International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 11 (2015) pp. 28427-28438

http://www.ripublication.com

Formant Estimation of Speech Using Discrete Hartley Transform

Madhukar.B.N.
Senior Assistant Professor, ECE Department,
New Horizon College of Engineering, Bangalore, INDIA.
madhukarbn890@gmail.com
Sanjay Jain
Professor and Head, ECE Department,
New Horizon College of Engineering, Bangalore, INDIA.
hod_ece@newhorizonindia.edu
P. S. Satyanarayana
Senior Professor, ECE Department,
Don Bosco Institute of Technology, Bangalore, INDIA.
pssvittala@yahoo.com

Abstract

This paper presents a new method for formant estimation of speech signal using Discrete Hartley Transformapproach. Hitherto, the formant estimation of speech was carried out in temporal and frequency domains using Discrete Fourier Transform (DFT) based approach. A new approach of accomplishing this in the frequency domain using DHT, rather than using the DFT, is proposed. DFT, being a complex transform, takes more computation time for finding the formant frequencies of the speech signal, but DHT being a real transform takes less computation time to do the same with less memory requirement. The relationship between DFT and DHT is made use of for computing the formants, rather than using the DFT directly. The usage of DHT method is found to be optimal than using the direct DFT approach thereby saving implementation cost substantially in the formant estimation of speech signals. The analysis is done by comparing the usage of DFT directly on the speech signal and then using DHT, there by seeing the performance of these mathematical transforms based on computation time, which is used as a performance metric for validating the veracity of performance of the two discrete transforms concerned.

Keywords: DHT, DFT, liftering, cepstrum, window, formant.

Introduction

Formant Estimation is one of the most important research areas in Digital Speech Processing. Research in this prominent field has been and still being carried out with stupendous amount of success. The term formant is defined in the literature in different ways. Gunnar Fant, a leading speech processing expert of the 20th century, in 1960, defined the peaks of the sound spectrum, $|P(f)|$as formants [2]. The Acoustical Society of America in 1994 defines formant as therange of frequencies of a complex sound, in which there is an absolute or relative maximum in the sound spectrum, and the frequency at the maximum corresponds to the formant frequency [3].Benade in 1976 defined formants as the peaks that are observed in the spectrum envelope. In Speech Processing, it is defined as an acoustic resonance of the human vocal tract. In phonetics, formant corresponds to the resonance. It can also point to the spectral maximum that the resonance produces [4]. Formants are measured as amplitude peaks in the frequency spectrum of the sound, using a spectrogram or a spectrum analyzer. For voiced speech, we get a measure of the vocal tract resonances. If we consider vowels spoken with a high fundamental frequency, the formant frequency may lie between the widely spaced harmonics and hence no corresponding peak is seen. A technical study of this leads to the amalgamation of the domains of Speech Processing and Acoustical Signal Processing[6].

Formants are the differentiable frequency components of human articulation. The information that human beings differentiate between vowelsare represented by the frequency content of the vowel sounds. These are the characteristic overtones that identify vowels to the listener.The formant with the lowest frequency is called F_1, the second F_2, and the third F_3. Most often the two first formants, F_1 and F_2, are usually used to remove any uncertainty associated with dissemination of the vowel.The relationship between the vowel quality,thus perceived, and the first two formant frequencies can be appreciated by listening to artificial vowels that are generated by passing a click train to simulate the glottal pulse train through a pair of bandpass filters for simulating the vocal tract resonances [5].

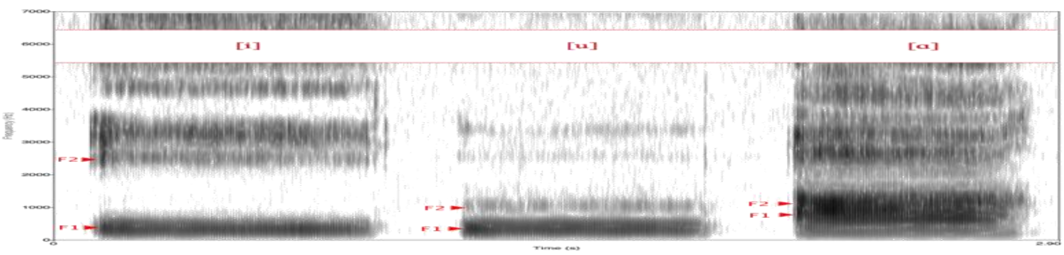

Figure 1: Spectrogram of American English vowels [i, u, a] showing the formants F_1 and F_2 [5].

The first two formants determine the quality of vowels in terms of the open/close and front/back dimensions. Thus the first formant F_1has a higher frequency for an open vowel (such as [a]) and a lower frequency for a close vowel (such as [i] or [u]); and the second formant F_2 has a higher frequency for a front vowel [7]. Vowels will almost always have four or more distinguishable formants. However, the first two formants are most important in determining the quality of the vowel and are displayed in terms of a plot of the first formant against the second formant. But this is insufficient to capture some of vowel qualityaspects like rounding [8].

Nasals usually have an additional formant around 2500 Hz. The [l] sound usually has an extra formant at 1500 Hz, whereas the English "r" sound is distinguished by virtue of a very low third formant that is below 2000 Hz [15].

Plosives and fricatives modify the placement of formants in the surrounding vowels. Bilabial sounds (such as /b/ and /p/ in "ball" or "sap") cause a lowering of the formants; velar sounds (/k/ in English) almost always show F_2 and F_3 , coming together in a velar pinch before the velar and separating from the same pinch as the velar is released. The time course of these changes in vowel formant frequencies are referred to as formant transitions [9].

If the fundamental frequency of the underlying vibration is higher than a resonance frequency of the system being used, then the formant communicated by that resonance is unrecovered. Control of resonances is an essential ingredient of vocal singing and this is popularly called overtone singing. In this, thesinger sings a low fundamental tone, and creates sharp resonances to cull upper harmonics. This in turn produces an edition of diverse tones being sung at one go. Spectrograms are used to visualize formants [10].

The Discrete Hartley Transform (DHT)

The Discrete Hartley Transform (DHT) of a discrete – time signal, $x(n)$, is defined as [1]

$$X_H(k) = DHT[x(n)] = \sum_{n=0}^{N-1} x(n) cas\left(\frac{2\pi kn}{N}\right) \ldots \quad (1)$$

$\nabla\, 0 \leq n, k \leq N-1.$

where, $cas\left(\frac{2\pi kn}{N}\right) = cos\left(\frac{2\pi kn}{N}\right) + sin\left(\frac{2\pi kn}{N}\right) \quad \ldots \quad (2)$

Also,we have the following formula given below,

$$cas\left(-\frac{2\pi kn}{N}\right) = cos\left(\frac{2\pi kn}{N}\right) - sin\left(\frac{2\pi kn}{N}\right) \quad \ldots \quad (3)$$

The Inverse Discrete Hartley Transform (IDHT) of $X_H(k)$ is defined as,

$$x(n) = IDHT[X_H(k)] = \frac{1}{N}\sum_{n=0}^{N-1} X_H(k) cas\left(\frac{2\pi kn}{N}\right) \quad \ldots \quad (4)$$

$\nabla\, 0 \leq n, k \leq N-1.$

Eqs. (1) and (4) are respectively called the Analysis and Synthesis Equations of DHT. Note that $cas\left(\frac{2\pi kn}{N}\right)$ is the kernel of DHT. It is a real term due to which the DHT is a real transform. The same kernel is made use of in the computation of the IDHT. Thus, unlike in the DFT, where there is a change of sign in the kernel in the synthesis equation, in the case of the DHT, there is no such thing which means that a single algorithm can be used to compute both the forward and backward transforms of

DHT, which is a major advantage over the conventional DFT. This is the equivalent to the case encountered in Continuous Hartley Transform. Hence, the DHT also is its own inverse. Thus, the DHT involves only real operations and hence, the computational load and memory requirement are considerably reduced by 50%, which stands out as a good merit over the normal DFT. Efficient algorithms called the Fast Hartley Transforms (FHT) are developed for computing the DHT and have been in use in many DSP applications [3].

Relationships Between DHT and DFT

First, the relationship between the DHT and DFT in the transform domain is considered followed by the relationship in the temporal domain.

A. Relationship between the Analysis Equations of DFT and DHT

The DFT of a discrete – time sequence $x(n)$ is defined [11] as

$$X(k) = DFT[x(n)] = \sum_{n=0}^{N-1} x(n)\, e^{-\frac{2\pi kn}{N}} \ldots \quad (5)$$

$\nabla\, 0 \leq n, k \leq N-1.$

Using Euler's Theorem, $e^{\pm j\gamma} = cis(\pm\gamma) = \cos(\gamma) \pm j\sin(\gamma)$, we get,

$$X(k) = \sum_{n=0}^{N-1} x(n)\left[\cos\left(\frac{2\pi kn}{N}\right) - j\sin\left(\frac{2\pi kn}{N}\right)\right] \ldots \quad (6)$$

Due to the absence of the j term in the expression for $X_H(k)$, DHT is purely real. Next, consider the following expression [12], $e^{j\alpha} = cis(\alpha) = \cos(\alpha) + j\sin(\alpha)$.

We knowby Euler's theorem.thatcos $\alpha = \frac{e^{j\alpha}+e^{-j\alpha}}{2}$ and $\sin\alpha = \frac{e^{j\alpha}-e^{-j\alpha}}{2j}$.

$$\Rightarrow cas\alpha = \cos\alpha + \sin\alpha = \frac{(1-j)e^{j\alpha}}{2} + \frac{(1+j)e^{-j\alpha}}{2} \quad \ldots \quad (7)$$

We make the following assumptions that

$$\chi = e^{j\alpha}, k = cas\alpha, \xi = \frac{(1-j)}{2}, \delta = \frac{1+j}{2} \quad \ldots \quad (8)$$

Then, we have the quadratic equation, $\xi\chi^2 - k\chi + \delta = 0 \ldots$ (9)

Solving for χ, we get,$\Rightarrow \chi = \frac{\rho}{2\xi} \pm \frac{1}{2\xi} j\,(\cos\alpha - \sin\alpha) \ldots$ (10)

Considering only the negative sign and simplifying, we get,

$$\chi = e^{j\alpha} = \frac{1+j}{2} cas\alpha + \frac{1-j}{2} cas(-\alpha) \ldots \quad (11)$$

Also, we can write $\chi^{-1} = \frac{1}{\chi} = e^{-j\alpha} = \frac{1+j}{2} cas(-\alpha) + \frac{1-j}{2} cas(\alpha) \ldots$ (12)

Now, the DFT of$x(n)$ is given by the well known equation

$$X(k) = DFT[x(n)] = \sum_{n=0}^{N-1} x(n)\, e^{-\frac{2\pi kn}{N}} \ldots \quad (13)$$

Making use of Eq. (12) in Eq. (13), we get,$X(k) = \frac{1+j}{2} X_H(-k) + \frac{1-j}{2} X_H(k)$

$$\therefore X(k) = \frac{1}{2}[X_H(k) + X_H(-k)] + j\frac{1}{2}[X_H(-k) - X_H(k)]$$

Since $X(k)$ is complex, we can write the above equation as,

$$X(k) = X_R(k) + jX_I(k) \ldots \quad (14)$$

Thus, the DFT coefficients in terms of DHT coefficients are given by the following two equations [1]

$$X_R(k) = \frac{1}{2}[X_H(k) + X_H(-k)] \quad \ldots \quad (15)$$

$$X_I(k) = \frac{1}{2}[X_H(-k) - X_H(k)] \quad \ldots \quad (16)$$

Subtracting Eq. (16) from Eq. (15), DHT coefficients that can be expressed in terms of DFT are got.

$$X_H(k) = X_R(k) - X_I(k) \ldots \quad (17)$$

B. Relationship Between The Synthesis Equations of DFT and DHT

If$x_H(n) = IDHT[X_H(k)]$, then it is possible to express the $IDHT$ of $X_H(k)$ in terms of the IDFT of $X(k)$ [14]

$$x_H(n) = x_R(n) - x_I(n) \ldots \quad (18)$$

Here, $x_R(n)$ and $x_I(n)$ are the real and imaginary parts of $x(n)$ respectively. One must bear in mind that $x_R(n) = IDFT[X_R(k)]$ and $x_I(n) = IDFT[X_I(k)]$ respectively.

DHT finds applications in a variety of domains in Communications Engineering, Signal Processing, Image Processing, Vibration Analysis, etc [10].

Cepstrum of Speech Signal and Formant Estimation

Cepstrum was invented jointly by B. P. Bogert, M. J. R. Healy, and J. W. Tukey in 1963. The name "cepstrum" was derived by reversing the first four letters of "spectrum." Operations on cepstra are labelled *quefrency analysis*, or *cepstral analysis*. A cepstrum is the result of taking the Inverse Fourier Transform (IFT) of the logarithm of the estimated spectrum of a signal [12].

Speech is considered to be the output of a system, the vocal tract $V(z)$to an input $p(n)$ which is either a periodic impulse train due to the vibration of vocal chords or the white noise due to the air flow. Speech can be broadly classified as voiced for the periodic impulse train input and as unvoiced speech for the white noise input. Speech is composed of excitation source and vocal tract system components[13]. The usage of cepstralprocessing of speech signal is to segregate its excitation and vocal tract components without any advanced knowledge about source and/or system. Voiced sounds are generated by exciting the time varying system characteristics with periodic impulse signals. The unvoiced sounds are generated by exciting the time varying system with a stochastic noise sequence. Hence, the resultant speech signal is the linear convolution of the corresponding excitation sequence and vocal tract filter characteristics. Let $a(n)$ and $h(n)$ be the excitation sequence and vocal tract filter sequences. Then, the speech signal is given by the one-dimensional discrete – time linear convolution of $a(n)$with $h(n)$ [14]

$$s(n) = a(n) \otimes h(n) \quad \ldots \quad (19)$$

where, the symbol '⊗' denotes discrete – time linear convolution.Taking Discrete – Time Fourier Transform (DTFT) on both sides of Eq. (19) yields the frequency spectrum [2],

$$S(\omega) = A(\omega).H(\omega) \quad \ldots \quad (20)$$

The speech signal is then deconvolved into the excitation and vocal tract components in the temporal domain. Taking absolute values forEq. (20), we get,

$$\Rightarrow |S(\omega)| = |A(\omega)|.|H(\omega)| \quad \ldots \quad (21)$$

Taking Naperian or Natural logarithms for Eq. (21), we obtain,

$$\ln|S(\omega)| = \ln|A(\omega)| + \ln|H(\omega)| \quad \ldots \quad (22)$$

The logarithmic operation is used for transforming the magnitude speech spectrum where the excitation and vocal tract components are multiplied, into a linear combination of them [15]. The segregation is done by taking the Inverse Discrete Fourier Transform (IDFT) of $\ln|S(\omega)|$, which yields the cepstral coefficients in the discrete – time domain. The fast varying components in the higher cepstral region correspond to the excitation components while the slow moving components in the lower cepstral region correspond to the vocal tract components.

$$c(n) = IDFT[\ln|S(\omega)|] = IDFT[\ln|A(\omega)| + \ln|H(\omega)|] \ldots \quad (23)$$

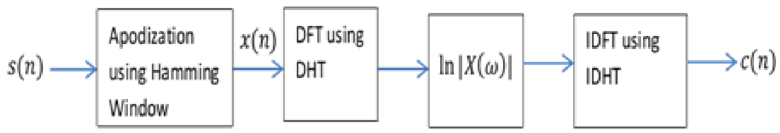

Figure 2: Cepstral Computation using DHT Approach [10].

Sometimes, variations occur in the lower cepstral region due to the vocal tract characteristics and the rapid varying strata of the cepstrum towards the upper cepstral region that is represented by the excitation characteristics of the short time speech segment [7]. Liftering is used for extracting vocal tract and excitation characteristics in the discrete – time domain.Liftering operation is similar to filtering operation in the frequency domain where a desired cepstral region for analysis is selected by multiplying the whole cepstrum by a Boxcar apodization function at the desired position. Liftering is a useful and meaningful process with the real cepstrum for obtaining an estimate of the log spectrum of either of the separated components [9]. That is, we can apply a useful linear operation to the real cepstrum. The output of this process in the quefrency domain is a real cepstrum. However, if the objective is to return to the original time domain with an estimate of the separated signal, the real cepstrum will fail, because its "linearizing" operation is not invertible. To complete this task, we would need a phase – preserving linearizing operation [13].

Low – timeliftering is used for estimating slow varying vocal tract characteristics from the computed cepstrum of the given speech signal. For accomplishing this, the low-time liftering window is used for extracting vocal tract characteristics. If T is the cut off length of the liftering window and $\frac{N}{2}$ is half the total length of the cepstrum, N, then, the low-time lifteringwindow function is represented mathematically as [4]

$$w_s(n) = \begin{cases} 1, 0 \leq n \leq T \\ 0, T \leq n \leq \frac{N}{2} \end{cases} \quad \ldots \quad (24)$$

A value of 15 or 20 is assigned to T. The vocal tract characteristics are computed by taking the product of the cepstral coefficients $c(n)$ with the low-time liftering window function.

$$c_s(n) = c(n).w_s(n) \quad \ldots \quad (25)$$

Application of DFT on the low – time liftered signal gives its log magnitude spectrum. This is nothing but the vocal tract spectrum of the given short – timespeech segment [6]. That is, we have,

$$\ln|H(k)| = DFT[c_s(n)] \quad \ldots \quad (26)$$

The computation of the DFT is done by using its relationship with DHT. We make use of the DHT indirectly to compute the DFT. The natural logarithm of the DFT is taken which is applied to a peak picking algorithm whose output gives the formants that is the desired output. The formant locations can be estimated by picking the peaks from the smooth vocal tract spectrum. Figure 3 shows the block diagram depicting the formant estimation technique [5].

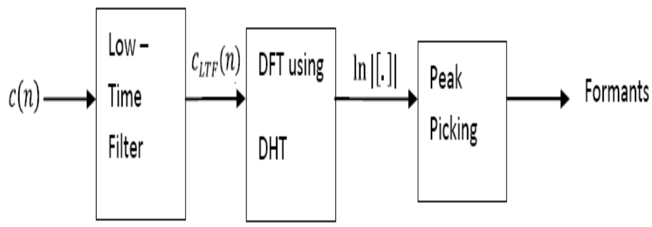

Figure 3: Block Diagram showing formant estimation using DHT approach [8].

Application of The Discrete Hartley Transform (DHT) In Digital Speech Processing

DHT, since it is a real transform, is very much useful in the cepstral analysis of speech. If $x(n)$ is the discrete speech signal, firstly, its N-point DHT, $X_H(k)$ is computed.Then, the DFT of $x(n)$ is calculated by exploiting the relationship between DFT and DHT, by using the following relations [11].

$$X_R(k) = \frac{X_H(k)+X_H(-k)}{2} \quad \ldots \quad (27)$$

$$X_I(k) = \frac{X_H(-k) - X_H(k)}{2} \quad \ldots \tag{28}$$

$$X(k) = X_R(k) + jX_I(k) \quad \ldots \tag{29}$$

We next express $X(k)$ in Steinmetz form or polar form as,

$$X(k) = |X(k)|e^{j\Theta(k)} \quad \ldots \tag{30}$$

We know that

$$\hat{x}(n) = IDFT[log_e\{X(k)\}] = IDFT[ln\{X(k)\}]$$

$$\Rightarrow \hat{x}(n) = IDFT\left[\ln\{|X(k)|e^{j\Theta(k)}\}\right]$$

$$\hat{x}(n) = IDFT[\ln|X(k)| + j\Theta(k)] = IDFT\left[\hat{X}(k)\right] \ldots \tag{31}$$

DHT is then computed by using the relation, $\hat{X}_H(k) = \hat{X}_R(k) - \hat{X}_I(k)$, where, $\hat{X}_R(k) = \ln|X(k)|$ and $\hat{X}_I(k) = \Theta(k)$.

$$\Rightarrow \hat{X}_H(k) = \ln|X(k)| - \Theta(k) \quad \ldots \tag{32}$$

Finally, the cepstral coefficients $\hat{x}(n)$, are computed by taking the IDHT of $\hat{X}_H(k)$, i.e.,

$$\hat{x}(n) = IDHT\left[\hat{X}_H(k)\right] \quad \ldots \tag{33}$$

The cepstra of voiced speech segment using DHT approach are shown in Figure4.Cepstral Analysis on speech signal is to segregate its excitation and vocal tract components. The input speech signal is converted into short-term segments of duration 15 – 20 msec. The frame size is maintained to 20 msec and then each frame is multiplied by Hamming Window. Then, the cepstral representation of short-term speech is computed by finding the IDFT of the log magnitude spectrum. In this work, instead of using the IDFT directly, we have used IDHT to compute IDFT [4].

Result Analysis and Validation

Figure 4 shows a 20 msec voiced frame of the sentence, "I like my mother," and its cepstrum. The vocal tract components are concentrated in the lower cepstral region and the excitation components are concentrated in the higher cepstral region. Also shown is $s(n)$ which is the voiced frame considered and $x(n)$ which is the windowed frame. Also shown is the cepstral coefficients $c(n)$. Note that $c(n)$ is symmetrical in the cepstral domain.Figure 5 shows the vocal tract characteristics of the cepstrum obtained by applying the procedure elucidated above. Figure 6 shows the formants locations obtained from the peaks in the vocal tract spectrum. The first peak corresponds to the first formant $F_1 = 752$ Hz, the second peak corresponds to the second formant $F_2 = 1{,}650$ Hz, and the last peak corresponds to the third formant $F_3 = 2{,}820$ Hz, respectively.

Conclusions

The DHT is two to three times faster than DFT because of no complex arithmetic being involved. All the four quadrants of the Hartley domain data must be used. In the computation of DHT and its inverse, only one quadrant of sines and cosines need to be calculated due to symmetry. Also, the multiplication in DHT is real but complex in the case of DFT. The DHT butterfly loop requires less memory space than the DFT because all of the data are stored in arrays of real numbers. The DFT butterfly loop on the other hand makes use of complex arrays that needs double the memory space of a real array. In the DFT computation, there is one multiplication and two additions of complex quantities that accruesto four multiplications and six additions of floating point numbers for each and every iteration. Also, the DHT has our multiplications and six additions of floating point numbers for each iteration.

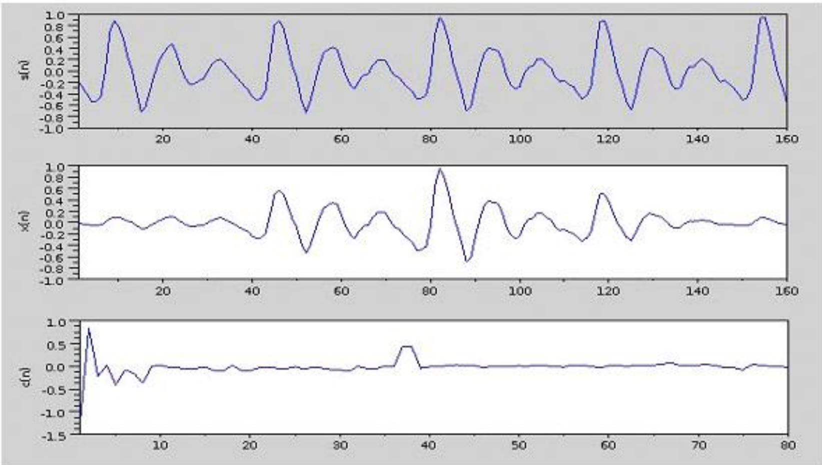

Figure 4: Cepstrum of voiced speech segment using DHT

The butterfly loop of the DHT loops from 2 to $\frac{N}{2}$, while the butterfly loop of the DFT loops from 1 to N. Since the DHT loops half the number of times as the DFT, the DHT algorithm has two multiplications and three additions for every four multiplications and six additions of the DFT. Also, the DHT has no multiplications for the zero and Nyquist frequencies, which is a major advantage of DHT over DFT.DHT needs less memory to store numbers than the DFT because DHT does not use complex numbers. The results of the DHT can be stored in the same memory space as the original data set, thus eliminating the need to allocate more disk space. Also converting from Hartley domain to Fourier domain and vice versa is a direct and simple procedure.

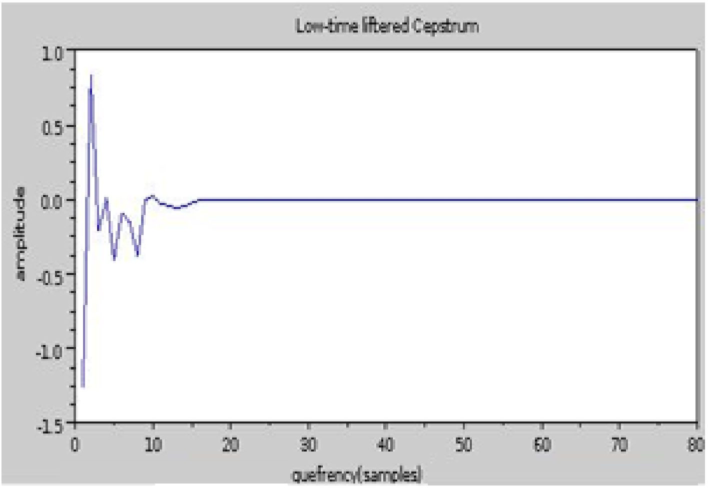

Figure 5: Vocal Tract Characteristics of The Cepstrum Obtained Using Low-Time Liftering Method

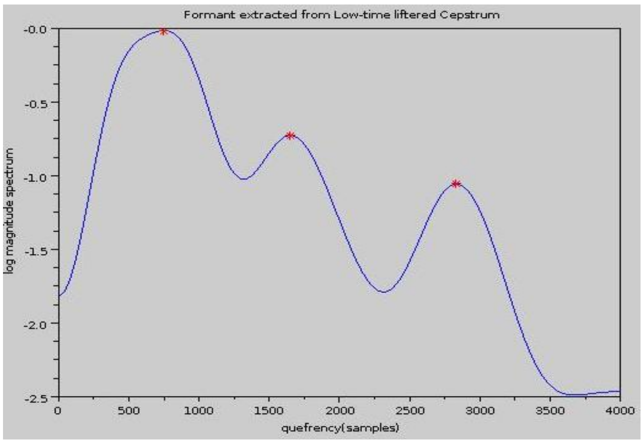

Figure 6: Formant Locations Obtained From Vocal Tract Spectrum

In this work carried out by the authors, formant estimation of voiced speech segment using both DFT and DHT approaches were considered. Usage of direct DFT approach required 1171.5085 complex additions and 585.75425 complex multiplications. Usage of DHT approach required 292.9 multiplications and 585.823 additions, which indicate that DHT has saved the computation load by 50%. Also, computation time is used as the performance metric for evaluating the performance of DHT and DFT approaches. DFT approach took 7.235 seconds for formant estimation whilst the DHT took 3.634 seconds for accomplishing the same task. This Thus, by using the DHT approach, the computation time, and the memory requirement are saved by 50% in the cepstral analysis of speech. This is a good improvement and the DHT based method is good when compared to the traditional DFT approach.The first formant frequency occurred at$F_1 = 752$ Hz, the second one at $F_2 = 1{,}650$ Hz, and the last one at $F_3 = 2{,}820$ Hz, respectively. The same results were obtained while DFT approach. This proves our claim that DHT yields faster results for formant estimation than compared to DFT approach.

Future Scope and Study

The present work is based on the concept of the basic Discrete Hartley Transform. Other variants of Hartley Transform and its amalgamation with Wavelet Transform need to be worked within the nature of quasiperiodicity and the nonstationarity of the speech signal. The behaviour of the speech signal spectrum to such robust algorithms needs to be analyzed.

Acknowledgement

The first two authors thank Dr. Mohan Manghnani, Chairman, New Horizon Education Institution, for providing the necessary infrastructure and encouraging us to carry out this research work. We are grateful to Dr. Manjunatha, Principal, New Horizon College of Engineering,forhis constant support to achieve ourendeavours. The third author is thankful likes to the Management of Don Bosco Institute of Technology for their wholehearted support in carrying out this work.

References

[1] Ronald N. Bracewell, "The Hartley Transform," 1st Edition, Oxford University Press, London, U.K., 1986.
[2] G. Fant, "Acoustic Theory of Speech Production," 1st Edition, Mouton & Co., The Hague, Netherlands, 1960.
[3] Chris Rowden (Editor), "Speech Processing," 1st Edition, The Essex Series in Telecommunications & Information Systems, McGraw-Hill Publishing Company, Berkshire, England, 1992.
[4] Lawrence R. Rabiner, and Ronald W. Schafer, "Digital Processing of Speech Signals," 1st Edition, Pearson India Inc., New Delhi, 2005.
[5] Douglas O'Shaughnessy, "Speech Communications: Human and Machine," 2nd Edition, Cambridge University Press India Private Limited, New Delhi, India, 2007.
[6] John R. Deller, Jr., John H.L. Hansen, and John G. Proakis, "Discrete-Time Processing of Speech Signals," IEEE Press, Wiley India Private Limited, New Delhi, 2014.
[7] A.M. Kondoz, "Digital Speech," 2nd Edition, Wiley India Private Limited, New Delhi, India, 2009.
[8] Lawrence R. Rabiner, and Biing-Hwang Juang, "Fundamentals of Speech Recognition," 1st Edition, Pearson India Inc., New Delhi, 2005.
[9] Thomas F. Quatieri, "Discrete-Time Speech Signal Processing," 1st Edition, Pearson India Inc., New Delhi, 2006.

[10] Ben Gold and Nelson Mogran, "Speech and Audio Signal Processing: Processing and Perception of Speech and Music," 1st Edition, Wiley India Private Limited, New Delhi, India, 2009.
[11] Athanasios Papoulis, "The Fourier Integral and its Applications," 1st Edition, McGraw-Hill Book Company Limited, New York, U.S.A., 1962.
[12] Shaila D. Apte, "Speech and Audio Processing," 1st Edition, Wiley India Private Limited, New Delhi, India, 2014.
[13] Nasir Ahmed and K. R. Rao, "Orthogonal Transforms in Digital Signal Processing," Springer – Verlag Private Limited, New York, U.S.A., 1975.
[14] Murray R. Spiegel, "Schaum's Outline of Fourier Analysis with Applications to Boundary-Value Problems," 1st Edition, Tata McGraw-Hill India Private Limited, New Delhi, India, 2007.
[15] J. L. Flanagan, "Speech Analysis, Synthesis and Perception," 2nd Edition, Springer – Verlag, New York, U.S.A., 1972.