Zero Time Windowing Analysis of Hypernasality in Speech of Cleft Lip and Palate Children

Akhilesh Kumar Dubey, S. R. Mahadeva Prasanna and S. Dandapat
Department of Electronics and Electrical Engineering
Indian Institute of Technology Guwahati, Guwahati-781039, India
Email: {d.akhilesh, prasanna, samaren}@iitg.ernet.in

***Abstract*—Hypernasality is a condition where the speech is highly nasalized resulting in significant reduction of intelligibility. Hypernasality happens due to severe leakage of air through nasal cavity because of either cleft palate and / or incomplete closure of velum. Hypernasality can be visualized as a phenomenon of introducing a nasal formant in the vicinity of first formant. To distinguish the nasal formant from the first formant, better spectral resolution is desirable. Traditional methods of formant extraction give average formant characteristics with 20-30 ms window size and poor resolution, when one pitch period is used. To overcome this poor resolution issue, especially for children speech case, this work proposes the use of zero time windowing (ZTW) technique. In ZTW, the speech signal is multiplied with a highly decaying impulse-like window of size approximately a pitch period. The loss in spectral resolution due to zero time windowing is restored by successive differentiation in frequency domain. Further the numerator of group delay is used to resolve closely spaced formants. The method when applied to hypernasal vowel sounds /a/, /i/ and /u/ gives accurate and better resolved location of nasal formant in the vicinity of first formant as compared to traditional methods.**

Keywords*—Hypernasality, CLP speech, nasal formant, zero time windowing, numerator of group delay.

I. Introduction

The cleft lip and palate (CLP) is an anatomical birth defect in children which causes hypernasality in speech. Hypernasality is due to cleft palate and velopharyngeal insufficiency i.e inability of velum to proper closing of velopharyngeal gap [1]. The nasal cavity gets coupled with the oral cavity causing excess of air flow through the nasal cavity even in case of vowel sounds. This causes nasalization of vowel sounds and hence reduction in speech intelligibility. Quantitative assessment of hypernasality helps in finding the result of multiple surgeries performed by plastic surgeons followed by speech therapy done by therapists. Generally two class of approaches are used for the assessment: Invasive and Non-invasive. The invasive approach uses the invasive instruments such as nasendoscopy, videofluoroscopy, etc to observe the movement of velopharyngeal part during speech production. The non-invasive approach can be further divided into clinical assessment and signal processing based technique. In clinical assessment, devices are used to measure pressure, vibration, volume velocity and nasal flow to assess hypernasality. The Nasometer is an example of such type of device. The signal processing based method uses the signal processing techniques to assess hypernasality. The invasive approaches are uncomfortable to children and clinical assessment requires costly instruments. Also the output information of the instruments requires an expert for the judgment. Alternatively, the signal processing based method is computer based which can be easily implemented in software, on portable devices, fast, cheaper and unbiased.

Acoustic Analysis on nasalized vowel sounds gives cue for the detection of hypernasality. The most important is the presence of an extra pole-zero pair in nasalized vowel sounds [2]. The concept of 'filling in' of the valleys in the spectrum above or below the frequency of the first formant by an extra pole-zero pair is observed in [3]. In [4], the presence of extra formant in low frequency (around 250 Hz) is discussed as an important cue for detection of hypernasality. The other cues which are observed in hypernasal speech are: broadening of formant bandwidths, specially the first formant and the presence of anti-formants in the spectrum. Based on these cues, the difference between the amplitude of first formant (A_1) and the amplitude of the extra formant (P_1) has been proposed as a promising measure for nasality detection [5]. The difference between the low-pass and band-pass profile of the Teager energy operator as a measure for nasal detection is proposed in [6]. Similarly, the distance between the cepstrum coefficients extracted from linear prediction coefficients of low and high order [7], detection of the extra peak around 250 Hz with the help of group delay method [4], use of mel-frequency cepstral coefficients (MFCC) and Teager energy operation to detect hypernasality in children speech [8], use of the shift in energy to the low frequency band as a cue for hypernasality detection [9], are some other methods proposed in the literature.

All above methods focused on the low frequency region, either to detect the extra formant introduced due to hypernasality or energy shift due to extra formant in this region. Traditional method of formant extraction like magnitude spectrum, linear prediction (LP) analysis and cepstrum analysis uses the window size of 20-30 ms to compute spectrum. Each of these methods have their own limitations. The magnitude spectrum has the effect of pitch and harmonics. In LP spectrum, the number of peaks depends on the order of the LP model. The cepstrum based spectrum depends on the size of low time liftering window. By adjusting these issues, these methods give the average characteristics of the vocal tract system within the window segment. These methods are not suitable for capturing changes in highly non-stationary cases especially, within the closed and open glottis phase of one pitch period. To observe these changes, the analysis window should be small enough (around one pitch period). The traditional methods when used with small window size, gives very poor resolution [10].

The poor resolution spectrum cannot be used for hypernasality analysis because they will not be able to resolve the closely spaced first formant and the extra nasal formant in vicinity of first formant. Even the group delay spectrum [11] cannot be used because it gives spiky peaks which masks the peaks due to actual formants. The modified group delay spectrum [4] is able to resolve the peaks but the accuracy is poor.

In this paper we therefore attempted the issue of extracting the vocal tract characteristics and hence the extra formant from very short (around 5 ms) segment of hypernasal speech. The operation is called zero time windowing (ZTW) [12]. The speech signal is multiplied with a window of short duration. The window is highly decaying impulse-like function and is an approximation to integration operation in the frequency domain. The loss in resolution is restored by successive differentiation in frequency domain. To resolve the closely spaced formants, the numerator of group delay function is used because of its additive property. Since the window size is small and impulse-like, the vocal tract characteristics can be found at each sampling instant. The proposed method therefore may be able to both extract vocal tract characteristics within a pitch period and also resolve closely spaced formants in case of hypernasal speech of CLP children.

The organization of the paper is as follows: Section II discusses brief overview of ZTW operation. Section III discusses the proposed ZTW method of hypernasality analysis. Sections IV discuss the different experiments. The summary and conclusions are given in Section V.

II. Zero-time windowing of speech

The frequency response of zero frequency resonator (ZFR) [13] is

$$|H(w)| = \left|\frac{1}{(1-z^{-1})^2}\right|_{z=e^{jw}} = \frac{1}{2(1-\cos w)} = \frac{1}{4\sin^2 w/2} \quad (1)$$

The filtering of speech signal through a ZFR is represented by following operation in time domain

$$y[n] = 2y[n-1] - y[n-2] + s[n] \quad (2)$$

where $s[n]$ is DC bias removed speech input and $y[n]$ is the output of the resonator. The frequency domain equivalence of above operation is to multiply the spectrum of speech $s[n]$ with a window function given by the frequency response of resonator $|H(w)|$. This is equivalent to integrating the speech signal twice in time domain to provide a sharp role off.

ZTW is a time domain operation analogous to zero frequency filtering (ZFF) operation in frequency domain. The time domain signal is multiplied with a window function similar in the shape of frequency response window of the ZFR. The window is called zero time window function and is given by

$$w_1[n] = \begin{cases} 0, & n = 0 \\ \frac{1}{(4\sin^2(\pi n/(2N)))}, & n = 1, 2, ..., N-1 \end{cases} \quad (3)$$

where N is the window length. This window function or analogous window in ZFR can be approximated to $\frac{1}{n^2}$ for smaller value of n and $N >> M$, which is equivalent to integrating the signal spectrum in frequency domain. M is the length of the speech segment and N is chosen to be the discrete Fourier transform (DFT) length. Window function gives more weight to the zero time samples and less weight to all other samples of the segmented signal.

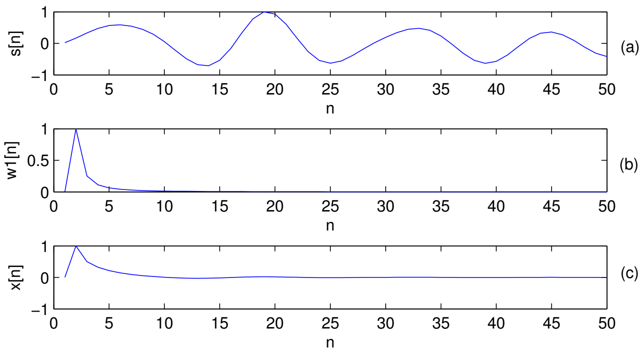

Fig. 1. Illustration of ZTW of speech signal.(a) Short segment of speech waveform. (b) ZTW function $w_1(n)$ (c) ZT windowed speech waveform $x(n)$

The concept of ZTW is shown in Fig. 1. Speech segment of length 5 ms is shown in Fig. 1(a). The ZTW function is shown in Fig. 1(b). The multiplication of the speech signal and ZTW function is shown in Fig. 1(c). The nature of windowed signal looks almost like an impulse, since the signal have high value of amplitude only around zero time $n = 0$ (reference time) and hence the name zero time windowing. Since the spectrum of windowed signal will show the characteristics of the signal around $n = 0$, such characteristics are also called instantaneous spectral characteristics.

III. Zero time windowing method for hypernasality analysis

The CLP is an anatomical birth defect that occurs in children. The cleft palate and the inability of the velum to completely close the velopharyngeal gap leads to coupling of nasal and oral cavities. This introduces extra nasal formant into the spectrum of oral vowel sounds which causes the nasalization of the vowels. The detection of extra peaks in nasalized vowels is the cue for the hypernasality. Since the vocal tract characteristics changes within a pitch period also, high temporal resolution is needed. The ZTW method is proposed for applications needing higher temporal resolution and hence can be used to detect the formant at each sampling instant.

In ZTW method, a segment of speech starting at each sampling instant is multiplied with a ZTW function $w_1[n]$. The various steps involved in extracting the instantaneous spectral characteristics and hence formants using ZTW operation are as follows:

(a) Consider the differenced speech signal $s[n]$ at a sampling frequency of f_s Hz . The signal is differenced to remove low frequency bias in the signal.

(b) Consider $s[n]$ of M samples starting from an arbitrary reference sampling instant set at $n = 0$ to end at $M - 1$.

(c) Take DFT length $N >> M$. Make the length of $s[n]$ equal to N by $N - M$ zero padding.

(d) Multiply N length $s[n]$ segment with a window function $w_1^2[n]$. This window function emphasizes the values near the $n = 0$ sampling instant.

(e) The truncation effect at the end of window ($M - 1$ sampling instant in time domain) results ripple in the frequency domain. This effect is reduced by using another window.

$$\begin{aligned} w_2[n] &= 2(1 + \cos(\pi n/M)) \\ &= 4\cos^2(\pi n/2M), \qquad n = 0, 1, ..., M - 1. \end{aligned} \tag{4}$$

(f) Compute N-point DFT of the double windowed signal, i.e. of $x[n] = w_1^2[n]w_2[n]s[n]$. The square magnitude spectrum of windowed signal is smooth due to equivalent four time integration in frequency domain.

(g) To highlight the spectral characteristics, the numerator of the group delay (NGD) function $g[k]$ of the windowed signal $x[n]$ is computed [14]. The NGD function $g[k]$ is the numerator part of the group delay function proposed in [10]. The group delay function for the signal $x[n]$ is given by [15]

$$\tau[k] = \frac{X_R[k]Y_R[k] + X_I[k]Y_I[k]}{X_R^2[k] + X_I^2[k]}, k = 0, 1, ..., N - 1. \tag{5}$$

where $X_R[k]$ and $X_I[k]$ are the real and imaginary parts of the N-point DFT $X[k]$ of $x[n]$ respectively and $Y_R[k]$ and $Y_I[k]$ are the real and imaginary parts of the N-point DFT $Y[k]$ of $y[n] = nx[n]$ respectively. The numerator of the group delay function $g[k]$ is given by

$$g[k] = X_R[k]Y_R[k] + X_I[k]Y_I[k], k = 0, 1, ..., N - 1. \tag{6}$$

The group delay function has high frequency resolution due to additive property ($\tau[k] \propto |H[k]|^2$ around the formants frequency) [16]. The spectral resolution further increases in NGD due to ignorance of denominator term ($g[k] \propto |H[k]|^4$ around the formants frequency) [10].

(h) To further highlight the spectral characteristics like peaks corresponding to formants, NGD function is differenced twice (DNGD) in frequency domain. This operation is opposite to twice integration operation due to multiplication of speech signal by ZTW function.

(i) The low amplitude spectral peaks in double differenced NGD spectrum gets suppressed due to valleys. The Hilbert envelope (HE) of the double differenced NGD spectrum is computed to remove the effect of spectral valleys. The resulting envelope is called HNGD spectrum.

The Hilbert envelope $a[n]$ of a sequence $e[n]$ is obtained as

$$a[n] = \sqrt{e^2[n] + e_h^2[n]} \tag{7}$$

where $e_h[n]$ is the Hilbert transform of the sequence $e[n]$, and is computed as follows:

$$e_h[n] = IDFT\{E_h(w)\} \tag{8}$$

where,

$$E_h(w) = \begin{cases} -jE(w), & 0 < w < \pi \\ jE(w), & -\pi < w < 0 \end{cases} \tag{9}$$

and $E(w)$ is the DTFT of the sequence $e[n]$

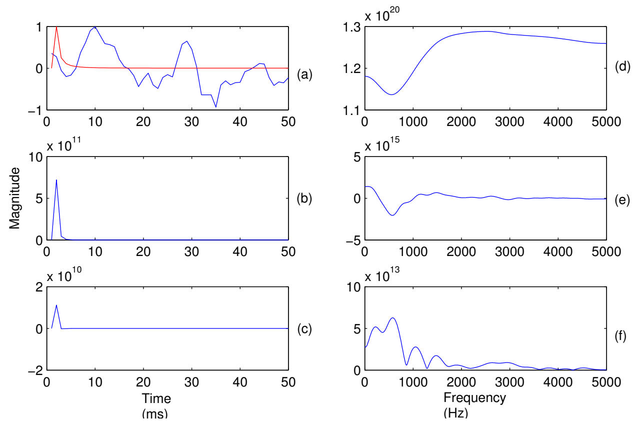

Fig. 2. Illustration of ZTW method for hypernasality detection. (a) Short segment (5 ms) of hypernasal vowel /a/ and ZTW function. (b) Combined window function $w(n) = w_1^2(n)w_2(n)$. (c) Windowed speech waveform $x(n) = s(n)w(n)$. (d) NGD spectrum of (c). (e) Double derivative of (d). (f) HE of (e)

The complete illustration of the ZTW method is shown in Fig. 2 where Fig. 2 (a) Shows the 5 ms speech waveform and ZTW function in the same figure. The highly exponential nature of the ZTW function can be observed. It gives high weight to the samples near zero time of the speech segment than the other samples in the segment. Multiplication with the ZTW function in time domain is equivalent to integration of the speech spectrum in frequency domain. The combined window function $w(n) = w_1^2(n)w_2(n)$ is shown in Fig. 2(b). The multiplication of combined window function with the speech signal in (a) is shown in Fig. 2(c). The NGD spectrum of Fig. 2(c) is shown in Fig. 2(d). The twice successive derivative of Fig. 2(d) is shown in Fig. 2(e). The Hilbert envelope of the Fig. 2(e) is plotted in Fig. 2(f) which is the final HNGD spectrum.

The HNGD spectrum shows many peaks. All the peaks are not formants. Only the dominant peaks are the formants. From the HNGD spectrum for the hypernasal vowel /a/ in Fig. 2 (f), we can see a dominant peak below 500 Hz and a dominant peak around 700 Hz. The dominant peak around 700 Hz is the first formant for the vowel /a/ and the dominant extra peak below 500 Hz is the clear cue for the hypernasality. The additional extra dominant peak below 500 Hz is due coupling of nasal cavity with the oral cavity in case of Cleft Palate with or without Cleft Lip.

IV. SPECTRAL COMPARISON OF HYPERNASAL SPEECH USING DIFFERENT METHODS

The significance of HNGD spectrum method for formant extraction can be tested on synthetic nasal speech signal generated by known formants and their bandwidths. The vocal tract filter model for speech production based on all pole model is given by

$$V(z) = \prod_{k=1}^{k=P} \frac{1}{1 - 2e^{-\pi B_k T}\cos(2\pi F_k T)z^{-1} + e^{-2\pi B_k T}z^{-2}} \tag{10}$$

where the F_k and B_k are the known formants and their bandwidths, respectively, and P is the number of formants. Here we have taken four formants at 400 Hz, 800 Hz, 1300 Hz, and 2100 Hz and 10% of these values as the corresponding bandwidth of formants. The speech signal is generated with sampling frequency 10 kHz. The formant at the frequency 400 Hz is responsible for the nasality in the generated synthetic speech.

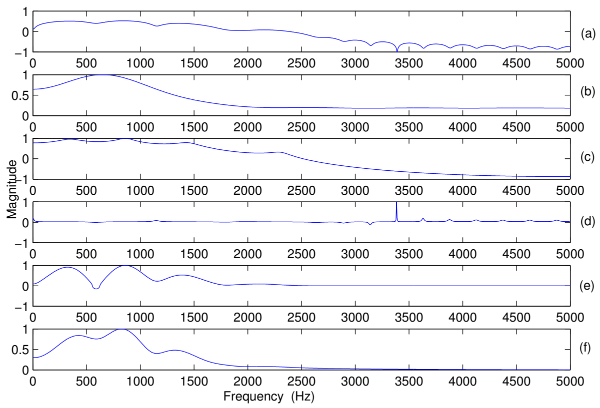

Fig. 3. Comparison of traditional methods, Group Delay method and HNGD method based on ZTW for the extraction of formants in case of synthetic nasal speech with window size of 4 ms. (a) Magnitude spectrum (b) Cepstrum spectrum (c) LP spectrum (d) Group Delay spectrum and (e) Modified Group delay spectrum (f) HNGD spectrum

Fig. 3 shows (a) magnitude spectrum, (b) cepstrum spectrum, (c) LP spectrum, (d) Group Delay spectrum, (e) Modified Group delay spectrum (MGD) and (f) HNGD spectrum for synthetic nasal speech with window size of 4 ms. The hanning window is used for traditional and group delay methods. The traditional methods give poor resolution with small window size of 4 ms. The magnitude spectrum is generally not used for formant extraction because of the effect of glottal excitation and the analysis window gets manifested on the spectrum of the vocal tract impulse response. Therefore the standard peak picking algorithm is not used for the formants extraction from the magnitude spectrum. The cepstrum based spectrum depends on the number of low time liftering coefficients. When the number of coefficients are less, it does not resolves the closely spaced formants at 400 Hz and 800 Hz as shown in Fig. 3 (b). When the number of coefficients are increased, it will resolve all formants but produces spurious peaks in cepstrum spectrum. The peak picking method will pick these peaks as formants. In the similar way, the LP spectrum is

TABLE I. PERFORMANCE OF FORMANT EXTRACTION BY LP SPECTRUM AND HNGD SPECTRUM WITH MEAN ABSOLUTE DEVIATION (MAD)

Formants (Hz)				MAD		
Actual	LP	MGD	HNGD	LP	MGD	HNGD
400	346	338	424	54	62	24
800	859	864	825	59	64	25
1300	1425	1357	1323	125	57	23
2100	2280	2158	2163	180	58	63

a model based spectrum which depends on the order of the model. When small window is used, poor estimation of autocorrelation coefficients occurs which affects the LP coefficients. Here the LP method is able to resolve all the formants with order 12 as shown in Fig. 3 (c). The formant peaks in group delay method gets masked by the several large spurious spikes at different locations as shown in Fig. 3 (d). These spikes are due to zeros added by pitch peaks, noise and window effect. To remove the spikes from the group delay spectrum, the modified group delay spectrum (MGD) is used. Fig. 3 (e) shows that the modified group delay is able to resolve the formants. Among these methods, LP spectrum (order 12), modified group delay spectrum (MGD) and the HNGD spectrum methods shown in Fig. 3 (f) are able to resolve the closely spaced formants at frequencies 400, 800 and 1300 Hz without spurious peaks. Both the LP method and MGD, uses the 20-30 ms window size for formant extraction. These methods give poor resolution when the window size is decreased to 4 ms. The poor resolution gives the deviation of peaks from the actual formant location which degrades the accuracy in formant detection.

To quantify the accuracy of the formants captured from LP spectrum, MGD spectrum and the HNGD spectrum, we measured the mean absolute deviation (MAD) of the captured formants from the known formants. MAD is the difference between actual and captured formant frequencies. The peak picking method is used for the capturing the formant frequencies from all spectrum. Four higher peaks are considered four formants in case of HNGD spectrum. Table I shows the values of actual formants and formants captured from three different spectra and their corresponding MAD. The MAD is more in case of LP spectrum specially in higher frequency formants. These higher frequency formant deviation is reduced in case of MGD spectrum with nearly same deviation in lower frequency. The MAD is least in case of HNGD spectrum for first three formants. The HNGD spectrum thus gives better resolution to capture formant locations accurately.

To analyze the response of traditional methods, group delay method and HNGD method for the original speech, let us consider the normal vowel speech /ae/ having three formants around 660 Hz, 1720 Hz and 2410 Hz. [17]. The sampling frequency is 10 kHz. Fig. 4 shows the plots for the case of normal vowel speech /ae/. The magnitude spectrum in Fig. 4 (a) is flat, indicating poor resolution. When the sufficient number of low time liftering coefficients are taken, the cepstrum based spectrum is able to resolve the formants with some spurious peaks as shown in Fig. 4 (b). The LP spectrum in Fig. 4 (c) is 14th order spectrum. This higher order spectrum is able to resolve the formants but the spurious peaks are also detected as formants. Fig. 4 (d) shows spiky nature of group delay spectrum which is removed in modified group delay spectrum in Fig. 4(e). The HNGD spectrum in Fig. 4(f) gives prominent peaks at formant locations. Thus the HNGD method is model free method with high resolution which can give the vocal tract characteristics even within a pitch period.

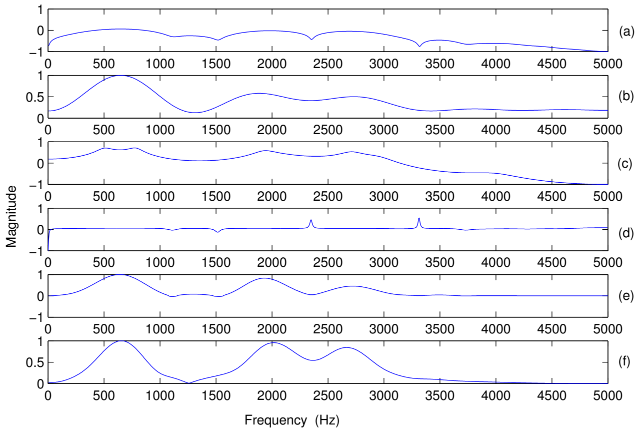

Fig. 4. Comparison of traditional methods, Group Delay method and HNGD method based on ZTW for the extraction of formants in case of normal vowel /ae/ with window size of 4 ms. (a) Magnitude spectrum (b) Cepstrum spectrum (c) LP spectrum (d) Group Delay spectrum (e) Modified Group Delay spectrum and (f) HNGD spectrum

The positioning of window relative to the sampling instant of the signal waveform also affects the signal spectrum. To overcome this effect, pitch synchronous analysis around glottal closure instants (GCIs) is generally used. But locating GCIs itself requires a specific method. The HNGD method uses the window size of approximately one pitch period which is highly decaying in nature. The energy of the windowed signal gets concentrated around the zero-time. This means we can find the vocal tract characteristics at each sampling instant. So HNGD method itself gives the vocal tract characteristics at the GCIs within each glottal cycle. The GCIs are high signal to noise ratio (SNR) instants. The HNGD spectrum at these instants are more prominent than the other instants.

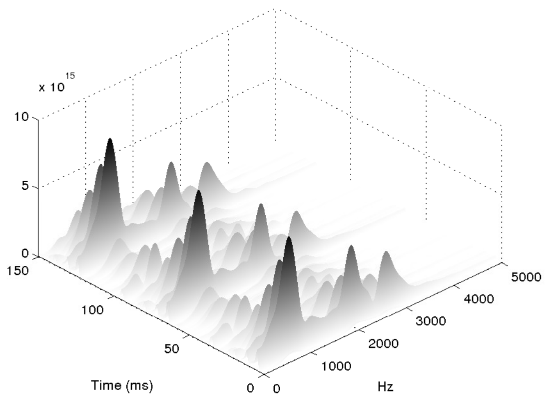

Fig. 5. HNGD spectrum at each instant within 3 pitch period

Fig. 5 shows the plot for the speech signal of three pitch periods. In each period, at one sampling instant, the HNGD spectrum is more prominent compared to other. These instants are GCIs. Because of high SNR, observing the HNGD spectrum at the GCI will give better peak for the hypernasality.

A. *Experimental Results and Discussion*

Experiments are performed to show the cue for hypernasality in CLP speech. The cue here is the addition of extra nasal formant in vicinity of first formant of the hypernasal vowels. The data used for the analysis are vowels /a/, /i/ and /u/ uttered by normal children and children having CLP. Data is sampled with a sampling frequency of 10 kHz. In the various experiments, the comparison between normal vowel sounds and hypernasal vowel sounds is shown in figures. The 3-D HNGD spectrum is plotted for each vowel sound at regular intervals around the GCIs.

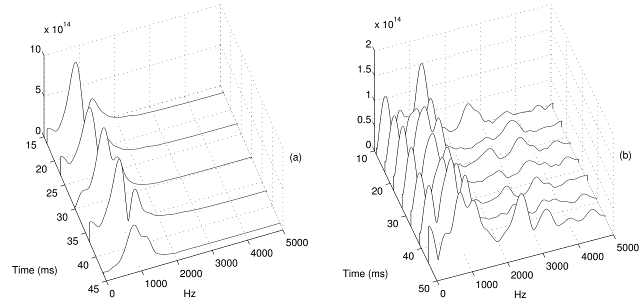

Fig. 6. Vowel sound /a/. (a) Normal vowel sound /a/ with first formant around 700 Hz. (b) Hypernasal vowel sound /a/ with first formant around 700 Hz and additional nasal formant below 500 Hz.

Fig. 6 shows the 3-D HNGD plot around GCI locations for the vowel sound /a/. Fig. 6 (a) shows the plot for the normal vowel sound. The first formant for the normal vowel /a/ lies around 700 Hz, which is shown as a dominant peak in HNGD spectrum. The normal vowels gets nasalized in case of CLP speech due to addition of nasal formant. In case of vowel /a/ this additional formant lies below 500 Hz. Fig. 6 (b) shows the plot for hypernasal vowel sound /a/. Figure shows the additional nasal formant in hypernasal vowel sound below 500 Hz and first formant around 700 Hz.

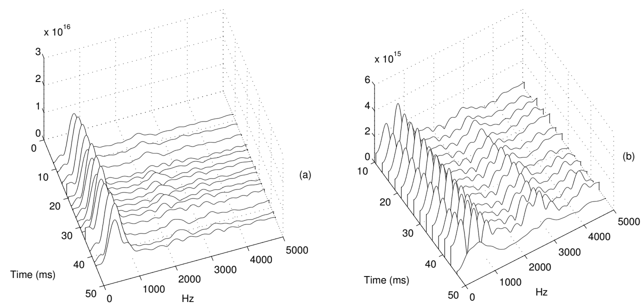

Fig. 7. Vowel sound /i/. (a) Normal vowel sound /i/ with first formant around 300 Hz. (b) Hypernasal vowel sound /i/ with first formant around 300 Hz and additional nasal formant around 1000 Hz.

In case of vowel sound /i/, the first formant lies around 300 Hz which is shown in Fig. 7 (a) as a dominant peak in HNGD spectrum. The addition of nasal formant in this case occurs around 1000 Hz. Fig. 7 (b) shows the first formant around 300 Hz and nasal formant around 1000 Hz for the hypernasal vowel sound /i/.

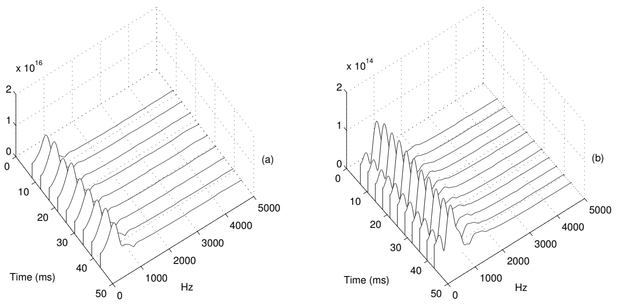

Fig. 8. Vowel sound /u/. (a) Normal vowel sound /i/ with first formant around 350 Hz. (b) Hypernasal vowel sound /u/ with first formant around 350 Hz and additional nasal formant around 800 Hz.

Similarly in case of vowel sound /u/, the first formant lies around 350 Hz which is shown in Fig. 8 (a) as a dominant peak in HNGD spectrum. The addition of nasal formant in this case occurs around 800 Hz. Fig. 8 (b) shows the first formant around 350 Hz and nasal formant around 800 Hz for the hypernasal vowel sound /u/. The additional nasal formants below 500 Hz in case of vowel /a/ and below 1000 Hz in case of vowels /i/ and /u/ is responsible for the nasalization of vowels and the HNGD method is well able to extract these formant and hence used for the detection of hypernasality.

V. SUMMARY, CONCLUSION AND SCOPE

In this paper we attempted to capture the changes in the vocal tract characteristics and hence the extra nasal formant from the hypernasal vowel speech of children having CLP. We used the impulse-like ZTW function. This high temporal resolution method gives accurate and resolves peak spectrum of hypernasal vowel sounds. The spectrum can be found at any sampling instant. The HNGD spectrum at each GCI gives better view for the analysis of hypernasality. Here we have used the additional nasal formant in the vicinity of first formant as the cue for the analysis of hypernasal speech. The other cues like the broadening of the formant bandwidth due to the addition of nasal formant and the presence of the anti-formants can also be used in future for the analysis and the detection of the hypernasal speech.

ACKNOWLEDGEMENTS

The authors would like to thank Prof Pushpavathi of AIISH Mysore for the mentoring and also sharing speech of CLP cases. This work is part of the ongoing project on development of software tools for the assessment of CLP speech.

REFERENCES

[1] K. Stevens, R. Nickerson, A. Boothroyd, and A. Rollins, "Assessment of nasalization in the speech of deaf children," *Journal of Speech, Language, and Hearing Research*, vol. 19, no. 2, pp. 393–416, 1976.

[2] G. Fant, *Acoustic theory of speech production*. The Hague, Netherlands: Mouton, 1960.

[3] S. Hawkins and K. N. Stevens, "Acoustic and perceptual correlates of the non-nasal-nasal distinction for vowels," *J. Acoust. Soc. Am.*, vol. 77, no. 4, pp. 1560–1574, Apr 1985.

[4] P. Vijayalakshmi, M. R. Reddy, and D. O'Shaughnessy, "Acoustic analysis and detection of hypernasality using a group delay function," *Biomedical Engineering, IEEE Transactions on*, vol. 54, no. 4, pp. 621–629, 2007.

[5] M. Y. Chen, "Acoustic parameters of nasalized vowels in hearing impaired and normal hearing speakers," *J. Acoust. Soc. Am.*, vol. 98, no. 5, pp. 2443–2453, Nov 1995.

[6] D. Cairns, J. H. Hansen, J. E. Riski *et al.*, "A noninvasive technique for detecting hypernasal speech using a nonlinear operator," *Biomedical Engineering, IEEE Transactions on*, vol. 43, no. 1, pp. 35–45, 1996.

[7] D. K. Rah, Y. I. Ko, C. Lee, and D. W. Kim, "A noninvasive estimation of hypernasality using a linear predictive model," *Annals of biomedical Engineering*, vol. 29, no. 7, pp. 587–594, 2001.

[8] A. Maier, F. Hönig, T. Bocklet, E. Nöth, F. Stelzle, E. Nkenke, and M. Schuster, "Automatic detection of articulation disorders in children with cleft lip and palate," *The Journal of the Acoustical Society of America*, vol. 126, no. 5, pp. 2589–2602, 2009.

[9] L. He, J. Zhang, Q. Liu, H. Yin, and M. Lech, "Automatic evaluation of hypernasality and consonant misarticulation in cleft palate speech," *Signal Processing Letters, IEEE*, vol. 21, no. 10, pp. 1298–1301, 2014.

[10] M. Anand Joseph, S. Guruprasad, and B. Yegnanarayana, "Extracting formants from short segments of speech using group delay functions," 2006.

[11] B. Yegnanarayana, D. Saikia, and T. Krishnan, "Significance of group delay functions in signal reconstruction from spectral magnitude or phase," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 32, no. 3, pp. 610–623, Jun 1984.

[12] Y. Bayya and D. N. Gowda, "Spectro-temporal analysis of speech signals using zero-time windowing and group delay function," *Speech Communication*, vol. 55, no. 6, pp. 782–795, 2013.

[13] K. Murty and B. Yegnanarayana, "Epoch extraction from speech signals," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 16, no. 8, pp. 1602–1613, 2008.

[14] B. Yegnanarayana, H. Murthy *et al.*, "Significance of group delay functions in spectrum estimation," *Signal Processing, IEEE Transactions on*, vol. 40, no. 9, pp. 2281–2289, 1992.

[15] L. Rabiner and B. H. Juang, "Fundamentals of speech recognition," 1993.

[16] B. Yegnanarayana, "Formant extraction from linear-prediction phase spectra," *The Journal of the Acoustical Society of America*, vol. 63, no. 5, pp. 1638–1640, 1978.

[17] G. E. Peterson and H. L. Barney, "Control methods used in a study of the vowels," *J. Acoust. Soc. Am.*, vol. 24, no. 2, pp. 175–194, 1952.