Theoretical Complex Cepstrum of DCT and Warped DCT Filters

R. Muralishankar Member, IEEE, Abhijeet Sangwan Student Member, IEEE, and Douglas O’Shaughnessy Fellow, IEEE

Abstract—In this letter, we derive the theoretical complex cepstrum (TCC) of the discrete cosine transform (DCT) and warped DCT (WDCT) filters. Using these derivations, we intend to develop an analytic model of the warped discrete cosine transform cepstrum (WDCTC), which was recently introduced as a speech processing feature. In our derivation, we start with the filter bank structure for the DCT, where each basis is represented by a finite impulse response (FIR) filter. The WDCT filter bank is obtained by substituting $z^{-1}$ in the DCT filter bank with a first-order all-pass filter. Using the filter bank structures, we first derive the transfer functions for the DCT and WDCT, and subsequently the TCC for each filter is computed. We analyze the DCT and WDCT filter transfer functions and the TCC by illustrating the corresponding pole-zero maps and cepstral sequences. Moreover, we also use the derived TCC expressions to compute the cepstral sequence for a synthetic vowel /aa/ where the observations on the theoretical cepstrum corroborate well with our practical findings.

I. INTRODUCTION

In an earlier work, we have introduced the warped discrete cosine transform cepstrum (WDCTC) as a new speech processing feature. Particularly, we demonstrated the better performance of the WDCTC over the mel-frequency cepstral coefficients (MFCC) in a vowel recognition and speaker-identification task [1]. Further, we reported some interesting facts about the WDCTC such as (i) good vowel class separability, (ii) low variance, (iii) good codebook representation, (iv) robustness to noise, and (v) better approximation towards a Gaussian distribution [2], [3].

The proposed algorithm for the computation of the WDCTC is briefly reviewed in Alg. 1 [1]. It is noted that the algorithm uses the warped discrete cosine transform (WDCT, [4]) which directly maps the time domain signal to the perceptual frequency domain via an appropriate choice of the warping factor [1], unlike the MFCC which employs the mel frequency filter banks to map the signal representation to a perceptual frequency domain. The use of the WDCT in computing the WDCTC gives a significant advantage in terms of the analyzability of the WDCTC as opposed to the MFCC, where analysis has proved to be a difficult task [5]. In this letter, we exploit this advantage by deriving the theoretical complex cepstrum (TCC) for the WDCTC.

Our approach towards deriving the TCC is as follows. Firstly, we consider the filter bank structure of the discrete cosine transform (DCT) and the WDCT where each basis is represented by a finite impulse response (FIR) filter as shown in Fig. 1 (a) and (b) [4]. Next, we obtain the transfer function for each of the DCT and WDCT filters and show the pole-zero maps along with the frequency responses to demonstrate the effects of warping. Finally, using these DCT and WDCT transfer functions, we derive the theoretical complex cepstrum (TCC) for each DCT and WDCT filter. In addition, we use the newly derived expressions for the WDCT and DCT TCC to compute the cepstral sequences for the synthetic vowel /aa/.

Algorithm 1 Algorithm to compute the WDCTC

1: Obtain an N-point WDCT, $X_{WDCT}(k)$, $0 \leq k \leq N - 1$ for a finite duration, real sequence $x(n)$, $0 \leq n \leq N - 1$.
2: Compute the phase $\zeta(k)$ of the WDCT coefficients as:
$$\zeta(k) = \frac{j\pi}{2}(\text{sgn}(X_{WDCT}(k)) - 1),$$
where $\text{sgn}$ is the sign of the WDCT coefficient.
3: Compute the WDCTC $\hat{x}(n)$ as:
$$\hat{x}(n) = \Re\{\text{IDCT}(\zeta(k) + \ln|X_{WDCT}(k)|)\},$$
where $\Re$ is the real part of the complex number.

II. POLE-ZERO MAP OF DCT AND WDCT FILTERS

Let the N-point DCT, $\{X(0), X(1), ..., X(N - 1)\}$ be given by
$$X(k) = U(k) \sum_{n=0}^{N-1} x(n) \cos \left(\frac{(2n + 1)k\pi}{2N}\right),$$
for $k = 0, 1, ..., N - 1$, where
$$U(k) = \begin{cases} \frac{1}{\sqrt{2}} & k = 0, \\ 1 & \text{otherwise}. \end{cases}$$

The $k^{th}$ row of the $N \times N$ DCT matrix can be viewed as a filter whose transfer function is given by
$$F_k(z^{-1}) = U(k) \sum_{n=0}^{N-1} \cos \left(\frac{(2n + 1)k\pi}{2N}\right) z^{-n},$$
that is, the $i^{th}$ coefficient of $F_k(z^{-1})$ is the $(k,i)^{th}$ element of the DCT matrix. $F_k(z^{-1})$ is a band-pass filter with center frequency at $\frac{1}{2} (\beta e^{\pi i})$, with the sampling frequency normalized to one. Letting $\theta = \frac{\pi}{\beta}$ and using Euler's trigonometric relations, (4) can be written as

$$F_k(z^{-1}) = \frac{U(k)}{2} \sum_{n=0}^{N-1} \left( e^{\frac{\pi i}{\beta} e^{i \theta} + e^{\frac{\pi i}{\beta} e^{-i \theta}}} \right) z^{-n}. \quad (5)$$

Using a closed form for the geometric series in (5) yields

$$F_k(z^{-1}) = \frac{U(k)}{2} \left\{ \frac{\beta e^{\pi i} z^{-N} - 1}{1 - e^{\pi i} z^{-1}} + \frac{1 - e^{\pi i} z^{-N}}{1 - e^{\pi i} z^{-1}} \right\}. \quad (6)$$

We rewrite $e^{i \pi N}$ as $(-1)^k$ in (6) and simplify to obtain the expression for the transfer function as:

$$F_k(z^{-1}) = \frac{U(k)}{2} \left( \frac{\beta e^{\pi i} z^{-N} - 1}{1 - e^{\pi i} z^{-1}} \right) z^{-1}. \quad (7)$$

To obtain the transfer function for the $k^{th}$ WDCT filter ($\tilde{F}_k(z^{-1})$), we substitute $z^{-1}$ with the first-order all-pass filter given by

$$A(z) = \frac{-\beta + z^{-1}}{1 - z^{-1}} \quad (8)$$

where $\beta$ is the warping parameter. The value of $\beta$ controls the nature and degree of warping. Using (8) in (7), we obtain

$$F_k(z^{-1}) = \frac{U(k)}{2} \left( \frac{\beta e^{\pi i} M_l(k) / N}{1 - e^{\pi i} A(z)} \right) \quad (9)$$

The $N$ roots of the term $1 - \beta e^{\pi i} M_l(k) / N$ in (7) are given by

$$z = e^{i \pi M_l(k) / N}, \quad l = 0, 1, ..., N - 1,$$

where

$$M_l(k) = \begin{cases} 2l + 1 & k \text{ is odd}, \\ 2l & k \text{ is even}. \end{cases}$$

Now, (7) can be rewritten as

$$F_k(z^{-1}) = \frac{U(k)}{2} \left( \frac{(1 - z^{-1}) \prod_{l=0}^{N-1} (1 - e^{i \pi M_l(k) / N})}{1 - e^{i \pi A(z)}} \right). \quad (10)$$

Similarly, (9) becomes

$$\tilde{F}_k(z^{-1}) = \frac{U(k)}{2} \left( \frac{(1 - A(z)) \prod_{l=0}^{N-1} (1 - e^{i \pi M_l(k) / N})}{1 - e^{i \pi A(z)}} \right). \quad (11)$$

Using the definition of the all-pass filter in (8), we simplify (11) to obtain

$$\tilde{F}_k(z^{-1}) = \frac{U(k)(1 + \beta e^{i \pi}) (1 - z^{-1})}{(1 + \beta e^{i \pi})(1 - e^{i \pi} A(z))} \prod_{l=0}^{N-1} (1 - \beta e^{i \pi M_l(k) / N}) \prod_{l=0}^{N-1} (1 - \beta e^{i \pi A(z)}) \frac{(1 - e^{i \pi M_l(k) / N})}{(1 - e^{i \pi} A(z))}. \quad (12)$$

The transfer functions of the DCT and WDCT filters are given by (10) and (12), respectively.

III. COMPLEX CEPSTRUM OF DCT AND WDCT FILTERS

To derive the TCC we apply the logarithm to (10), i.e.,

$$\ln(F_k(z^{-1})) = \ln \left( \frac{U(k)}{2} \right) + \ln(1 - z^{-1}) + \sum_{l=0}^{N-1} \ln(1 - e^{i \pi M_l(k) / N} / (1 - e^{i \pi} z^{-1}) - \ln \left( \frac{1 - e^{i \pi} z^{-1}}{1 - e^{i \pi} z^{-1} - \beta e^{i \pi}} \right). \quad (13)$$

The power series expansion for the term $\ln(1 - \alpha z^{-1})$ is [6],

$$\ln(1 - \alpha z^{-1}) = -\sum_{n=1}^{\infty} \frac{\alpha^n}{n} z^{-n}, \quad |z| > |\alpha|,$$
where the region of convergence includes the unit circle. Using the above-mentioned power series expansion in (13), we get

\[
\ln(F_k(z^{-1})) = \ln\left(U(k) \cos\left(\frac{\theta}{2}\right)\right) - \sum_{n=1}^{\infty} \frac{z^{-n}}{n} - \sum_{l=0}^{N-1} \frac{e^{i\theta n}}{n} z^{-n} + \sum_{n=1}^{\infty} \frac{e^{-i\theta n}}{n} z^{-n}.
\]

Taking the inverse z-transform, we obtain the TCC for the \(k^{th}\) filter as

\[
f_k(n) = \begin{cases} 
\ln\left(U(k) \cos\left(\frac{\theta}{2}\right)\right) + \sum_{l=0}^{N-1} \frac{e^{i\theta n}}{n} & n = 0, \\
\frac{1}{2} \left(1 + \sum_{l=0}^{N-1} e^{i\theta n} \right) - 2 \cos(\theta n) & n > 0.
\end{cases}
\]

Similarly, we obtain the TCC of the \(k^{th}\) WDCT filter as

\[
\tilde{f}_k(n) = \begin{cases} 
\ln\left(U(k) (1 + \beta) \cos\left(\frac{\theta}{2}\right)\right) + \sum_{l=0}^{N-1} \ln(1 + e^{i\theta n} / N) & n = 0, \\
\frac{1}{\beta} \left(1 + \sum_{l=0}^{N-1} \frac{(1 + e^{i\theta n} / N)^n}{(1 + e^{i\theta n} / N)^n} \right) & n > 0.
\end{cases}
\]

IV. TCC OF A SYNTHETIC VOWEL

Let the input speech signal be represented as \(X(z)\) in the \(z\)-domain. Using the source-filter model, we get

\[
X(z) = V(z)E(z),
\]

where \(V(z)\) and \(E(z)\) are the \(z\)-domain representations of the vocal-tract response and excitation signal, respectively. Then the output of the \(k^{th}\) DCT and WDCT filter bank is given by

\[
Y_k(z) = X(z)F_k(z)
\]

and

\[
\tilde{Y}_k(z) = X(z)\tilde{F}_k(z),
\]

respectively, where \(k = 0, 1, \ldots, N - 1\). Applying logarithm and inverse z-transform, (18) and (19) become

\[
\hat{y}_k(n) = \hat{x}(n) + \hat{f}_k(n)
\]

and

\[
\tilde{\hat{y}}_k(n) = \tilde{\hat{x}}(n) + \tilde{\hat{f}}_k(n),
\]

respectively, where \(n = 0, 1, \ldots, N - 1\). Here, \(\hat{y}_k(n)\) and \(\tilde{\hat{y}}_k(n)\) are output cepstral sequences of the DCT and the WDCT. The decimator present in each branch of the filter bank acts as a multiplexer which selects consecutive samples from each of the filter output sequences. Hence, the TCC can be obtained from DCT and WDCT cepstral sequences as

\[
\hat{c}_k = \hat{y}_k(n) \delta(n - k)
\]

and

\[
\tilde{\hat{c}}_k = \tilde{\hat{y}}_k(n) \delta(n - k),
\]

respectively, where \(k, n = 0, 1, \ldots, N - 1\).

In this letter, we use the above technique to obtain the TCC for a synthetic vowel. An intelligible vowel sound can be synthesized by appropriately choosing the first three formant frequencies \(F_1, F_2\), and \(F_3\). Also, one could use cascaded digital resonators to produce vowel sounds by choosing resonant frequencies equal to formants. Hence, the problem of deriving the TCC for a vowel sound reduces to determining the TCC for a digital resonator. In order to derive the TCC for a digital resonator, we define the phase and magnitude of the poles of the digital resonator to be \(\pm \omega_0\) and \(r_0\), respectively, and obtain the corresponding transfer function as

\[
H_{\omega_0,\omega_0}(z) = \frac{1}{(1 - r_0 e^{j \omega_0 z^{-1}})(1 - r_0 e^{-j \omega_0 z^{-1}})}.
\]

Taking logarithm and inverse z-transform of (24), we get the complex cepstrum of the digital resonator as

\[
\tilde{h}_{\omega_0,\omega_0}[n] = \begin{cases} 
\frac{2 r_0^2}{n} \cos(\omega_0 n) & n > 0, \\
0 & n = 0.
\end{cases}
\]

Now, the transfer function for the vocal-tract response of a vowel sound is easily obtained as a product of the three digital resonator transfer functions corresponding to each formant,

\[
V(z) = H_{\omega_1,\omega_1}(z)H_{\omega_2,\omega_2}(z)H_{\omega_3,\omega_3}(z).
\]

Hence, the complex cepstrum of the vowel is now merely the sum of the individual complex cepstrum sequences corresponding to the formants and the excitation, i.e.,

\[
\tilde{x}[n] = \frac{2}{\pi} \left\{ \sum_{i=1}^{N_h} \tilde{h}_{\omega_i,\omega_i}[n] + \tilde{e}[n] \right\},
\]

where \(\tilde{e}[n]\) is the complex cepstrum of the excitation, and \(N_h\) is the number of harmonics available within \(\frac{\omega_0}{\Delta \phi}\).

V. RESULTS AND DISCUSSION

The pole-zero maps for the DCT and WDCT filters are shown in Fig. 2 for the \(4^{th}\) filter of 16-length transform. It is observed in Fig. 2 that the peaks in the frequency response correspond to a pole-zero cancellation. The warping effect is illustrated in Figs. 2(c,d), (e) and (f) for two values of the warping parameter \((\beta = 0.56\) and \(-0.56\)). Note that the warped filters have \(N = 1\) real poles, which boosts
the magnitude response in the region of higher resolution. Figure 3 shows the cepstral sequence for the DCT and WDCT filters using two values of the warping parameter ($\beta = 0.56$ and $-0.56$) where the effect of warping is evident. It may be useful to note that Smith and Abel have already shown that the warping for $\beta = 0.56$ closely resembles the psychoacoustic Bark scale for a sampling frequency of 16kHz [7]. Hence, $\beta = 0.56$ is employed in computing the WDCTC speech feature, and is also used in analyzing the synthetic vowel /aa/ in what follows.

We used synthetic vowel /aa/ with the formant frequencies given by 700Hz, 1220Hz and 2600Hz. Its fundamental frequency was 100Hz. The complex cepstral sequence of vowel /aa/ with DCT and WDCT is obtained using Eqs. (22) and (23), respectively and is shown in Fig. 4. Figures 4(a) and (c) show cepstral sequences obtained using DCT and WDCT. The sequences are decaying with a factor equal to $\frac{1}{56}$. Further, pitch peaks can be seen in Figs. 4(a) and (c) and are exactly equal to 100Hz (160th sample in the quefrency domain). We can also observe clear separation between vocal-tract response and excitation in the quefrency domain for the WDCT. Figures 4(b) and (d) show the spectrum obtained from the first 18 cepstral coefficients of (a) and (c). Further, from Figs. 4(b) and (d), we can see that the dynamic range of the WDCT spectrum is less than the DCT. It is useful to note that a similar observation about the low variance of the WDCTC was made in practical experiments [2]. It is also observed that the higher resolution in low frequency provides a closer match to the peaks in the spectrum and given formant positions for the WDCT than the DCT.

VI. CONCLUSION

We have presented the TCC and the transfer functions for the DCT and WDCT filters, and illustrated the corresponding pole-zero map, frequency responses and TCC sequences. Our analysis of the DCT

and WDCT filters has thrown light on some interesting properties such as the distribution of zeros on the unit circle and the pole-zero cancellation, which illustrate the spectral properties of the DCT and WDCT bases. Further, we have also shown the TCC for a synthetic vowel which reflects the property of low variance observed in an earlier work. The results presented in this letter are significant as they provide an opportunity to theoretically analyze the WDCTC which has already shown improved performance over the MFCC in an earlier work.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their helpful and constructive comments during the review process.

REFERENCES