DCT BASED PSEUDO COMPLEX CEPSTRUM

R. Muralishankar and A. G. Ramakrishnan

Department of Electrical Engineering, Indian Institute of Science, Bangalore-560012, INDIA.
sripad@ee.iisc.ernet.in, ramkiag@ee.iisc.ernet.in

ABSTRACT

We propose two algorithms, which effectively use Discrete Cosine Transform (DCT) to obtain complex cepstrum, which we call as the Discrete Cosine Transformed Cepstrum (DCTC). The first algorithm uses the relation between Discrete Fourier Transform (DFT) and DCT II. Computing the complex cepstrum using Fourier Transform needs unwrapped phase. The calculation of the unwrapped phase is sometimes difficult due to the presence of multiple zeros and poles near or on the unit circle. Since DCT is a real function, its phase can only be 0 or π and phase unwrapping is done by representing negative by $\exp(-j\pi)$ and positive by 1. The motivation for the second algorithm is to obtain DCTC by representing DCT sequence itself by magnitude and phase components. Phase unwrapping is performed in the same way as in the case of first algorithm. We tested DCTC by simulating the impulse response of a system that has multiple poles and zeros near or on the unit circle and we have shown that DCTC provides a closer match to the theoretical complex cepstrum than the DFT based complex cepstrum. DCTC has been used to obtain the pitch contour of syllables, words and sentences. It is shown that the first few coefficients' spectral envelope matches with the envelope of the signal spectrum under consideration. We are also exploring the applicability of DCTC to speaker recognition.

1. INTRODUCTION

The common speech-production model views speech as the output of a linear, time-varying system excited by either quasi-periodic pulses or random noise. Speech signal is the result of convolving excitation and vocal tract impulse response. It is possible to separate or "deconvolve" the components using linear filter because the convolved signals have very different spectra. Homomorphic deconvolution has proven useful in a variety of fields such as speech analysis and synthesis, marine and earth seismology, biomedical, radar and acoustic system analysis. Its application was introduced by Oppenheim and Schafer [1] to speech analysis, echo detection and removal. Considerable work covering different applications has been performed by Childers et al. [2]. Homomorphic deconvolution is usually realized using Fourier transform (FT), and is very much data dependent. Unfortunately, the nature of FT restricts its general applicability. Precise calculation of the unwrapped phase is sometimes difficult due to spectral notching or multiple bands with low signal to noise ratios. Complex cepstrum [3] 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 [3]. This does not require unwrapped phase calculations and no specific windows are used to precondition the signal in order to produce a more accurate

representation of the complex cepstrum. In [4], it is shown that if the original signal is symmetrical, the Discrete Fourier Transform (DFT) for cepstral analysis can be replaced by Discrete Cosine Transform (DCT). This principle is applied to the evaluation of the real and complex pseudocepstrum of speech signals. In both the cases, it is found that the use of DCT does not degrade the information contained in the cepstrum, while substantially reducing the computational complexity.

In this paper, we propose two new methods of estimating complex cepstrum through DCT, which is termed as Discrete cosine transformed cepstrum (DCTC). In the first method, forcing half sample symmetric extension for the signal under consideration, we obtain the relation between DFT and DCT II. This relation simplifies the phase unwrapping algorithm. Since DCT is a real function, its phase can only be 0 or π. Log_e operation has been carried out on the sign (represented as $\exp(-j\pi)$ for negative and 1 for positive) as well as the magnitude of the DCT. After phase unwrapping, the resulting sequence of length M is concatenated with conjugated and flipped version of the same. Now the modified sequence of length 2M is inverse Fourier transformed to obtain DCTC representation of the signal. In the second method, no assumptions are made regarding the signal, and DCTC has been obtained using DCT-IDCT combination. Phase unwrapping is same as that of the first method.

2. MATHEMATICAL FORMULATION OF DCTC

Consider a sequence with a rational Z-transform of the form

$$X[z] = \frac{A \, z^r \, \prod_{k=1}^{m_i}(1 - a_k z^{-1}) \, \prod_{k=1}^{m_o}(1 - b_k z)}{\prod_{k=1}^{p_i}(1 - g_k z^{-1}) \, \prod_{k=1}^{p_o}(1 - d_k z)} \qquad (1)$$

The factors a_k's and g_k's are the zeros and poles inside the contour of radius γ, respectively, and $|a_k|$, $|b_k|$ and $|g_k|$, $|d_k|$ are all less than γ and b_k's and d_k's are zeros and poles outside the contour γ [1]. The complex cepstrum exists if $\ln(X[z])$ is analytic. Only a sequence $x[n]$ with positive mean and zero mean phase derivative will have a unique analytic function $\ln(X[z])$ [6]. Thus, a general discrete time mixed phase sequence requires time alignment and adjusting the constant A to be positive in order for the complex cepstrum to exist. By definition, the Z-transform of the cepstrum sequence of c is

$$C[z] = \ln(Y[z]) \qquad (2)$$

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

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

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 [1]. However, the goal of this paper is to obtain a DCTC representaion using an orthogonal transform (DCT).

2.1. Algorithm I

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

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

Simplifying DFT of $y[n]$,

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

Here the terms inside the summation correspond to

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

which is same as the DCT of $x[n]$ except for the scaling factors. Denoting it by $X_{DCT}[k]$,

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

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

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

where

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

and

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

Therefore,

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

$Y[k]$ can be written as $\exp(-j\theta[k]) |Y[k]|$. Therefore

$$\exp(-j\theta[k]) |Y[k]| = 2 \exp\left(\frac{j\pi k}{2M} + \xi[k]\right) |X_{DCT}[k]| \qquad (9)$$

Taking logarithm on both sides, we have

$$\ln |Y[k]| - j\theta[k] = \ln(2) + \ln |X_{DCT}[k]| + \frac{j\pi k}{2M} + \xi[k] \qquad (10)$$

Separating real and imaginary parts of Eq. 10, we have,

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

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

Here, $0 \leq k \leq M - 1$. Construct a $2M$ length sequence, $\zeta[k]$ from M length RHS sequence, as shown below:

$$\zeta[k] = \begin{cases} Re[k] - jIm[k] & : \ k < M \\ Re[2M - 1 - k] + jIm[2M - 1 - k] & : \ k \geq M \end{cases}$$

Pseudo complex cepstrum is computed for the symmetrically extended signal $y[n]$ as follows:

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

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

2.2. Algorithm II

Instead of using the relation between DFT and DCT II for the symmetrically extended signal, this approach uses DCT and corresponding IDCT [5] to obtain complex cepstrum of a signal. Consider a real sequence $x[n]$, defined for $0 \leq n \leq N - 1$ and zero else where. Taking N-point DCT of the above sequence, we have $X[k]$ defined for $0 \leq k \leq N - 1$ and can be written as

$$X[k] = \exp\left(\frac{-j\pi}{2}((\text{sgn}\{X[k]\} - 1))\right) |X[k]| \qquad (11)$$

Taking logarithm on both sides

$$\ln\{X[k]\} = \frac{-j\pi}{2}(\text{sgn}\{X[k]\} - 1) + \ln |X[k]| \qquad (12)$$

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

$$\widehat{x}[n] = Re\{IDCT[\frac{-j\pi}{2}(\text{sgn}\{X[k]\} - 1) + \ln |X[k]|]\} \qquad (13)$$

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 [2]. (Nevertheless, some authours consider the linear phase component of the complex cepstrum by assuming that the Fourier transform of $\log \exp(jr\omega)$ substitutes for the Z-transform of $\log z^r$ on the unit circle). The linear phase component must be removed before the complex cepstrum calculation. Otherwise, it introduces rapid decaying oscillations in the complex cepstrum [2] since its Fourier transform is

$$\widehat{x}_{lin\ ph}(n) = \begin{cases} 0, & n = 0 \\ \frac{-r}{n}\cos(n\pi) = (-1)^{n+1}\frac{r}{n}, & n \neq 0 \end{cases} \qquad (14)$$

Here, r is equal to the number of zeros of the Z-transform outside the unit circle. If this number is large, then the $\widehat{x}_{lin\ ph}(n)$ term can be large and may mask echo peaks in the complex cepstrum. The presence of a linear phase term may influence the choice of filter selected in the cepstral domain since each point will have contributions from the linear phase component that may change the 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

$$X(z) = A z^r \prod_{k=1}^{m_i}(1 - a_k z^{-1}) \prod_{k=1}^{m_o}(1 - b_k z) \qquad (15)$$

$X(z)$ is an mth order polynomial in z. We assume that m_i zeros are inside the unit circle and m_o zeros are outside the unit circle. Hence, we obtain ($|\alpha_k| < 1, |\beta_k| > 1$) and on simplifying Eq.15, we get the relation between linear phase and the number of zeros

outside the unit circle, m_o (Proof of this relation is given in [3]). Thus we get

$$X(z) = A\, z^{-m_o} \prod_{k=1}^{m_i}(1 - \alpha_k z^{-1}) \prod_{k=1}^{m_o}(1 - \beta_k z) \qquad (16)$$

where,

$$A = (-1)^{m_o} x(0) \prod_{i=1}^{m_o} \beta_i \qquad (17)$$

The contribution of the constant term A to the phase is an integer multiple of π. The formula for the constant term A is valid for a stable sequence with a rational transfer function as well.

4. VALIDATION OF DCTC USING THEORETICAL COMPLEX CEPSTRUM

We simulate the impulse response of a linear system whose zero /pole locations are known. The purpose of this example is to show that the DCTC obtained by our method compares well with the theoretical complex cepstrum (TCC) [1]. In Fig. 1(a), a complex pole pair is placed near the unit circle ($|p_{12}| = 0.99$) as is a complex zero pair ($|z_{12}| = 0.99$). It is shown in [3] that their proximity to the unit circle presents severe problems to the unwrapped phase calculation used in the FT method. The system impulse response x, 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 pole /zero locations, the TCC can be calculated:

$$\widehat{x}(n) = \ln|A|, \quad n = 0 \qquad (18)$$

$$\widehat{x}(n) = -\sum_{k=1}^{m_i} \frac{a_k^n}{n} + \sum_{k=1}^{p_i} \frac{g_k^n}{n}, \quad n > 0 \qquad (19)$$

$$\widehat{x}(n) = \sum_{k=1}^{m_o} \frac{b_k^n}{n} - \sum_{k=1}^{p_o} \frac{d_k^n}{n}, \quad n < 0 \qquad (20)$$

The coefficients a_k, g_k, b_k, and d_k are defined in Eq. 1 and the TCC is shown in Fig. 1(c). Removing linear phase component in the signal by shifting the sequence x by $m_o = 7$ samples to the left produces the aligned sequence y. The FT cepstrum is shown in Fig. 1(d). Here, because of the presence of multiple poles and zeros on or very near the unit circle, the resultant phase unwrapping is not accurate. The DCTC obtained using the first algorithm is presented in Fig. 1(e) (solid line) along with the TCC. The DCTC obtained using the second algorithm is presented in Fig. 1(f), which is found to be a close approximation of the TCC. We emphasize here that no specific window was used, only simple truncation. In our algorithms, DCT has been used for getting DCTC and it matches closely with the TCC compared to FT based complex cepstrum.

5. APPLICATIONS OF DCTC

Simple speech model considers voiced sounds to be produced by quasi-periodic pulses of air which in turn cause the vocal card to vibrate producing glottal pulses that excite the vocal tract to finally produce speech. For nonnasal sounds the vocal tract is modeled as an all pole filter over short time intervals. The glottal source is modeled with zeros in the z-domain again over short time intervals. It is easy to achieve the deconvolution of the pulse (impulse) train with the composite convolution of the glottal impulse

response and vocal tract response since these two time sequences occupy different frequency ranges in the cepstral domain. Based on this deconvolution, we can estimate the pitch period. This is accomplished by longpass liftering the cepstrum and then following the inverse process. Pitch period can also be measured directly from the DCTC by measuring the time interval from the origin to the first peak, as shown in Fig. 2(a). To test the effectiveness in pitch detection, above algorithms are applied to syllables, words and sentences to obtain the respective pitch contours. The signals are analyzed with a frame length of 30 ms with an overlap of 20 ms. Fig. 2(c) and (d) show an utterance /niilamegha/ and its pitch contour obtained using our method. Another application is estimation of the envelope of the speech spectrum. The speech spectrum is generally quite scalloped due to the speaker's pitch. The pulse train can be liftered from the cepstrum by a shortpass lifter. After inverse processing it, we obtain an estimate of the envelope of the speech spectrum. An example is shown in Fig. 2(b).

As another application, we used DCTC as a feature for speaker identification. We used a 50-speaker subset of the TIMIT database. The average duration of each sentence in the database is around 3 sec. The training data consisted of 2 sentences per speaker. Testing was carried out on 4 sentences per speaker. Speech was first downsampled to 8 kHz. The duration of each frame was 30 ms, with a 20 ms overlap between successive frames. Each frame was hamming windowed and procesed using our algorithms. The feature vectors consisted of the first 13 DCTC coefficients. Each speaker was modeled using a 30-length vector quantization codebook. The results were on par with Mel-frequency cepstral coefficients. 44 of the 50 speakers were correctly identified. When Fisher's discriminant analysis was carried out on the features, all the 50 speakers were correctly identified.

6. CONCLUSIONS

This paper presents the application of DCT to the computation of the DCTC. As compared to the case of FT based complex cepstrum, phase unwrapping is easy in the case of DCTC and the latter also matches more closely with TCC. The potential applications of DCTC in speech processing have been explored through pitch detection and speaker recognition.

7. REFERENCES

[1] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1975.

[2] D. G. Childers, D. P. Skinner and R. C. Kemerait, "The cepstrum: A guide to processing," *Proc. IEEE,* vol. 65, no. 10, pp. 1428-1443, Oct. 1977.

[3] Radomir T. Sokolov, "*Time-domain cepstral transformations,*" Ph. D Thesis, Michigan Technological University, 1989.

[4] H. Hessanein and M. Rudko, "On the use of Discrete Cosine Transform in cepstral analysis," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-32, no. 4, pp. 922-925, Aug. 1984.

[5] S. A. Martucci, "Symmetric convolution and the discrete sine and cosine transforms,"*IEEE Transactions on Signal Processing,* Vol. 42, No. 5, pp. 1038-1051, May. 1994.

[6] T. F. Quatieri, "Phase estimation with application on speech analysis-synthesis," Ph.D. dissertation, Dept. of Electrical Eng., M. I. T., Cambridge, MA, Nov. 1979.
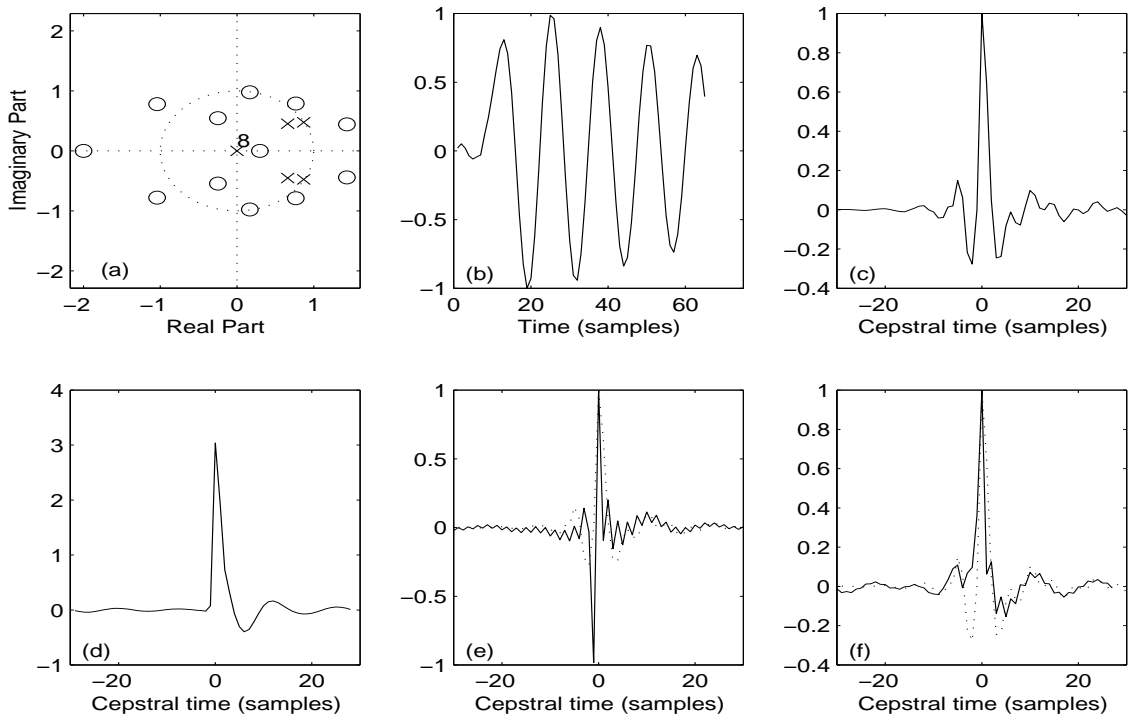
Fig. 1. (a) Pole-zero plot of the simulated linear system. (b) Truncated system impulse response. (c) TCC of the simulated linear system. (d) Complex cepstrum using Fourier Transform. (e) DCTC obtained using first algorithm (solid line) overlapped with TCC of the simulated linear system.(f) DCTC obtained using second algorithm (solid line) overlapped with TCC of the simulated linear system.
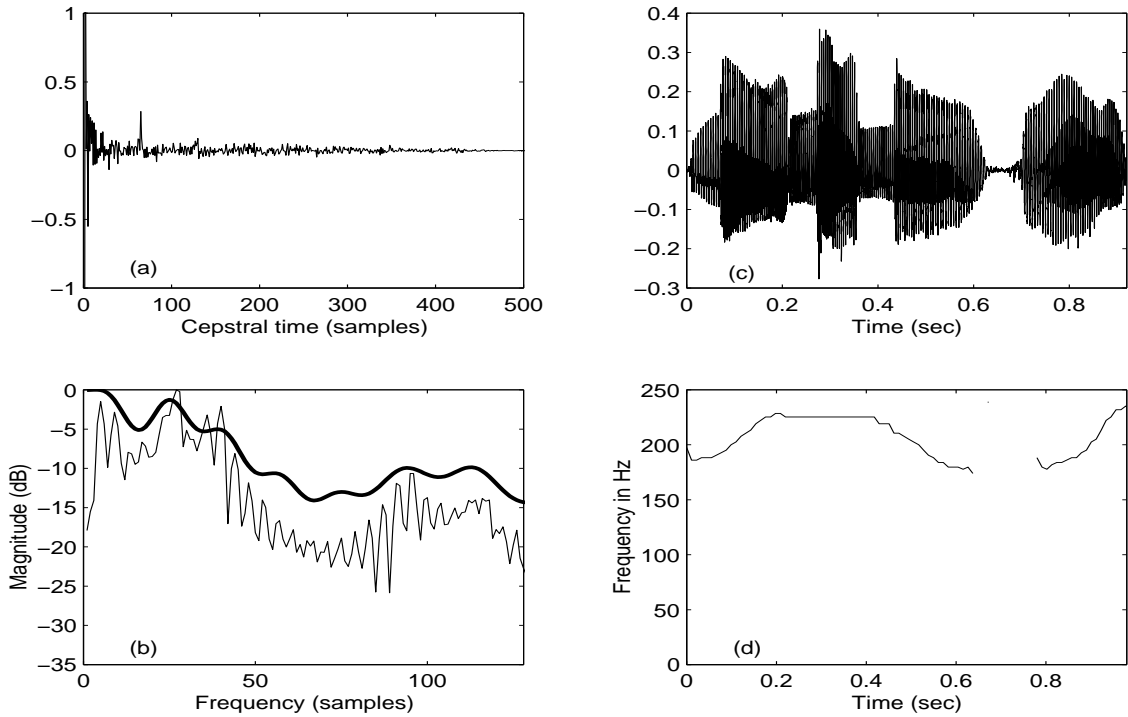

Fig. 2. (a) DCTC for one frame. (b) Its spectrum (considering first 18 coefficients), overlapped with the signal spectrum. (c) An utterance /niilamegha/. (d) Its pitch contour obtained using DCTC.