{if } \left| X_t^k \right|^2 > P_t^k \\ 0, \qquad \text{otherwise} \end{cases}. \tag{12}$$

The obtained filter can be considered as a spectral subtraction method in the domain of sliding DCT. In general, spectral subtraction methods [1], while reducing the wide-band noise, introduce a new "musical"' noise due to the presence of remaining spectral peaks. To attenuate the "musical" noise, one can suggest oversubtraction of the power spectrum of noise by introducing a nonzero power spectrum bias. Finally, the MMSE estimation of the processed speech in the domain of the sliding DCT can be written as follows:

$$\hat{A}_t^k = \begin{cases} \left(1 - \dfrac{P_t^k}{\left| X_t^k \right|^2} \right) X_t^k, \text{if } \left| X_t^k \right|^2 > P_t^k + B^k \\ 0, \qquad \text{otherwise} \end{cases}, \tag{13}$$

where B^k is a speech-dependent bias value.

The filtered speech signal can be obtained with use of (8). It also follows from (8) that in the estimation only the spectral coefficients with even indices are involved.

A test speech signal recorded in helicopter environment is presented in Fig.1. The data was sampled at 16.00kHz. In our tests the window length of 761 samples is used. The sliding squared DCT coefficients averaged over all positions of the running window for noisy speech is presented in Fig. 2. The power spectrum of noise is obtained by actual measurement from background noise in intervals where speech is not presented. It is shown in Fig. 3. We see the difference in spectral distributions of the speech and the helicopter noise, which will help us to suppress the helicopter noise.

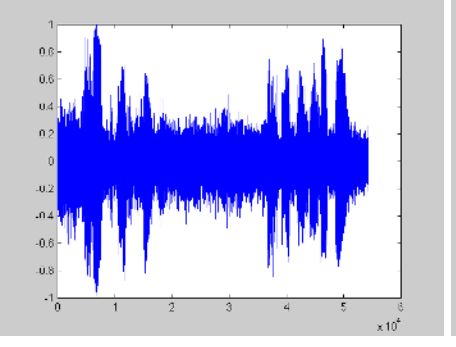

Fig. 1. Time wavefront of helicopter speech

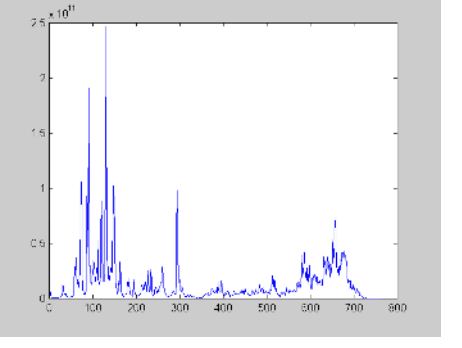

Fig. 2. Average squared DCT magnitude of noisy speech

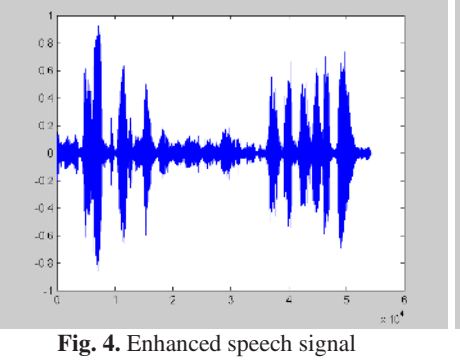

Fig. 4. Enhanced speech signal

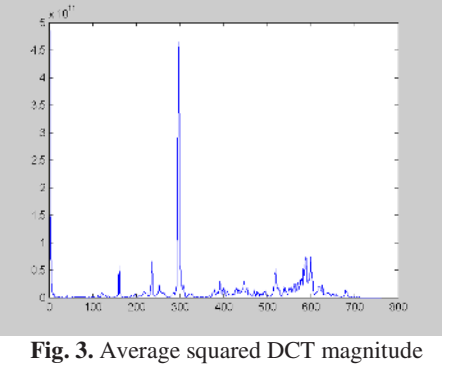

Fig. 3. Average squared DCT magnitude of noise

The result of filtering by using the proposed filter is shown in Fig. 4. It is clearly that the system is capable of significant noise reduction. Numerous formal subjective tests are shown that the helicopter noise can be made imperceptible by proper choice of the filter parameters in (13).

In this section we derived a filter for noise suppression on assumption that speech is always was presented in the measured data. However, if a given frame of data consists of noise alone, then obviously a better suppression filter can be used [5, 6]. In general, an optimal algorithm should include a detector of voiced and unvoiced speech signals. After detecting, different strategies of processing should be applied to voiced and unvoiced speech signals.

4 Conclusions

In this paper, we have presented a new technique for enhancing speech degraded by additive noise. The technique utilizes the sliding DCT. A MMSE estimator in the domain of the sliding DCT has been derived. In order to provide speech processing in real time, a fast recursive algorithm for computing the sliding DCT has been suggested. The algorithm requires essentially less operations of multiplication and addition comparing with known fast DCT algorithms. Extensive testing has shown that background noise such as in the helicopter cockpit can be significantly reduced by proper choice of suppression parameters.

Acknowledgment. This work was supported by the grant 36077-A from CONACYT

References

1. Boll S.F., Suppression of acoustic noise in speech using spectral subtraction, IEEE Trans. Acoust. Speech Signal Process., Vol. ASSP-27, (1979) 113–120.
2. Lim J.S. and Oppenheim A.V., Enhancement and bandwidth compression of noisy speech, Proc. IEEE, Vol. 67, (1979) 1586–1604.
3. McAulay R.J., Malpass M.L., Speech enhancement using a soft-decision noise suppression filter, IEEE Trans. Acoust. Speech Signal Process., Vol. ASSP-28, (1980) 137–145.
4. Ephraim Y. and Malah D., Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator, IEEE Trans. Acoust. Speech Signal Process., Vol. ASSP-32, No. 6, (1984) 1109–1121.
5. Ahmed M.S., Comparison of noisy speech enhancement algorithms in terms of LPC perturbation, IEEE Trans. Acoust. Speech Signal Process., Vol. 37, No. 1, (1989) 121–125.
6. Chen Y.M. and O'Shaughnessy D., Speech enhancement based conceptually on auditory evidence, IEEE Trans. Signal Process., Vol. 39, No.9, (1991) 1943–1953.
7. Ephraim Y., A Bayesian estimation approach for speech enhancement using hidden Markov models, IEEE Trans. Signal Process., Vol. 40, (1992) 725–735.
8. Hu H.T., Robust linear prediction of speech signals based on orthogonal framework, Electronics Letters, Vol. 34, No 14, (1998) 1385–1386.
9. Oppenheim A.V., Shafer R.W., Discrete-time signal processing, Prentice Hall, Englewood Cliffs, NJ (1989).
10. Vitkus R.Y., and Yaroslavsky L.P., Recursive algorithms for local adaptive linear filtration, in: Mathematical Research., Eds.: Yaroslavsky L.P., Rosenfeld A., and Wilhelmi W., Academy Verlag, Berlin, (1987) 34–39.
11. Jain A.K., A sinusoidal family of unitary transforms, IEEE Trans. Pattern Anal. Machine Intell., Vol. PAMI-1, No 4, (1979) 356–365.
12. Wang Z., Fast algorithms for the discrete W transform and for the discrete Fourier transform, IEEE Trans. Acoust. Speech Signal Process., Vol. ASSP-32, No 4, (1984) 803–816.
13. Kober V. and Cristobal G., Fast recursive algorithms for short-time discrete cosine transform, Electronics Letters, Vol. 35, No 15, (1999) 1236–1238.
14. Hou H.S. , A fast recursive algorithm for computing the discrete cosine transform, IEEE Trans. Acoust. Speech Signal Process., Vol. ASSP-35, No 10, (1987) 1455–1461.
15. Britanak V, On the discrete cosine computation, Signal Process., Vol. 40**,** No 2-3, (1994) 183–194.