Comparative Analysis for Discrete Sine Transform as a suitable method for noise estimation

Swati Dhamija[1] and Dr. Priyanka Jain[2]

[1] Quality Engineer, Aircom International, Gurgaon

[2] Assistant Professor, Delhi Technical University, New Delhi

Abstract

When the speech signal corrupted by noise is processed using different transforms like the Discrete Fourier Transform, Discrete Cosine Transform and the Discrete Sine Transform , a comparative analysis proves that the Discrete Sine Transform (DST) is most suitable for de-noising and therefore reconstruction of the original speech. The experimental results show that the discrete sine transform coefficients at the low frequency regions are predominantly speech, and at the high frequency regions are predominantly noise. According to this, a new noise estimation method based discrete sine transform is proposed in this paper. The usage of DST coefficients reveals that the energy distribution throughout the segment of speech is uniform.

Keywords: *Digital Speech Processing, Discrete Sine Transform, Discrete Cosine Transform, Noise Estimation.*

1. Introduction

Digital speech processing is an important part of the communication domain, in the course of speech technology. However, because of the presence of noise, the speech which the receiver gets is not the original speech. In order to obtain as pure as possible speech signal, we need to estimate the noise, but, because the noises are root in a large number of origins and are a mass of kinds, it is almost impossible to obtain the completely pure speech from the noised speech. The mainstream algorithms for speech enhancement nowadays are mainly based on Fourier Transform (DFT).

DST, DFT, DCT as well as K-L transform are reduced into an identical transformation family in comparison with the first order Markov process called the “sine family”. This means that the DST has the similar nature with these three kinds of transformations. DFT, DCT as well as K-L transformation are all applied to the speech enhancement and there is no correlated literature on sine transform.

The binding energy ability of DST is not as strong as DFT and DCT but that does not imply that the strongest is the best. The experiment shows that DST is better in the processing of white noise as compared to DFT and DCT.

2. Definition and Characteristics of DST

The Discrete Sine Transform (DST) is a Fourier-related transform similar to the discrete Fourier Transform (DFT), but using a purely real matrix. It is equivalent to the imaginary parts of a DFT of roughly twice the length, operating on real data with odd symmetry (since the Fourier transform of a real and odd function is imaginary and odd), where in some variants the input and/or output data are shifted by half a sample.

For an assigned sequence x(n), its DST transform and inverse transform are defined separately as:

$$X[k] = \sqrt{\frac{2}{N+1}} \sum_{n=1}^{N} x(n) \sin \frac{nk\pi}{N+1} \qquad (1)$$

$$x(n) = \sqrt{\frac{2}{N+1}} \sum_{k=1}^{N} X[k] \sin \frac{nk\pi}{N+1} \qquad (2)$$

Where n=1, 2, 3…N; k=1, 2, 3…N.

DST, which was originally, developed by Jain (type I) and Kekra and Solanka (type II), belong to the family of unitary transform. This unitary family includes the DST, discrete cosine transforms (DCT) and the discrete Fourier transforms (DFT).

Figure 1 presents the waves of a length of speech signal and the waves after three kinds of transforms, for convenience of comparison only the absolute parts are drawn for DFT. We can see from the figure that DCT is the best in binding energy. The transformed speech energy mainly concentrates on the low frequency area and the noise mainly concentrates on the high frequency area.

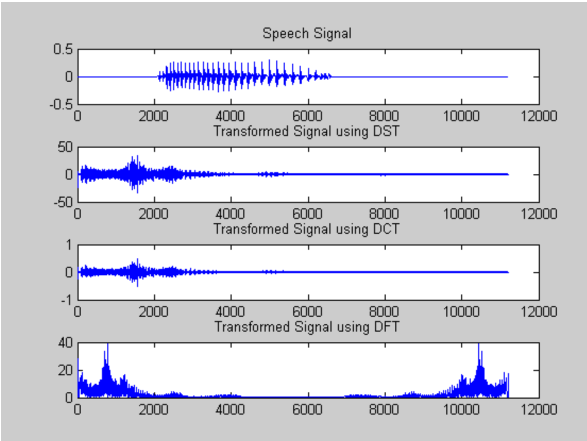

Figure 1. Speech signal transformations

3. Reconstruction of Original Speech

The inverse of the transformed signal results in the signal itself. However, the conversion of the signal to its transform and then inverse of the transformed signal to reconstruct the original speech waveform causes a shift in the coefficients thereby introducing noise in the original signal. The speech signal mainly concentrates on the low frequency part, and the noise energy mainly concentrates on the high frequency unit, so the estimation of noise is mainly done the high frequency area.

Figure 2 represents the waves of the reconstructed speech signal after performing the three inverse transformations.

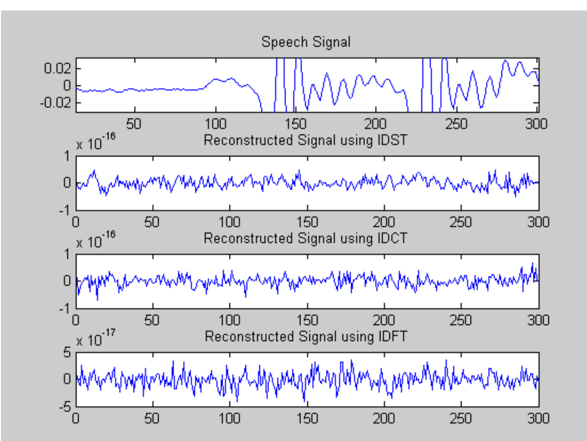

Figure 2. Reconstructed speech

3.1 Calculation of SNR

After the reconstruction of the signal using the three different transforms, we can calculate the Signal-to-Noise Ratios (SNR).The lower the signal to noise ratio is, the more accurate the noise estimate is.

We can compute the SNR as follows, where x(n) is the clean speech, $\tilde{x}(n)$ is the de-noised speech:

$$SNR = 10 * \log \frac{\sum_{n=0}^{N-1} x(n)^2}{\sum_{n=0}^{N-1} [x(n) - \tilde{x}(n)]^2} \tag{3}$$

The computed values of SNR for the three transforms are as follows for a speech signal with length N, which is set to 11220.

Table 1: SNR Comparisons

Transformation Used	SNR value
DST	683.3081
DCT	703.4444
DFT	701.4479

The SNR values computed in Table1 imply that DST has the lowest value. This proves that the noisy components are maximum due to sine transformation, thereby making it suitable for noise estimation.

3.2 Noise Estimation using DST

For noise estimation, take the discrete sine transform on signal with length L, then take estimation in the last N/2 sine coefficients, that is the high frequency region.

$$\sigma_{v1}^2 = Var(The last N/2 of DST coefficients) \tag{4}$$

The discrete sine transform is an orthogonal transformation, so it cannot keep the energy of the signal to be invariable after transformation; therefore the variance we estimated from (4) is not the true noise variance. We add the white Gaussian noise to a speech with length L, which L is set to 11220, then take the DST of the signal and compute the estimated noise by the formula:

$$\sigma_v^2 = 2\sigma_{v1}^2 / L \tag{5}$$

By estimating the noise we get the following data:

Table 2 Estimated Variance and Errors.

SNR	Actual Variance	Estimated Variance	Error
-15	9.0092e+004	16.0593	9.0076e+004
-10	2.7287e+004	4.8640	2.7282e+004
-5	9.1577e+003	1.6324	9.1561e+003
0	2.8220e+003	0.5030	2.8215e+003
5	879.5745	0.1568	879.4177
10	283.8823	0.0506	283.8317
15	88.1087	0.0157	88.0930

The set of experiments done in Table2 were then repeated for DCT and DFT. Figure3 shows the noise estimation using the three different transformations.

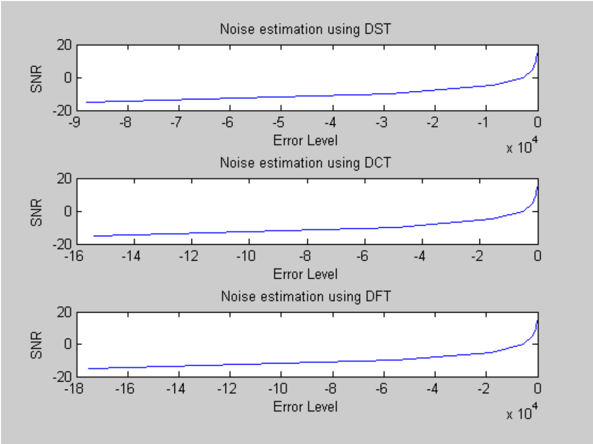

Figure 3. Noise estimation using various transformations

Figure4 is the error chart between estimated noise level and actual noise level under different signal to noise ratio, it can be seen from the chart that the lower the signal to noise ratio is, the more accurate the noise estimation is.

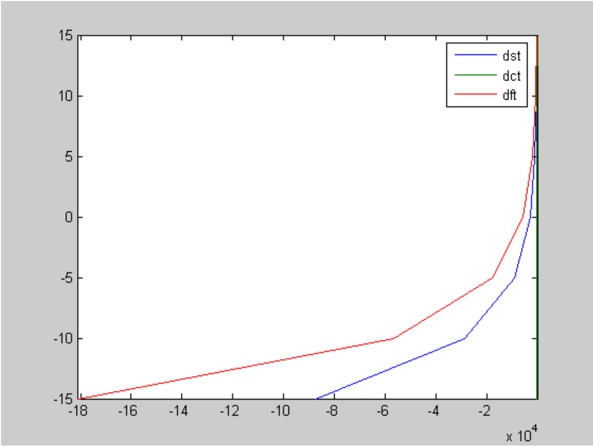

Figure 4. Error chart

4. Experiment

The proposed estimation method was tested on the speech data of the word "Hi", spoken by a male. The segment length of the speech signal was about 10ms. The speech data used was sampled at 11220 Hz.
The speech signal was the processed using different transformations like the DST, DCT and DFT. Inverse transformations were then applied to the transformed signals to reconstruct the original speech. Signal to noise ratios were computed for all the three waveforms. White Gaussian noise of variable SNR values were then added to the original speech for noise estimation.

5. Conclusions

This paper adopts discrete sine transform as the method that generates better results for noise estimation as compared to DCT and DFT. The improvement is noticeable for white noise. In the future research, speech reconstruction using DST can be considered; the further improvement of the effect of de-noising may be seen.

References

[1] Jain AK. , "Fast Karhunen–Loeve transform for a class of stochastic processes", IEEE Transactions on Communications1976; COM-24:1023–1029.
[2] Kekra HB, Solanka JK.,"Comparative performance of various trigonometric unitary transforms for transform image Coding", International Journal of Electronics 1978; 44:305–315.
[3] Proakis JG, Manolakis DG, Digital Signal Processing (3rd edn). Prentice-Hall: Englewood Cliffs, NJ, U.S.A., 1996.
[4] Priyanka Jain, Balbir Kumar and Shail Bala Jain. "Discrete sine transform and its inverse - realization through recursive algorithms", International Journal of Circuit Theory and Applications; Volume 36 Issue4.
[5] Xueyao Li, Hua Xie, and Bailing Cheng, "Noisy Speech Enhancement Based on Discrete Sine Transform", 2006 First International Multi-Symposiums on Computer and Computational Sciences, ISBN: 0-7695-2581-4.
[6] Ing Yann Soon , Soo Ngee Koh , Chai Kiat Yeo, "Noisy speech enhancement using discrete cosine transform", Speech Communication 24 (1998) 249-257.
[5] John G. Proakis, Dimitris G.Manolakis, "Digital signal processing: principles, algorithms and applications", Prentice Hall, March 2002.
[6] Rabiner and Gold, "Theory And Applications Of Digital Signal Processing", Prentice Hall, Inc., 1975