Pseudo Complex Cepstrum Using Discrete Cosine Transform

R. Muralishankar and A. G. Ramakrishnan

Department of Electrical Engineering, Indian Institute of Science, Bangalore-560012, INDIA.

Abstract

Two new algorithms are proposed, which obtain pseudo complex cepstrum using Discrete Cosine Transform (DCT). We call this as the Discrete Cosine Transformed Cepstrum (DCTC). In the first algorithm, we apply the relation between Discrete Fourier Transform (DFT) and DCT. Computing the complex cepstrum using Fourier transform needs the unwrapped phase. The calculation of the unwrapped phase is difficult whenever multiple zeros and poles occur near or on the unit circle. Since DCT is a real function, its phase can only be 0 or π and the phase is unwrapped by representing the negative sign by $\exp(-j\pi)$ and the positive sign by $\exp(j0)$. The second algorithm obviates the need for DFT and obtains DCTC by representing the DCT sequence itself by magnitude and phase components. Phase is unwrapped in the same way as the first algorithm. We have tested DCTC on a simulated system that has multiple poles and zeros near or on the unit circle. The results show that DCTC matches the theoretical complex cepstrum more closely than the DFT based complex cepstrum. We have explored possible uses for DCTC in obtaining the pitch contour of syllables, words and sentences. It is shown that the spectral envelope obtained from the first few coefficients matches reasonably with the envelope of the signal spectrum under consideration, and thus can be used in applications, where faithful reproduction of the spectral envelope is not critical. We also examine the utility of DCTC as feature set for speaker identification. The identification rate with DCTC as feature vector was higher than that with linear prediction-derived cepstral coefficients.

1 Introduction

Homomorphic deconvolution is a method based on the principle of generalized superposition, which has useful applications in a variety of research fields, such as speech analysis and synthesis, marine and earth seismology, medicine (EEG), radar and acoustic system analysis. It was first applied to speech analysis by Oppenheim and Schafer (Oppenheim and Schafer, 1968). Childers et al. (Childers et al., 1977) performed considerable work covering different aspects and applications of homomorphic deconvolution and filtering. Homomorphic deconvolution systems have usually been realized using the Fourier transform (FT), and the utility of the output of deconvolution is very much data dependent.

Unfortunately, the nature of FT dictates the possible uses of the method and restricts its general applicability. It is well known that precise calculation of the unwrapped phase is sometimes difficult (unreliable or even impossible) due to spectral notching or multiple bands with low signal to noise ratios. Calculating the complex cepstrum without phase unwrapping or integration methodology has been introduced in (Bernar and Watt, 1985). However, it is still based on FT and therefore continues to have the limitations of complex cepstrum obtained using FT.

Complex cepstrum (Sokolov, 1989), entirely based on time-domain calculations, avoids or minimizes the problem associated with the FT method. Explicit transformations of an ordinary, mixed phase time sequence into its complex cepstrum time sequence and vice versa are derived in (Sokolov, 1989). This does not require the calculation of the unwrapped phase and no specific windows are used to precondition the signal in order to produce a more accurate representation of the complex cepstrum. In (Hessanein and Rudko, 1984), it is shown that if the original signal is defined to be symmetrical, the Discrete Fourier Transform (DFT) used in cepstral analysis can be replaced by Discrete Cosine Transform (DCT). This principle is applied to evaluate the real and complex pseudocepstra of speech signals. In both the cases, it is found that the use of DCT retains the information contained in the cepstrum, while substantially reducing the computational complexity. However, to the best knowledge of the authors, there has been no new work on the use of DCT for computation of the complex

cepstrum in the past decade.

DCT coefficients have been very successfully deployed as features for character recognition (Dhanya and Ramakrishnan, 2002; Vijay Kumar and Ramakrishnan, 2002). In addition, DCT being a real orthogonal transform, can be advantageously used for resampling or interpolation. In earlier research work, we exploited this to get a good technique for pitch modification (Muralishankar et al., 2003) and period normalization of ECG cycles for a superior compression performance (Ramakrishnan and Saha, 1997). Though DFT and DCT are both sinusoidal transforms, DCT has outperformed DFT in certain applications, such as JPEG (Joint Photographic Experts Group). DCT intrinsically gives higher spectral resolution. We have shown that this results in better performance in applications such as, pitch detection in noisy speech using Spectral Autocorrelation Function (Muralishankar and Ramakrishnan, 2000). Salahuddin et al. (**?**) reported that DCT-based soft thresholding technique, when used for speech enhancement, resulted in better output SNR and speech quality, than wavelet based methods. Similarly, we strongly feel that DCT will have better performance in other speech applications, compared to DFT. Accordingly, we have explored the relative advantages of computing complex cepstrum using DCT, expecting that this might reduce the error arising in the computation due to DFT. In addition, the presence of the binary phase information along with the magnitude might lead to better performance of DCT based complex cepstrum as a feature vector, compared to one that has only magnitude information.

In this work, we introduce two new methods for obtaining pseudo cepstrum (Muralishankar and Ramakrishnan, 2002). In the first method, forcing half-sample symmetric extension for the signal under consideration, we obtain the relation between DFT and DCT. A major difficulty in complex cepstral analysis is the necessity to unwrap the phase in order to make it a continuous function. The relation between DFT and DCT permits the definition of a simpler phase unwrapping algorithm. Since the bases of the cosine transform are real functions, the principal value of their phase can only be 0 or π. Accordingly, we represent the phase as $\exp(-j\pi)$ for negative sign and $\exp(j0)$ for positive sign. With this representation, Log_e operation has been carried out on the sign as well as magnitude of the DCT to obtain log spectrum. After phase unwrapping, the sequence is concatenated with its mirror image to obtain a sequence of

twice the length. This is inverse Fourier transformed. We call this cepstrum, discrete cosine transformed cepstrum (DCTC). In the second method, no assumption is made regarding the signal, and straightaway, the pseudo-cepstrum has been calculated using DCT and IDCT. The phase unwrapping algorithm is the same as in the case of the first method.

2 Mathematical formulation of the DCTC

In a homomorphic system for deconvolution, the characteristic system transforms a time function (signal) from a convolutional space into its image in an additive (cepstral) vector space, i.e., there is a functional time relation connecting these signals (Oppenheim and Schafer, 1989). We now investigate this relation. Consider a sequence $x[n]$ with a rational Z-transform of the form

$$X(z) = \frac{Az^r \prod_{k=1}^{m_i}(1 - a_k z^{-1}) \prod_{k=1}^{m_o}(1 - b_k z)}{\prod_{k=1}^{p_i}(1 - g_k z^{-1}) \prod_{k=1}^{p_o}(1 - d_k z)} \tag{1}$$

The set of values of a_k and g_k are, respectively, the zeros and poles inside the unit circle, b_k and d_k are zeros and poles outside the unit circle, and $|a_k|$, $|b_k|$, $|g_k|$, and $|d_k|$ are all less than 1 (Oppenheim and Schafer, 1968). The complex cepstrum exists if $\ln\{X(z)\}$ is analytic. Only a sequence $x[n]$ with positive mean and a phase derivative with zero mean will have a unique analytic function $\ln\{X(z)\}$ (Quatieri, 1979). Thus, a general, discrete-time, mixed phase sequence $x[n]$ requires time alignment and adjusting the constant A to be positive in order for the complex cepstrum to exist. By definition, the Z-transform $C(z)$ of the cepstrum sequence $c[n]$ is

$$C(z) = \ln\{Y(z)\} \tag{2}$$

where $Y(z)$ is the Z-transform of $y[n]$, obtained by time aligning $x[n]$. After taking the derivative of Eq. 2 with respect to z and taking the inverse Z-transform, we obtain

$$ny[n] = \sum_{k=-\infty}^{\infty} kc[k]y[n-k] \tag{3}$$

This nonlinear difference equation is an implicit relation between y and c, and for minimum or maximum phase sequences, it can be reduced to implicit recurrence expressions (Oppenheim and Schafer, 1968). However, the goal of this paper is to obtain a cepstral representation using the orthogonal transform, DCT.

2.1 DCTC-1

Let $x[n]$ be a real sequence, defined for $0 \leq n \leq M-1$. Consider its half-point, symmetrically extended sequence $y[n]$:

$$y[n] = \begin{cases} x[n] & 0 \leq n \leq M-1 \\ x[2M-1-n] & M \leq n \leq 2M-1 \end{cases}$$

The 2M-length DFT of $y[n]$ can be simplified as (Rao and Yip, 1990),

$$Y[k] = 2\exp(\frac{j\pi k}{2M}) \sum_{n=0}^{M-1} x[n] \cos\frac{(2n+1)\pi k}{2M} \tag{4}$$

Here, the terms inside the summation correspond to

$$X_{DCT}[k] = \sum_{n=0}^{M-1} x[n] \cos\frac{(2n+1)\pi k}{2M} \tag{5}$$

which is the DCT of $x[n]$, except for the scaling factors. Incorporating this in Eq.4,

$$Y[k] = 2\exp(\frac{j\pi k}{2M}) \; X_{DCT}[k] \tag{6}$$

We can write $X_{DCT}[k]$ as

$$X_{DCT}[k] = \exp(\xi[k]) \; |X_{DCT}[k]| \tag{7}$$

where

$$\xi[k] = \frac{j\pi}{2}(sgn(X_{DCT}[k]) - 1)$$

and

$$sgn(p) = \begin{cases} 1, & for \quad p \geq 0 \\ -1, & for \quad p < 0 \end{cases}$$

Therefore,

$$Y[k] = 2\exp(\frac{j\pi k}{2M} + \xi[k]) \;\; |X_{DCT}[k]| \tag{8}$$

$Y[k]$ can be written as $\exp(j\theta[k]) \;\; |Y[k]|$. Taking natural logarithm of the right-hand side of the equation, we have

$$\ln(2) + \ln|X_{DCT}[k]| + \frac{j\pi k}{2M} + \xi[k] \tag{9}$$

Separating the real and imaginary parts of Eq. 9, we have,

$$Re[k] = \ln(2) + \ln|X_{DCT}[k]|$$

$$Im[k] = \frac{\pi}{2}(\frac{k}{M} + (sgn\{X_{DCT}[k]\} - 1))$$

Here, $0 \leq k \leq M-1$. We construct $\zeta[l]$, a sequence of length $2M$, from the above M-length sequence, as shown below:

$$\zeta[l] = \begin{cases} Re[l] + jIm[l] & : \quad 0 \leq l \leq M-1 \\ (Re[2M-1-l] + jIm[2M-1-l])^{\star} & : \quad M \leq l \leq 2M-1 \end{cases}$$

DCTC-1 is computed for the symmetrically extended signal $y[n]$ as follows:

$$\widehat{y}[n] = \{IDFT(\zeta[l])\}$$

where $(0 \leq n \leq 2M-1)$. $\widehat{y}[n]$ is real because of the symmetry of $\zeta[l]$.

2.2 DCTC-2

Instead of using the relation between the DFT and the DCT for symmetrically extended signals, this approach uses DCT and the corresponding IDCT (Martucci, 1994) to obtain the DCTC-2. Consider a finite duration, real sequence $x[n]$, defined for $0 \leq n \leq M-1$ and zero elsewhere. Taking M-point DCT of the above sequence, we have $X_{DCT}[k]$ defined for $0 \leq k \leq M-1$. Using Eq.7 and expanding $\xi[k]$, this can now be written as

$$X_{DCT}[k] = \exp(\frac{j\pi}{2}(sgn\{X_{DCT}[k]\} - 1))\ |X_{DCT}[k]| \tag{10}$$

Taking natural logarithm on both sides,

$$\ln\{X_{DCT}[k]\} = \frac{j\pi}{2}(sgn\{X_{DCT}[k]\} - 1) + \ln|X_{DCT}[k]| \tag{11}$$

Then we obtain the pseudo-complex cepstrum of $x[n]$ as,

$$\widehat{x}[n] = Re\{IDCT[\frac{j\pi}{2}(sgn\{X_{DCT}[k]\} - 1) + \ln|X_{DCT}[k]|]\} \tag{12}$$

3 Linear phase component in complex cepstrum

For the complex cepstrum to exist, it is necessary for the phase function to be continuous and be an odd function on the unit circle. It does not exist (is not defined) if linear phase is present, since $\log z^r$ does not have a Laurent expansion near $z = 0$, and therefore, the phase function is not continuous (Childers et al., 1977). Nevertheless, some authors consider the linear phase component of the complex cepstrum by assuming that the Fourier transform of $\log e^{jr\omega}$ substitutes for the Z-transform of $\log z^r$ on the unit circle. The linear phase component must be removed before computing the complex cepstrum. Otherwise, it introduces rapid decaying oscillations in the complex cepstrum (Childers et al., 1977) since its Fourier transform is

$$\widehat{x}_{linear\ phase}[n] = \begin{cases} 0, & n = 0 \\ \frac{-r}{n}\cos(n\pi) = (-1)^{n+1}\frac{r}{n}, & n \neq 0 \end{cases} \tag{13}$$

Here, r is equal to the number of zeros outside the unit circle. If this number is large, the $\widehat{x}_{linear\ phase}[n]$ term can be large. The oscillations, introduced in the complex cepstrum due to the presence of the linear phase component in the signal, may mask the inherent periodicities of the signal reflected in the complex cepstrum. The presence of a linear phase term may influence the choice of the filter in the cepstral domain, since each point will have contributions from the linear phase component that changes sign from sample to sample. For a finite length sequence, there are no poles so that the denominator of Eq. 1 is unity and its Z-transform is of the form,

$$X(z) = Az^r \prod_{k=1}^{m_i}(1 - a_k z^{-1}) \prod_{k=1}^{m_o}(1 - b_k z) \tag{14}$$

Writing $X(z)$ as a polynomial in z and using the proof given in (Sokolov, 1989), we see that the linear phase factor depends on the number of zeros outside the unit circle, m_o, and that the contribution to the phase due to the constant term A, is an integer multiple of π.

4 Validation of DCTC using theoretical complex cepstrum

We simulate a linear system to show that the DCTC obtained by our method compares well with the theoretical complex cepstrum (TCC) (Oppenheim and Schafer, 1968). In Fig. 1(a), a complex pole pair and a complex zero pair ($|z_{12}| = 0.99$) are placed near the unit circle. It is shown in (Sokolov, 1989) that poles and zeros, proximal to the unit circle, present severe problems to the calculation of the unwrapped phase in the FT based method for cepstrum. The system impulse response $h[n]$, truncated to 65 points, is shown in Fig. 1(b). A complete representation (not shown) would portray a decaying oscillatory function with a duration of several hundred samples. Given the locations of the poles and zeros, the TCC can be calculated:

$$\widehat{h}(n) = \begin{cases} \ln|A|, & n = 0 \\ -\sum_{k=1}^{m_i} \frac{a_k^n}{n} + \sum_{k=1}^{p_i} \frac{g_k^n}{n}, & n > 0 \\ \sum_{k=1}^{m_o} \frac{b_k^n}{n} - \sum_{k=1}^{p_o} \frac{d_k^n}{n}, & n < 0 \end{cases} \tag{15}$$

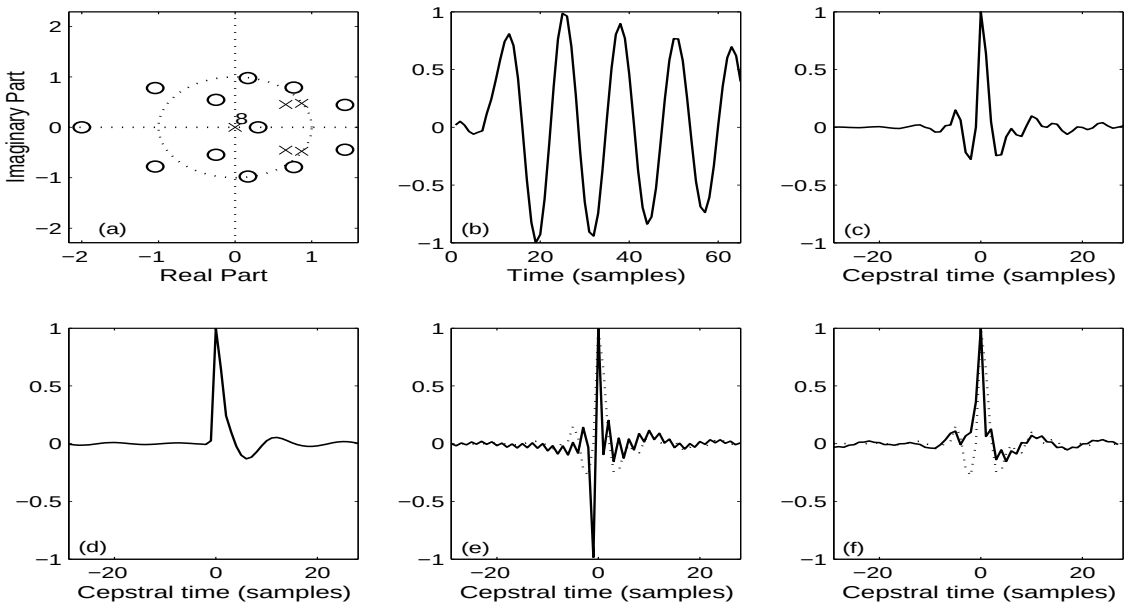

Figure 1: (a) Pole-zero plot of the simulated linear system. (b) System impulse response truncated to 65 points. (c) TCC of the simulated linear system. (d) Complex cepstrum using Fourier Transform. (e) DCTC-1 (solid line) overlapped with TCC. (f) DCTC-2 (solid line) overlapped with TCC.

The coefficients a_k, g_k, b_k, and d_k are defined in Eq.1 and the cepstrum is shown in Fig. 1(c). Removing the linear phase component in the signal by shifting the time sequence $x[n]$ by $m_o = 7$ samples to the left produces the aligned sequence $y[n]$. The DCTC-1 is presented in Fig. 1(e) (solid line) along with the TCC. We emphasize here that no window has been used and only a simple truncation has been performed. Fig.1(d) illustrates the FT cepstrum. Here, phase unwrapping has not been accurately achieved because of the presence of multiple poles and zeros very near the unit circle. The DCTC-2 is displayed in Fig. 1(f), which is a closer approximation to the TCC, than the other two. Thus, it is evident that DCTC matches the TCC better than the FT based complex cepstrum.

5 Applications of DCTC

The DCTC, as proposed, is a concept requiring further work to improve its effectiveness in applications. It is still worthwhile, however, to explore its strengths and weaknesses in its present form. In the subsections that follow, we employ DCTC for spectral estimation, pitch

detection and speaker identification and report our results.

5.1 Estimation of spectral envelope

As a first step, we explore the applicability of DCTC for estimating the envelope of any speech spectrum. Speech spectra are generally quite scalloped, owing to the changes in the pitch of the speaker. The pulse train can be liftered from the cepstrum by a bandpass lifter. After inverse processing, we obtain an estimate of the envelope of the speech spectrum. Figs. 2 and 3 compare the spectral estimates of utterances of a male with fundamental frequency $\approx$100 Hz, obtained using DCTC-1 and DCTC-2, with those obtained employing DFT and linear predictive cepstral coefficients (LPCC). DCTC and LPCC are obtained from a Hamming windowed speech segment of 32.5 msec duration. Spectra are estimated using the first 13 coefficients of DCTC of both voiced and unvoiced speech samples. LPCC spectrum is obtained using 12^{th} order LPC analysis. Figs. 2(a) to 2(d) and 2(e) to 2(f) display the spectra obtained using DCTC-1 for voiced and unvoiced speech samples, respectively. Figs. 3(a) to 3(d) and 3(e) to 3(f) show spectra obtained using DCTC-2 for voiced and unvoiced speech samples, respectively. Figs. 4 and 5 give the corresponding results for a female speaker with pitch frequency $\approx$ 290 Hz.

From the above results, it is observed that the first few DCTC coefficients have the information about the vocal tract response. The figures show that the spectra, estimated by the DCTC, are not as good as those estimated by LPCC, in terms of exactly following the envelope. Thus, if one is interested in the best estimation of spectral envelope, then, DCTC may not be the most appropriate method. Hence, DCTC cannot be used in its current form for applications such as speech coding, where faithful reproduction of the spectral envelope is extremely critical. However, the positions of the peaks are correctly estimated by both the DCTC techniques. Further, the spectra estimated by both the DCTC approaches are smoother than those estimated by LPCC as can be observed from Figs. 2 to 5. This feature is a favorable attribute for application to speaker identification. In order to study this, we have employed DCTC as the feature set in speaker identification experiments and studied whether there is any improvement in performance to that obtained with LPCC.

5.2 Speaker identification

We have employed DCTC as feature set for text-independent speaker identification. Such applications model the speaker individuality by employing features that convey vocal tract information. This makes the complex cepstrum a natural choice as a feature for speaker identification. However, due to the difficulties encountered in the FT-based approach for obtaining the complex cepstrum, most practical systems use either Mel-frequency cepstral coefficients (MFCC) or LPCC. Our experiments with DCTC, MFCC and LPCC demonstrate the viability of DCTC for speaker identification, though future enhancements could further improve its performance.

We have performed speaker identification experiments on a limited subset (50 male speakers) of the TIMIT database. The average duration of each sentence in the database is around 3 sec and the sampling frequency is 16 kHz. Two sentences per speaker are used as the training data and 3 sentences per speaker are used for testing. The speech data is first downsampled to 8 kHz after antialias filtering. Analysis frames of length 31.25 ms are employed, with a frame rate of 100/sec. Each frame is Hamming windowed and processed to obtain DCTC, using the algorithms explained earlier. The first cepstral coefficient is excluded and the next 13 coefficients form the feature vector. To obtain MFCC, a bank of 24 triangular shaped filters is used. The log energy outputs of these filters are transformed to cepstral coefficients using DCT. The feature vector is formed from 13 coefficients, dropping the first coefficient. LPCC is obtained using 12^{th} order LP-analysis. The first 13 LPCC coefficients, excluding the gain term, form the feature vector.

Since the training data is of 2 sec duration, and the frame size is 32.5 ms (frames overlap by 10 ms), we roughly have 200 feature vectors from the training data. Thus, each speaker is modeled using a 30-length vector quantization codebook, consisting of code vectors of DCTC, MFCC or LPCC coefficients, each of dimension 13. The codebooks are trained using k-means clustering algorithm (Duda et al., 2002), employing Euclidean distance measure. Speakers are identified by evaluating the distortion between the features of the test speech sample and the models in the speaker database. The results of the experiments are evaluated as in (Reynolds and Rose, 1995). The test speech of each speaker is processed to produce a sequence of feature

Table 1: Comparison of speaker identification performance of different features. The figures given in the first row represent the relative number of correctly identified frames. The numbers in the second row give the number of speakers identified as a percentage.

Feature	LPCC	DCTC-1	DCTC-2	MFCC
Frame Identification Rate (%)	65.1	67.0	68.9	73.4
Speaker Identification Rate(%)	86	88	90	94

vectors. This sequence is then divided into overlapping segments of 100 feature vectors each, with an overlap of 90 feature vectors between successive segments. Each segment is considered a separate test utterance. If N_T is the total number of segments and N_C is the number of correctly identified segments, then the identification performance is evaluated in percentage as $\frac{N_C}{N_T} \times 100$. Thus, a segment based performance metric, rather than direct speaker recognition performance, has been used.

Here, the length of test data from each speaker is approximately equal and so the performance evaluation is not biased towards any particular speaker. While there may be variations in performance with respect to individual speakers, the evaluation is aimed to track the average performance of the system for different speaker identification tasks, allowing a common basis of comparison (Reynolds and Rose, 1995). Since our evaluation follows the scheme presented in (Reynolds and Rose, 1995), the obtained recognition performance is generally low for all the feature sets employed, compared to the figures available in the recent literature.

The results are given in Table 1. It must be noted that the results are only indicative, since they are based on experiments using data from limited number of speakers. From the results, it can be seen that MFCC gives the best performance among all the features. Both DCTC-1 and DCTC-2 perform comparably with LPCC. This demonstrates the viability of DCTC as an alternative for LPCC for speaker identification. Obviously, MFCC incorporates the perceptual model and thus results in a "weighted" distance measure, which ought to perform better than unweighted distance measures. Thus, by appropriately incorporating the auditory perceptual model into DCTC, we can expect a performance of DCTC comparable to MFCC.

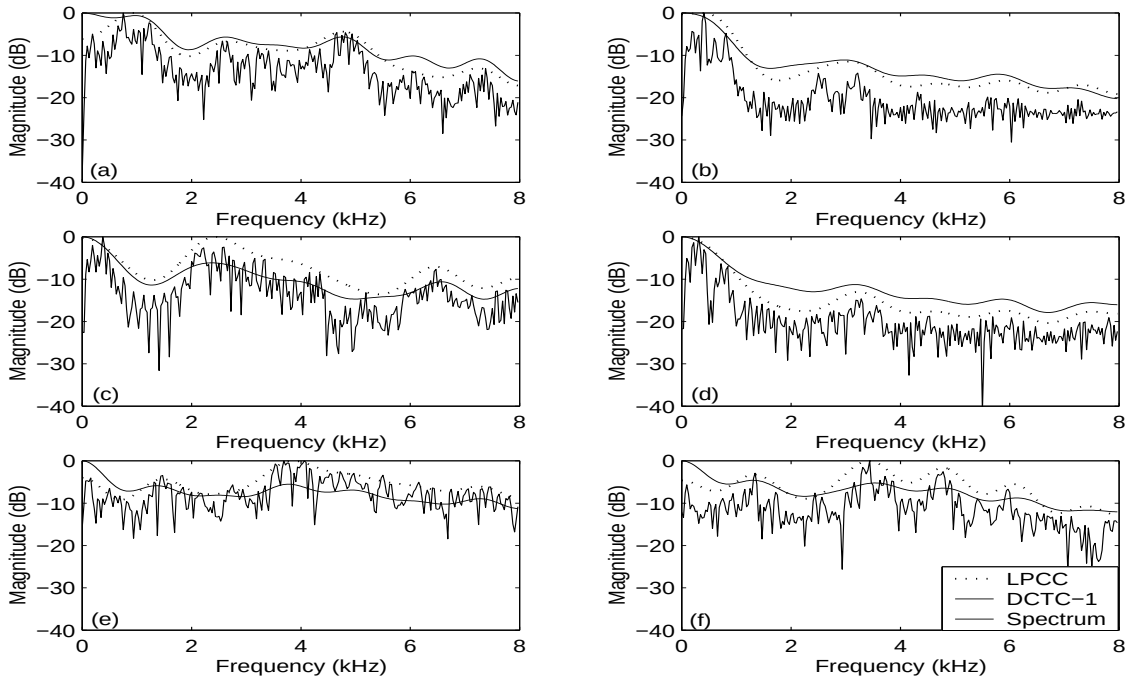

Figure 2: Spectral match for male speech using DCTC-1, (a) to (d) Voiced speech. (e) and (f) Unvoiced.

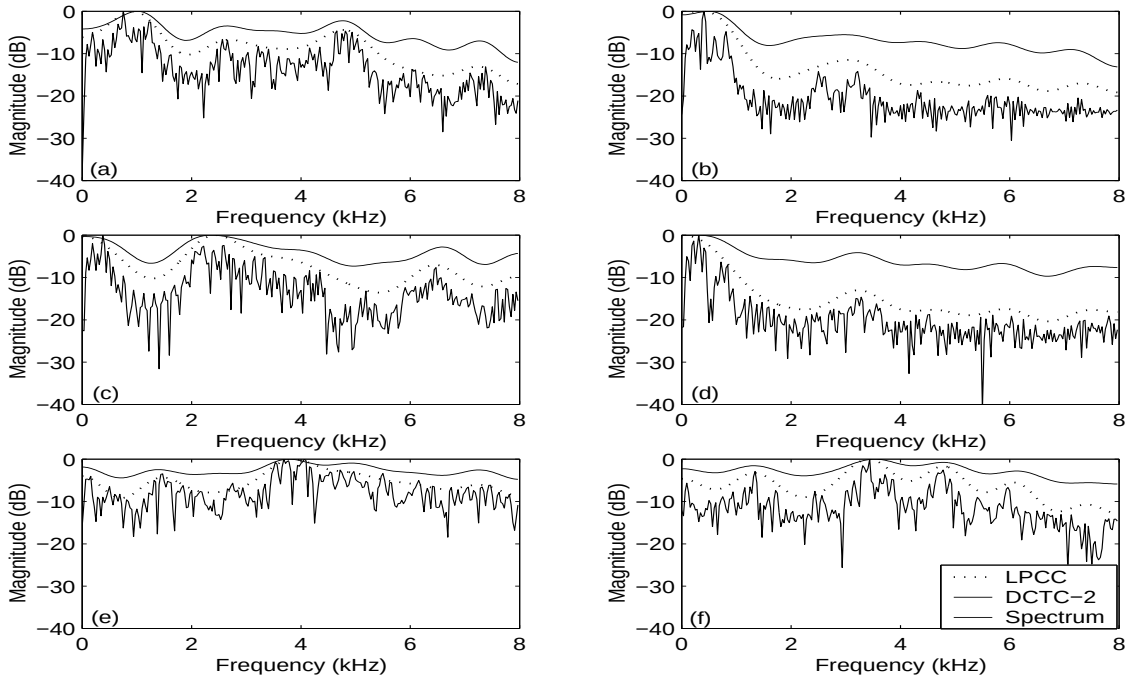

Figure 3: Spectral match for male speech using DCTC-2, (a) to (d) Voiced speech. (e) and (f) Unvoiced.

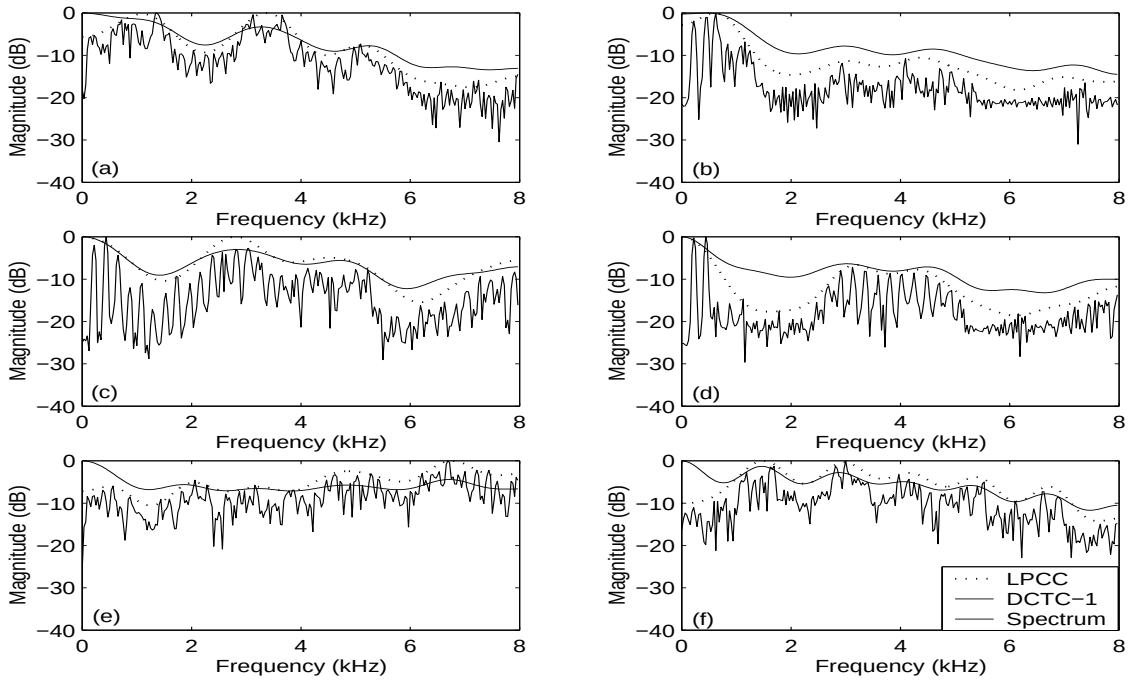

Figure 4: Spectral match for female speech using DCTC-1, (a) to (d) Voiced speech. (e) and (f) Unvoiced.

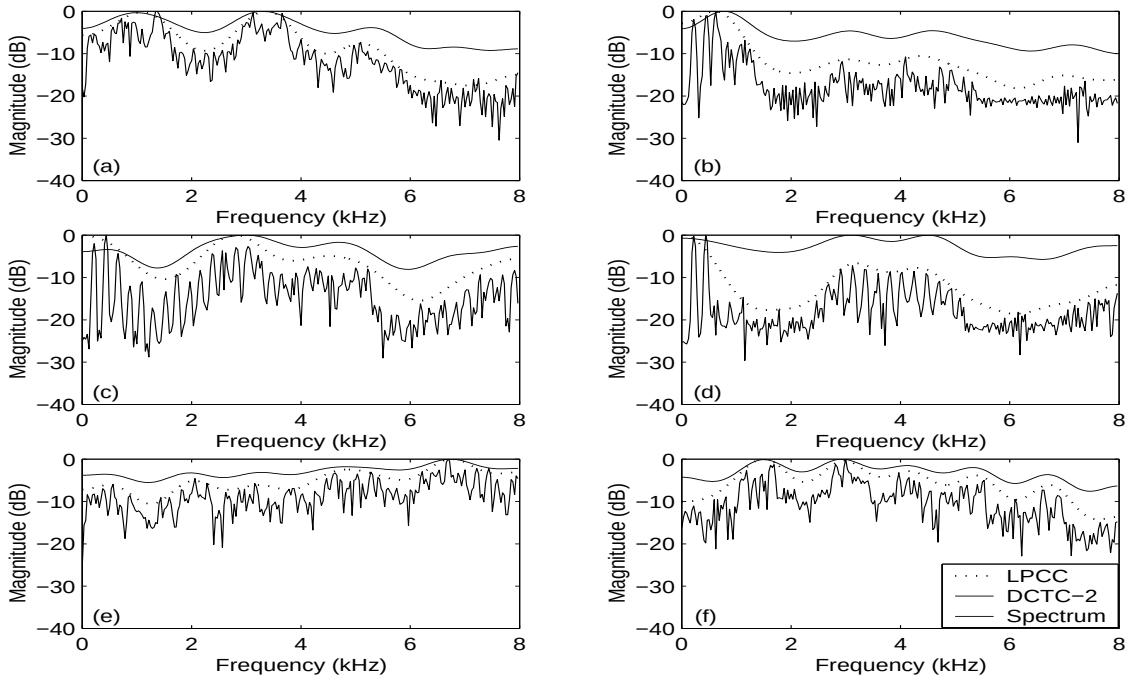

Figure 5: Spectral match for female speech using DCTC-2, (a) to (d) Voiced speech. (e) and (f) Unvoiced.

5.3 Pitch detection

Finally, we investigate the effectiveness of DCTC for pitch extraction. In speech analysis, we usually estimate the parameters of an assumed speech-production model. The most common model views speech as the output of a linear, time varying system (the vocal tract) excited by either quasi periodic pulses or random noise. It is useful to separate or "deconvolve" the system and the excitation components. This is possible in the case of speech, because the convolved signals have very different spectra (O'Shaugnessy, 2000). Cepstral deconvolution transforms a product of two spectra into a sum of two signals. If the resulting summed signals are spectrally different, they can be separated by linear filtering. We know that the formant structure varies slowly in frequency compared to the pitch harmonics or noise. So, the contributions can be linearly separated. The pitch period can be estimated by bandpass liftering the cepstrum and then following the inverse process. It can also be measured directly from the cepstral domain by measuring the time interval from the origin to the first peak.

Fig. 6 shows DCTC-1, DCTC-2 and real cepstral sequences, respectively, of a Kannada utterance, /*niilamegha*/. Only the real cepstrum, rather than the complex cepstrum has been shown, because the intention is to locate the peaks in the cepstral sequences. In order to measure the pitch period from the results of the two DCTC algorithms, the interval from the origin to the first peak is measured. Thus, we obtain the pitch contour of the word. Fig. 7(a) shows the utterance /*niilamegha*/. Its pitch contours obtained using DCTC-1 and DCTC-2 are plotted in Figs. 7(b) and 7(c), respectively, along with the result of autocorrelation method obtained using Praat (Boersma and Weenink, 2003). Here, we have used a frame length of 30 ms and a frame shift of 10 ms to obtain the pitch contours. A pitch range of 50-600 Hz has been assumed here to pick the first peak in the cepstral domain.

6 Conclusion

We have shown the advantages of using DCT for computing the pseudo complex cepstrum. Compared to the FT-based complex cepstrum, phase unwrapping is easy in the case of DCTC and the latter also matches more closely with the theoretical complex cepstrum. Several

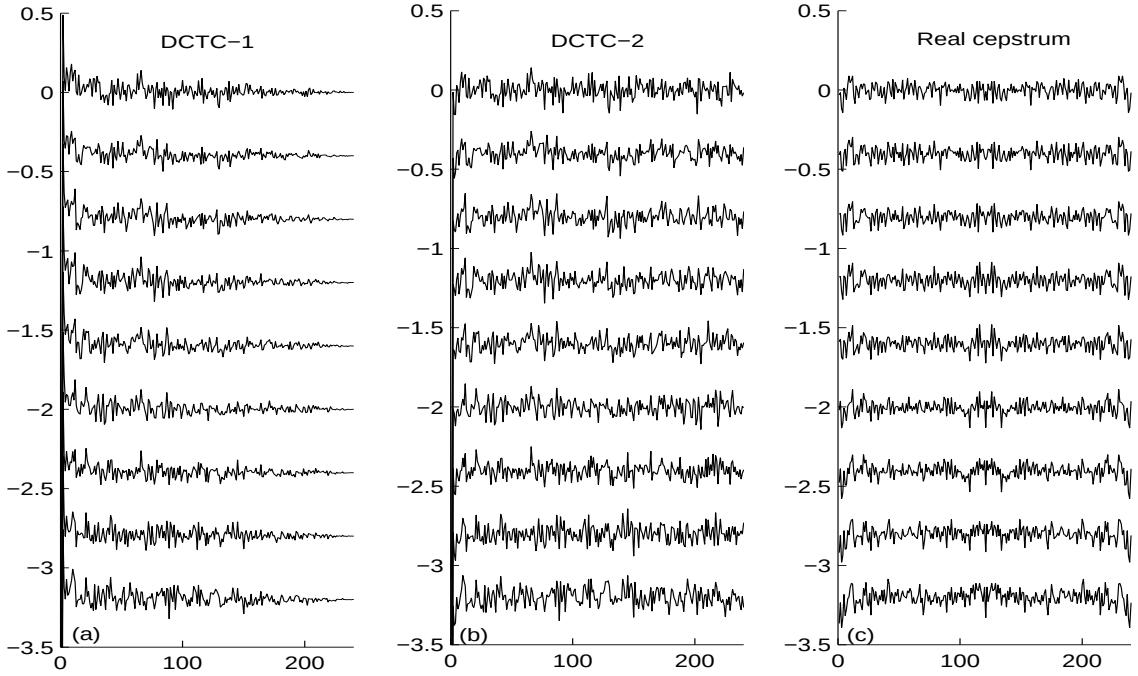

Figure 6: Cepstral sequences of an utterance */niilamegha/* for successive frames. (a) DCTC-1. (b) DCTC-2. (c) Real cepstrum.

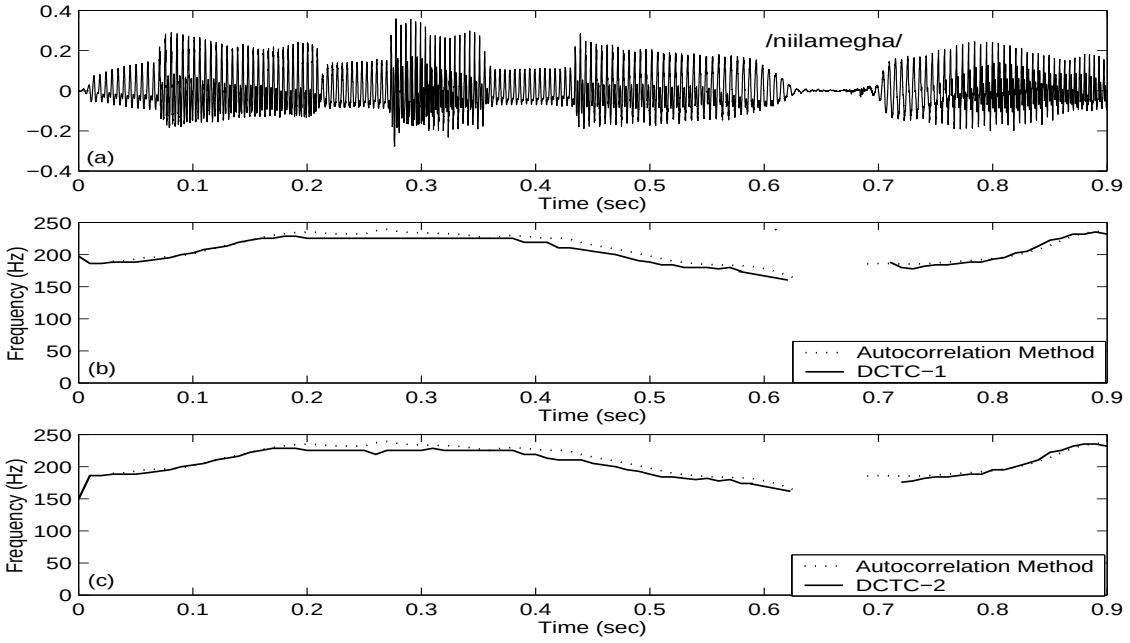

Figure 7: Comparison of pitch detection performance of DCTC with that of autocorrelation based method. (a) An utterance */niilamegha/.* (b) Its pitch contour obtained using DCTC-1 (solid line) and autocorrelation methods (dotted). (c) Its pitch contour obtained using DCTC-2 (solid line) and autocorrelation methods (dotted).

applications of DCTC in speech processing have been examined through pitch detection and speaker identification trials. We have demonstrated the usefulness of DCTC in obtaining the pitch contours. Speaker identification performance using DCTC features is comparable or better than that obtained by LPCC. This demonstrates the viability of DCTC as an alternative for LPCC for speaker identification. DCTC has also been shown to be effective in obtaining the overall spectral envelope. Even though the spectral envelope fitting by DCTC is not as good as that of LPCC, it scores over the latter in higher discriminability, when it comes to speaker identification performance, as shown by the results obtained. The added binary phase information could be the key in deciding this improvement (Provenzale et al., 1992). The results show that DCTC has some promise as an alternative feature for speech, with future enhancements, such as incorporation of auditory perceptual model.

7 Acknowledgement

We thank Dr. S. V. Narasimhan, National Aerospace Laboratories, Bangalore, for his invaluable suggestions. Thanks are also due to Mr. A. Vijayakrishna, IISc, for his help and suggestions in our work. This work has been completed as part of the project, *Algorithms for Kannada Speech Synthesis*, funded by the Ministry of Communication and Information Technology, Government of India. The authors are grateful to the anonymous reviewers for causing considerable improvement in the presentation of this paper.

References

Bernar, J. B. and Watt, T. L. (1985). Calculating the complex cepstrum without phase unwrapping or integration. *IEEE Transactions on Acoustic Speech Signal Processing*, 33:1014–1017.

Boersma, P. and Weenink, D. (2003). *Praat: doing phonetics by computer.* http://www.fon.hum.uva.nl/praat/.

Childers, D. G., Skinner, D. P., and Kemerait, R. C. (1977). The cepstrum: A guide to processing. *Proceedings of the IEEE*, 65:1428–1443.

Dhanya, D. and Ramakrishnan, A. G. (2002). Optimal feature extraction for bilingual OCR. In Document Analysis Systems V, editor, *Daniel Lopresti, Jianying Hu and Ramanujan Kashi*, pages 25–36, Berlin Heidelberg. Springer-Verlag.

Duda, R., Hart, P., and Stork, D. G. (2002). *Pattern classification*. J. Wiley, New York.

Hessanein, H. and Rudko, M. (1984). On the use of discrete cosine transform in cepstral analysis. *IEEE Transactions on Acoustic Speech Signal Processing*, 32:922–925.

Martucci, S. A. (1994). Symmetric convolution and the discrete sine and cosine transforms. *IEEE Transactions on Signal Processing*, 42:1038–1051.

Muralishankar, R. and Ramakrishnan, A. G. (2000). Robust pitch detection using dct based spectral autocorrelation. *Proceedings of International Conference on Multimedia Processing, Chennai*, pages 129–132.

Muralishankar, R. and Ramakrishnan, A. G. (2002). DCT based pseudo complex cepstrum. In *Proceedings of the IEEE, ICASSP*, pages I:521–524.

Muralishankar, R., Ramakrishnan, A. G., and Prathibha, P. (2003). Modification of pitch using DCT in the source domain. *in Press, Speech Communication*.

Oppenheim, A. V. (1969). A speech analysis-synthesis system based on homomorphic filtering. *Journal of the Acoustical Society of America*, 45:458–465.

Oppenheim, A. V. and Schafer, R. W. (1968). Homomorphic analysis of speech. *IEEE Transactions on Audio and Electroacoustics*, AU-16:221–226.

Oppenheim, A. V. and Schafer, R. W. (1989). *Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall.

O'Shaugnessy, D. (2000). *Speech Communications-Human and Machine*. 2nd Ed. Piscataway, NJ: IEEE Press.

Provenzale, A., Smith, L. A., Vio, R., and Murante, G. (1992). Distinguishing between low-dimensional dynamics and randomness in measured time series. *Physica D*, 58:31–49.

Quatieri, T. F. (1979). *Phase estimation with application to speech analysis-synthesis.* PhD thesis, Department of Electrical Engineering, Massachussets Institute of Technology, Cambridge, MA, USA.

Ramakrishnan, A. G. and Saha, S. (1997). ECG coding by wavelet-based linear prediction. *IEEE Transactions on Biomedical Engineering*, 44(12):1253–1261.

Rao, K. R. and Yip, P. (1990). *Discrete Cosine Transform, Algorithms, Advantages, Applications.* Academic Press.

Reynolds, D. A. and Rose, R. C. (1995). Robust text-independent speaker identification using gaussian mixture speaker models. *IEEE Transactions on Speech and Audio Processing*, 3:72–83.

Schafer, R. W. (1968). *Echo removal by discrete generalized linear filtering.* PhD thesis, Department of Electrical Engineering, Massachussets Institute of Technology, Cambridge, MA, USA.

Schroeder, M. R. (1981). Direct (nonrecursive) relations between cepstrum and predictor coefficients. *IEEE Transactions Acoustic Speech Signal Processing*, 29:297–301.

Sokolov, R. T. (1989). *Time-domain cepstral transformations.* PhD thesis, Michigan Technological University.

Vijay Kumar, B. and Ramakrishnan, A. G. (2002). Machine recognition of printed Kannada text. In Document Analysis Systems V, editor, *Daniel Lopresti, Jianying Hu and Ramanujan Kashi*, pages 37–48, Berlin Heidelberg. Springer-Verlag.