

Frequency warping control parameter determination biology essay

[Science](#), [Biology](#)



VijayaBhaskar Reddy Tupili, ECE-612, UIN: 00957451

Abstract---In this paper, the primary goal is to design a MATLAB based simulator for processing or to enhance the speech which is degraded under noisy environments by using a technique called warped discrete cosine transform(WDCT) is proposed. To achieve these results, sample speeches were obtained. These were modeled as an autoregressive (AR) process and represented in the frequency domain by the WDCT filter. A warped frequency warping filter along with the conventional DCT is applied to the input frequency to get the effective characteristics of the input frequency. The warping parameter which is used to control the frequency is adjusted according to the spectral distribution in each frame. To obtain the frequency warping control parameter, a split band approach is employed in which the global soft decision for speech is performed in each band separately. The original speech signal and the reconstructed speech signal obtained from the output of the filter were compared. The idea of this comparison is to pursue an output speech signal, which is similar to the original one. It was concluded that WDCT is a good constructing method for speech.

Index Terms " Discrete cosine transform (DCT), frequency warping control parameter, Laguerre filter, speech enhancement, split-band global soft decision, warped DCT (WDCT).

I. INTRODUCTIONSpeech processing has been a growing and dynamic field for more than two decades and there is every indication that this growth will continue and even accelerate. the interest in the noise speech enhancement

is increased day by day since the performance of the system seriously degrades with the presence of noise in it. In order to achieve this speech enhancement many approaches have been investigated, some of those are, Wiener filtering, spectral subtraction, soft decision estimation, and minimum mean-square error(MMSE)[1]-[2]. To eliminate noise from noisySpeech. The above methods use Discrete Fourier Transform (DFT) in frequency domain. However, DCT is proved to be more better at enhancing noise signals than DFT because of its high energy compaction capability than DFT which can be used to provide higher resolution without increasing the length of the DCT. To adjust the frequency distribution a method is devised to warp the input frequency. This paper is organized as follows. A WDCT definition and implementation is given in section-II. Speech enhancement algorithm and determination of frequency warping parameter is given in section- III. Experimental result and discussions are given in section- IV and finally conclusion is given in the section-V. II. Warped Discrete Cosine Transform: Here, we review an N-point WDCT of the input vector $[x(0); x(1); \dots; x(N-1)]'$. The N-point DCT, $\{X(0), X(1), \dots, X(N-1)\}$ is defined by,

,

Where, In the N-point DCT matrix the kth row can be viewed as a filter and its transfer function is given as, The (k, n)th element of the DCT matrix is the nth coefficient of the $F_k(Z)$. If the sampling frequency is normalized to 1 then $F_k(Z)$ is a band pass filter with a center frequency equal to $(2k+1)/2M$. For low-frequency inputs such as voiced sounds, the output magnitude of $F_k(Z)$ for small k is generally larger and it enables data compression by giving

more emphasis to the lower band outputs than the higher band ones. For higher values of k the magnitude of the output of the $F_k(Z)$ is large if the input consists of high-frequency components which is desired feature for noise removal. Discrimination of speech is affected by few factors.

Specifically, frequency selectivity is one of the important aspect in the discrimination of speech. Frequency selectivity is defined as the important frequency regions with high spectral magnitude areas that listeners want to hear with intensive care. Higher resolution needs to be provided in selected areas like high spectral magnitude region because of this reason. To increase the frequency resolution one of the possible way is to increase the total size of the DCT which depends on the distance between the two consecutive sampling points in the frequency domain. speech quality will be improved by increasing the frequency resolution. But the drawback with this method is it increases the complexity in embedded systems. The listener's fatigue occurs from a large amount of simulation in the noise-only frequency regions especially for higher frequency resolution and for low spectral magnitude.

For the above reasons, an appropriate method based on the input frequency warping is used without increasing the size of DCT and is used to adjust the input speech spectral distribution. An all pass transform with a stable all pass filter $A(z)$ is proposed to replace Z^{-1} to warp the frequency axis and is given as, $A(z) = \frac{z^{-1} + \alpha}{1 + \alpha z^{-1}}$ Where, α is known as the Laguerre filter which is the control parameter for the warping of frequency response and is widely used in algorithms of various signal processing. The resulting transfer function $F_k(A(Z))$ is now as infinite impulse response (IIR) filter is defined as, $F_k(A(z)) = \cos(A(z))n$ Implementation of WDCT: The filter bank method is considered

for the implementation. The filtering and decimation result for the finite impulse response (FIR) filter with M -tap finite impulse gives the inner product of the input vector and the filter coefficient vector. This is again equal to the inner product of the filter coefficients of the DFT and the conjugate DFT of the input that have $F_k(e^{j\omega})$ which consists of the sampled values for $\omega = 0, (2\pi/M), (4\pi/M), \dots, ((M-1)k\pi/M)$. The inner products of the inverse discrete Fourier transform (IDFT) and the input vector of the sampled sequence $F_k(Ae^{j\omega})$ approximates the result of filter along with $F_k(A(e^{j\omega}))$. Figure 1: Frequency responses (a) 4-point DCT filter Bank (b) WDCT filter bank with $\alpha = 0.25$ (c) WDCT filter bank with $\alpha = -0.25$. For different values of α , the warped filter banks frequency responses are shown in fig. 1. In that figure, for positive α low band is more emphasized. In the high band for negative α , the modeling and spectral characteristics are more appropriate. We apply positive α for the speech signal especially with low frequency components for the above reason. Similarly, for the signals with high frequency components we suggest negative α . Frequency-Warped Speech

Enhancement: For a given speech signal x it is assumed that a noise signal n is added, and their sum is denoted by y , considering the M -point DCT gives us, $Y_k(t) = X_k(t) + N_k(t)$, $K = 0, 1, \dots, M-1$. Where ' k ' k th frequency bin, M is the total number of frequency components and ' t ' is the frame index in the time domain respectively. If a noisy speech signal frame is given then the basic assumption used for the speech enhancement approach is given as follows: H_0 : speech absent: $Y(t) = N(t)$ H_1 : speech present : $N(t) + X(t)$ in which $Y(t) = [Y_0(t), Y_1(t), \dots, Y_{M-1}(t)]^T$, $N(t) = [N_0(t), N_1(t), \dots, N_{M-1}(t)]^T$, and $X(t) = [X_0(t), X_1(t), \dots, X_{M-1}(t)]^T$ are the DCT coefficients of noisy

speech, noise and original clean speech respectively. This speech enhancement technique is mainly used to approximate $\{X_k(t), K= 0, 1, \dots, M-1\}$ and $\{Y_k(t), K= 0, 1, \dots, M-1\}$. The MMSE estimator for X_k is given from Gaussian expression as follows Where, in which λ_s, k and λ_n, k are the variances of clean speech and voice respectively. In the performance of speech enhancement the strong estimation of $\lambda_s, k, \lambda_n, k$ and $???? k$ play an vital role. If the DCT is employed then the multiplication in a transform domain specially corresponds to the time domain filtering. In the case of WDCT, the linear convolution can be carried similarly. When the real Gaussian assumption is adopted then we can note that the MMSE estimator reduces to the Wiener filter. The “ musical tone” is the major drawback for the Wiener filter with spectral-domain-based speech enhancement. This is a random frequency which occurs due to the under estimation of noise power. Frequency-warped domain also consists of similar properties. The soft-decision-based speech enhancement algorithm is used to overcome this drawback. The basic frame work in speech enhancement is adopted for the extra configuration. Split-Band Global Soft Decision: For each split frequency band a statistical model is assumed for the determination of frequency-warping control parameter α . For the determination of α , the whole frequency range is divided into high-band and low-band regions. In the M DCT coefficients, the first ‘ m’ coefficients are assigned to low-band and the remaining (M-m) are assigned to high-band region. Experimentally chosen value $m = (2/3)M$ is used here. The probability density functions (pdfs) in the high-band region for the noisy speech conditioned on H_0 and H_1 are given as, From the above statistical assumptions the likelihood ratio is given as

follows, The high-band global speech presence probability (HB-GSPP) can be derived by applying Baye's rule to the above ratio as follows: Where $Y_H(t) = [Y_m(t), Y_{m+1}(t), \dots, Y_{M-1}(t)]$. We assume that the spectral component in each frequency bin is statistically independent and can be converted as follows: Where, $q_H = P_H(H1)/P_H(H0)$. The low-band global speech presence probability (LB-GSPP) can be calculated in the same way as HB-GSPP such that Where, $Y_L(t) = [Y_0(t), Y_1(t), \dots, Y_{m-1}(t)]$ and $q_L = P_L(H1)/P_L(H0)$.

Frequency-Warping Control Parameter Determination: For the given image-compression algorithm to minimize the reconstruction error we choose frequency-warping control parameter. The optical warping parameter should be determined before the recognition stage in the speaker normalization by compensating the various vocal track lengths in order to improve the speech recognition accuracy. We have to choose efficient warping parameter necessarily in an online fashion as we cannot apply these architectures directly to speech enhancement. The frequency-warping control parameter α should be considered based only on the input speech in each frame. The application of voiced or unvoiced (V/UV) decision and selecting α depending on the decision is the straight forward method. In the high band region with high energy, the positive value is given to the voice sound and the negative value is given to the unvoiced sound. In this we discuss a method to determine α by using only the HB-GSPP and LB-GSPP. In this method we assume that the positive α is chosen for the voiced sound which is more concentrated in the low-band region and for the high-band region speech signal which has most of its energy chooses negative α . To avoid abrupt discontinuity in spectral components we use soft-decision scheme which is

more helpful and this method is described as follows: Where, $P_{min} = 0.2$ and $\alpha \in [\alpha_{min} = (-0.02), \alpha_{max} = (0.02)]$. From different experimental tests the values of α_{min} and α_{max} are determined. Experimentally the optimized value for α is chosen as the higher values of α lead to signal degradation according to this experiment. If LB-GSPP is sufficiently small, then HB-GSPP approaches one and we can easily find out $\alpha(t)$ as it becomes α_{min} . When HB-GSPP is low then LB-GSPP increases and $\alpha(t)$ approaches α_{max} . The trajectory of α is shown in figure 2. We can say that α is negative with mostly high frequency components during the speech parts and α is positive for the voiced periods from the results. Figure 2: trajectory of α with corresponding speech waveform. Solid line denotes the $\hat{\alpha}(t)$ and dotted line denotes the $\alpha(t)$ with $\lambda_p = 0.2$ respectively. The temporal smoothing technique $\alpha(t)$ is applied to avoid rapid variation such that: Where $\hat{\alpha}(t)$ is the smoothed control parameter and λ_p is the smoothing parameter. For each of α , a WDCT matrix is required to implement WDCT-based speech-enhancement technique. To acquire specific value of α , it requires large computation and we can precompute the WDCT matrices and store them for particular value of α . For that computation, $[\alpha_{min}, \alpha_{max}]$ is uniformly divided into 16 regions and a WDCT matrix is constructed such that each region will have a center value. We can quantize α in 16 steps while considering the time-invariant characteristics of speech signal which helps to reduce the memory size. To transform the data WDCT matrix corresponding to that region is applied and the $\hat{\alpha}(t)$ region to which it belongs should also be identified during the speech enhancement. Experiments results and Discussions: According to the noise suppression in the IS-127, a 13ms length trapezoidal window was

applied to the input signal for every 10 ms. the blocking effect of the DCT can be decreased by the overlapping of the adjacent frames(3 ms). after zero-padding frame by frame, each frame of the signal can be transformed to the corresponding spectrum by applying it to WDCT. The figure below shows the experimental results obtained. TABLE ISEGMENTAL SNR RESULTS FOR THE WDCT AND DCT-BASED SPEECH-ENHANCEMENT METHODS.

NOISEBABBLECARWHITESNR(db)510155101551015DCT8. 6812. 6216. 989. 2813. 4517. 9310. 8114. 3818. 20WDCT8. 8612. 8317. 099. 3013. 4717.

9311. 1214. 7418. 53WIFig. 3. Comparison of speech segment under the white noise at SNR = 5dB (a) Clean speech. (b) Noisy signal.(c)Noisy speech signal (d) Enhanced speech by DCT. (e) Enhanced speech by WDCT. By varying the SNR ratio, babble, white, and car noises are electronically added to the clean speech waveforms. From the experimental results it is proved that WDCT is better at enhancing the noisy signal than the conventional DCT signal. CONCLUSIONIn this paper, an approach for speech enhancement using WDCT is proposed. Where WDCT is formed by cascading the adjustable all pass IIR filter with the conventional DCT , which results in the input signal transform. Split band analysis is used for the determination of warping control paramenter. Since WDCT matrices are predefined in a prescribed set of frequencies, with a little more computational burden the WDCT performs much better than the conventional DCT.