INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

Hypernasality Severity detection using Constant Q Cepstral Coefficients

Akhilesh Kumar Dubey, S. R. Mahadeva Prasanna, S. Dandapat

[1]Indian Institute of Technology Guwahati, Guwahati, India
[2]Indian Institute of Technology Dharwad, Dharwad, India

(d.akhilesh, prasanna, samaren)@iitg.ernet.in

Abstract

In this work, detection of hypernasality severity in cleft palate speech is attempted using constant Q cepstral coefficients (CQCC) feature. The coupling of nasal tract with the oral tract during the production of hypernasal speech adds nasal formants and anti-formants in low frequency region of vowel spectrum mainly around the first formant. The strength and position of nasal formants and anti-formants along with the oral formants changes as the severity of nasality changes in hypernasal speech. The CQCC feature is extracted from the constant Q transform (CQT) spectrum which employs geometrically spaced frequency bins and maintains a constant Q factor for across the entire spectrum. This results in a higher frequency resolution at lower frequencies and higher temporal resolution at higher frequencies. The CQT spectrum resolves the nasal and oral formants in low frequency and captures the spectral changes due to change in nasality severity. The CQCC feature gives the overall classification accuracy of 83.33 % and 78.47 % for /i/ and /u/ vowels corresponding to normal, mild and moderate-severe hypernasal speech, respectively using multi-class support vector classifier.

Index Terms: Hypernasality, constant Q cepstral coefficients, constant Q transform, cleft palate.

1. Introduction

The cleft palate (CP) is a congenital craniofacial disorder. The speech of cleft palate (CP) children get affected due to structural abnormalities, inadequate functioning of velopharyngeal port and mis-learning [1]. The CP speech is universally reported in terms of presence or absence of resonance disorder, nasal air emission and/or turbulence, consonant production errors and voice disorder [2]. Hypernasality is an important disorder belonging to the resonance disorder category where excess nasality is heard in the speech due to the coupling of nasal tract with the oral tract during the production of voice sounds, especially vowels. The plastic surgery done by the surgeons to correct the structural abnormalities may not be enough to restrict the coupling of nasal tract due to velopharyngeal insufficiency and mis-learning [3]. Hence, nasality remains present in repaired CP children. The intelligibility of CP speech gets reduced due to hypernasality. The evaluation of hypernasality helps plastic surgeons and speech-language pathologists (SLPs) in the proper diagnosis of CP children.

In a clinical environment, the hypernasality evaluation is done perceptually by expert SLPs and the decision of perceptual evaluation is confirmed by some instrumental method. The confirmation is done because the perceptual decision may sometimes vary among the SLPs [4] due the abnormalities in pitch, loudness, voice quality and/or articulation occurring in conjunction with hypernasality [5] affect the perception of nasality in hypernasal speech [6]. The instrumental method of evaluation can be categorized into direct and indirect method [7]. The X-Ray (Cephalometry), videofluoroscopy, nasendoscopy, accelerometry, and nasometry are some important instrumental method of hypernasality evaluation. But the limitations of these techniques like radiation effect, invasiveness, requirement of addition sensing device and inability of nasometer to provide nasality score for prerecorded speech data, motivates the researcher to propose a noninvasive, effective, objective method of hypernasality evaluation based on the acoustic analysis of speech signal.

In literature, several hypernasality detection works have been done by the researchers by spectral analyses of /a/, /i/ and /u/ vowels present in the speech. These works are based on the acoustic cues for nasalized vowels proposed in the literature. The presence of nasal formants in low-frequency region around the first formant (F_1), reduction in strength of F_1 and hence broadening of first formant due to the presence of nasal anti-formants and flattening of overall spectrum are some important acoustic cues for nasalized vowels [8],[9],[10]. The important works on hypernasality detection are based on Teager energy operator (TEO) based feature [11], TEO feature with frequency cepstral coefficient (MFCC) feature [12], pitch-adaptive MFCC feature [13], linear prediction cepstral coefficient (LPCC) feature [14] and features extracted from high spectral resolution group delay spectrum [4] and zero time windowing technique [15], [16]. Besides that the feature set obtained from acoustic, noise and cepstral analysis, nonlinear dynamic and entropy measurements [17], [18], [19], based on energy distribution [20], [21] and using vowel space area (VSA) [22] are also used. The hypernasality detection is also done using recorded sentences speech database using jitter, shimmer, MFCC, bionic wavelet transform entropy and bionic wavelet transform energy features.[23].

All the above works on hypernasality detection give good accuracy ranging from the 70% to 90% for the normal and hypernasal speech classification. However, doctors and SLPs are more interested in severity grading of hypernasal speech onto a 4-point scale because it gives the information about the velopharyngeal gap size in children with CP. The 4-point scale is described as 0 = nasality within normal limits (Normal speech), 1 = mild hypernasality, 2 = moderate hypernasality, and 3 = severe hypernasality [2]. Besides high importance, very few works like the relation between spectral characteristics and perceived hypernasality using one-third octave spectra [24], hypernasality severity detection using formant feature and Gaussian Mixture Model (GMM) [21] and hypernasality severity analysis using zero time windowing [25] are done for the severity grading of hypernasality.

In this work, detection of hypernasality severity in high vowels /i/ and /u/ corresponding to normal, mild and moderate-severe hypernasal speech is attempted using constant Q cepstral coefficients (CQCC) feature. As per the guideline given

http://dx.doi.org/10.21437/Interspeech.2019-2151

Table 1: *Description of database in terms of number of normal, mild and moderate-severe hypernasal stimuli recorded and the number of /i/ and /u/ phonemes*

Stimulus	Number of stimuli			Number of phonemes		
	Normal	Mild	Moderate-severe	Normal	mild	Moderate-severe
/pipi/	190	180	186	380	360	372
/pupu/	166	162	190	332	324	380

in [2] regarding the speech samples for severity rating hypernasal speech, only high vowels are considered in this work. The CQCC feature is extracted from the spectrum obtained from perceptually motivated time-frequency analysis called as the constant Q transform (CQT). The transform employs geometrically spaced frequency bins which ensures a constant Q factor across the entire spectrum. This results in high resolution at lower frequencies along with high temporal resolution at higher frequencies. The CQT is contrary to the Fourier transform which imposes regular spaced frequency bins. The spectral analysis of hypernasal speech shows the presence of nasal formants and anti-formants in the vicinity of F_1. The strength and position of nasal formants and anti-formants along with oral formants varies as the severity increases from normal to mild to moderate-severe. The high resolution at lower frequencies of CQT based spectrum resolves the nasal and oral formants, and captures the spectral changes present in normal, mild and moderate-severe hypernasal speech. Hence the CQCC feature may capture the nasality severity information in a better way, which may enhance the accuracy of hypernasality severity detection. The multi-class support vector machine (SVM) classifier is used for severity detection.

The rest of the paper is organized as follows. In Section 2, gives the description about the database. In Section 3, spectral analysis of hypernasal speech is explained. Section 4 describes the constant Q transform. Section 5 gives the results and discussion and finally section 6 contains the summary and conclusion of the work.

2. Speech database

In this work, data is collected from three groups of children. The first group is the control normal (CN) group containing 15 children with normal speech. The second and third groups are CP groups each containing 15 children with repaired CP and having respectively, mild and moderate-severe hypernasality in their speech. Out of 15 children of each group, 9 are boys and 6 are girls. None of the children who participated in data collection has any history of hearing impairment disorder. The age range of children lie between 7-12 years. The data is recorded in the sound-treated room of All Indian Institute of Speech and Hearing (AIISH), Mysore, India [26] using Bruel & Kjaer sound level meter (SLM) microphone. The native language of all children is Kannada, hence the data is recorded in the Kannada language which is a Dravidian language spoken in the southern part of India. During the time of recording, the instructor first utters the word and then the child repeats the same. The recording is done at sampling frequency 44.1 kHz, 16 bps in .WAV format, which is down-samples at 16 kHz for the analysis. The stimuli considered in this work are words /pipi/ and /pupu/ in which the vowel /i/ and /u/ immediately follow the pressure consonant /p/. The stimuli were designed by SLPs of AIISH as per the suggestion given in [2]. Table 1 shows the number of normal, mild and moderate-severe hypernasal stimuli recorded and the number of /i/ and /u/ phonemes annotated in the stimuli. The annotation of vowel regions in stimuli is done by SLPs using Wavesurfer tool [27].

Table 2: *Intra rater reliability estimation*

Vowel	Pair of raters	Cohen's kappa	Correlation coefficient
/i/	1 and 2	0.75	0.78
	2 and 3	0.67	0.69
	1 and 3	0.74	0.76
/u/	1 and 2	0.72	0.75
	2 and 3	0.68	0.70
	1 and 3	0.73	0.75

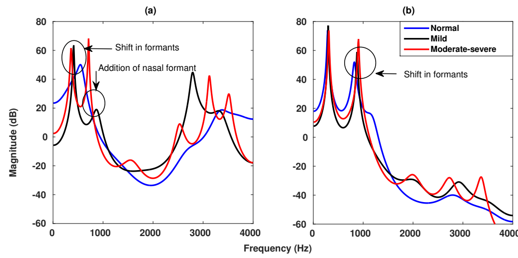

Figure 1: *LP magnitude spectra of /i/ and /u/ vowels for normal, mild, and moderate-severe hypernsality speech. (a) is for /i/ vowel and (b) is for /u/ vowel*

To assess the hypernasality severity of each stimulus from CP group, the perceptual test is conducted by the three SLPs from the AIISH. The SLPs are highly trained and having the experience of around five years in the field of CP speech evaluation. The stimuli are randomized in order and SLPs are asked to rate the hypernasality in words using the 4-point severity rating scale. The severity rating agreement between each pair of SLP is compared using Cohens kappa and Spearmans rank correlation coefficient which is shown in Table 2.

3. Spectral analysis of hypernasal speech

The vowel spectrum gets affected in hypernasal speech due to the addition of nasal formant and anti-formant pairs at the natural frequencies of the nasal tract and the sinuses. The natural frequency of nasal tract lies in the frequency range of 450 to 650 Hz and 1800 to 2400 Hz [28], whereas it lies around 400 Hz and 1300 Hz for the sinuses [29]. For /i/ vowel where F_1 is present below 500 Hz and F_2 above 2000 Hz, the nasal formant around 400 Hz enhances the strength of F_1 and the formant around 450 to 650 Hz gives additional formant between F_1 and F_2. The strength of F_1 and additional formant get enhanced, and their frequency location also gets shifted as the severity of nasality increases. For /u/ vowel where both F_1 and F_2 are below 1000 Hz, the nasal formants enhances the strength of both the formants and shift their frequency locations as the severity of nasality increases. Fig. 1 shows the spectral changes in hypernasal speech as the nasality of severity increases from normal to mild to moderate-severe. Fig. 1 (a) is shown for /i/ vowel and (b) is for /u/ vowels. The enhancement in strength of F_1, additional nasal formant and F_2 along with the shifting in formants frequencies as the severity of nasality increases in hypernasal speech can be observed from the Fig. 1. Hence this analysis shows that the spectral characteristics of hypernasal speech changes as the severity of nasality increases.

4. Constant Q transform

The CQT is a perceptually motivated time-frequency analysis. It first introduced by Youngberg [30] and Boll is refined by Brown [31] for music signal processing. In this time-frequency approach, the octaves and central frequencies of each filter are geometrically distributed. The approach gives a higher frequency resolution for lower frequencies and a higher temporal resolution for higher frequencies and this is in contrast to the

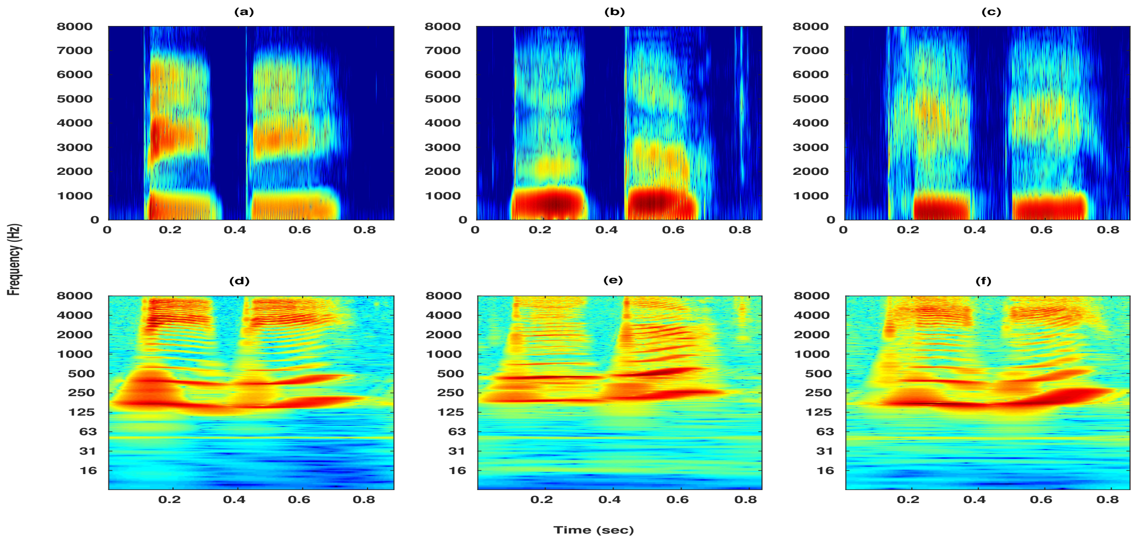

Figure 2: *Comparison of CQT spectrogram with the STFT spectrogram for vowel /i/. (a)-(c) show the STFT spectrograms of normal, mild and moderate-severe hypernasal speech respectively whereas the (d)-(e) show for the CQT spectrograms.*

fixed time-frequency resolution of short-term Fourier transform (STFT). In STFT the bandwidth of each filter is constant, hence the Q factor increases as the central frequency increases from low to high frequencies. The STFT lacks high frequency resolution at lower frequencies and high temporal resolution at higher frequencies. The CQT has been used widely for the analysis, classification and separation of audio signals. [32], [33].

4.1. Computation of CQT

The CQT of a discrete time signal $x(n)$ is denoted by $X^{CQ}(k,n)$ and it is defined as [34]

$$X^{CQ}(k,n) = \sum_{j=n-\lfloor N_{k/2} \rfloor}^{n+\lfloor N_{k/2} \rfloor} x(j) a_k^*(j - n + N_{k/2}) \qquad (1)$$

where $k = 1, 2, ..., K$ is the frequency bin index, N_k are variable window size and a_k^* is the complex conjugate of $a_k(n)$ which is given by

$$a_k(n) = \frac{1}{C}(\frac{n}{N_k}) exp[i(2\pi n \frac{f_k}{fs} + \Phi_k)] \qquad (2)$$

where f_k is the central frequency of the k^{th} bin, f_s is the sampling frequency, $\omega(t)$ is the window function, Φ_k is a phase offset and C is scaling factor given by:

$$C = \sum_{l=-\lfloor N_{k/2} \rfloor}^{\lfloor N_{k/2} \rfloor} \omega(\frac{l + N_k/2}{N_k}) \qquad (3)$$

The central frequencies f_k are given by $f_k = f_1 2^{\frac{k-1}{B}}$, where f_1 is the central frequency of lowest-frequency bin and B is the number of bins per octave. The Q factor is given by

$$Q = \frac{f_k}{f_{k+1} - f_k} = (2^{\frac{1}{B}} - 1)^{-1} \qquad (4)$$

The window lengths N_k is given by $N_k = \frac{fs}{f_k} Q$. Further a parameter $\gamma = \Gamma = 228.7 * (2^{\frac{1}{B}} - 2^{\frac{-1}{B}})$ is added at which the bandwidths equal a constant fraction of the ERB critical bandwidth.

4.2. Hypernasality severity analysis using CQT spectrum

Fig. 2 compares the spectrograms obtained from STFT (top) and CQT (bottom) for normal (a), mild (b) and moderate-severe (c) hypernasal speech. It can be observed from Fig. 2 that the energy in hypernasal speech is mainly present in low frequency region and it increases as the severity of nasality increases. This is due to the addition of nasal formants in low frequency region whose strength increases with the increase of nasality severity. It can also be observed that the STFT spectrogram is unable to resolve the nasal and oral formants in low frequency due to its resolution limitation where the CQT spectrogram can resolve them due to its higher resolution at lower frequencies. Hence the CQCC feature extracted from the CQT spectrum can effectively capture the nasality evidence in normal, mild and moderate-severe hypernasal speech.

4.3. Constant Q cepstral coefficients

Fig. 3 shows the steps of CQCC extraction from the signal $x(n)$ [34]. Unlike the traditional way where the cepstral coefficients are computed from the logarithmic magnitude square spectrum, the CQCC cannot be directly computed from it. This is because k bins in $X^{CQ}(k)$ are geometrically spaced which should be covered to linear space to take the discrete cosine transform (DCT). This scale conversion is done by the uniform resampling step shown in Fig. 3. In this step, downsampling operation over the first k low frequency bins and upsampling operation for the remaining high frequency $K - k$ bins are performed. Let the distance between f_k and $f_1 = f_{min}$ is defined as:

$$\Delta f^{k \leftrightarrow 1} = f_k - f_1 = f_1(2^{\frac{k-1}{B}} - 1) \qquad (5)$$

where $k = 1, 2, ..., K$ is the frequency bin index. For linear sampling, consider a period T_l which is equivalent to determine a value of $k_l \in 1, 2, ..., K$ such that $T_l = \Delta f^{k \leftrightarrow 1}$, and it can be obtained by splitting the first octave into d equal parts with period T_l to get K_l as:

$$\frac{f_1}{d} = f_1(2^{\frac{k-1}{B}} - 1) \rightarrow k_l = Blog_2(1 + \frac{1}{d}) \qquad (6)$$

Thus the new frequency rate is given by:

$$F_l = \frac{1}{T_l} = [f_l(2^{\frac{k_l - 1}{B}} - 1)]^{-1} \qquad (7)$$

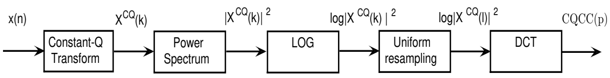

Figure 3: *Block diagram showing the steps of CQCC feature extraction*

Table 3: *Hypernasality severity accuracy for /i/ vowel*

Feature	Accuracy (%)	Confusion Matrix			
CQCC	83.33	Hypernasality	Normal	Mild	Moderate-severe
		Normal	92.19	3.13	3.69
		Mild	9.09	65.91	25.00
		Moderate-severe	2.08	10.42	87.50
MFCC	76.28	Hypernasality	Normal	Mild	Moderate-severe
		Normal	95.31	1.56	3.13
		Mild	40.91	36.36	22.73
		Moderate-severe	2.08	10.42	87.50
formant	46.15	Hypernasality	Normal	Mild	Moderate-severe
		Normal	60.94	7.81	31.25
		Mild	70.45	2.27	27.27
		Moderate-severe	29.17	4.17	66.67

The new frequency rate contains d uniform samples in the first octave, $2d$ in the second and $2^j d$ in the $(j-1)^{th}$ octave. The CQCC can now be extracted from the uniformed sampled spectrum as:

$$CQCC(p) = \sum_{l=1}^{L} log \left| X^{CQ}(l) \right|^2 cos \left[\frac{p(l-\frac{1}{2})\pi}{L} \right] \quad (8)$$

where $p = 0, 1, ..., L1$ and l are the newly resampled frequency bins.

5. Hypernasality severity detection using constant Q cepstral coefficients feature

In this section, the hypernasality severity detection is performed using CQCC feature and the result is compared with the result obtained from baseline MFCC feature and formant feature. The formant feature is proposed in [21], where F_1 and the number of formants in the spectrum is used as the feature.

5.1. Experimental setup

The 13-dimensional MFCC and 2-dimensional format features are extracted for each frame of speech. The frame size of 20 ms and the frame shift of 10 ms is used for the framing of the speech. For applying the CQT, the maximum frequency of $F_{max} = F_s/2$, where $F_s = 8kHz$, the minimum frequency $F_{min} = F_{max}/2 \simeq 15Hz$, $B = 96$, $\gamma = \Gamma$ and $d = 16$ are taken and 19-dimensional CQCC feature is computed. The SVM classifier is accomplished for multi-class classification using one-versus-one strategy with radial basis kernel (RBF). The 5-fold cross validation of the entire train database is done to find the optimum value of the kernel parameters c and γ. Randomly selected 12 normal, 12 mild, and 12 moderate-severe children data are used for SVM training and remaining 3 normal, 3 mild, and 3 moderate-severe children data for is used for testing.

Table 4: *Hypernasality severity accuracy for /u/ vowel*

Feature	Accuracy (%)	Confusion Matrix			
CQCC	78.47	Hypernasality	Normal	Mild	Moderate-severe
		Normal	82.00	8.00	10.00
		Mild	20.00	65.00	15.00
		Moderate-severe	11.11	3.70	85.19
MFCC	72.22	Hypernasality	Normal	Mild	Moderate-severe
		Normal	74.00	14.00	12.00
		Mild	17.50	60.00	22.00
		Moderate-severe	12.96	7.41	79.63
formant	65.97	Hypernasality	Normal	Mild	Moderate-severe
		Normal	90.00	2.00	8.00
		Mild	32.50	22.50	45.00
		Moderate-severe	16.67	7.41	75.92

5.2. Result

Table 3 and Table 4 show performance result of hypernasality severity detection for /i/ and /u/ vowel respectively. The performance is presented at the phoneme level and it is in terms of the overall accuracy and confusion matrix. Phoneme level results are derived from the frame level results. It is done by using the class labels given by SVM classifier on majority basis i.e. a phoneme will belong to a particular class if majority of its frames belong to that particular class. The individual accuracies for CQCC, MFCC and formant features are shown in each table. It can be observed from the result tables that CQCC feature gives an accuracy of 83.33 % and 78.47 % for vowels /i/ and /u/, respectively which is better than the accuracy obtained from baseline MFCC and formant features. The low accuracy for baseline features is due to misclassification of mild hypernasal speech. The spectral characteristics of mild hypernasal speech lie between two stream levels normal and moderate-severe, and it has a very minute spectral difference in terms of strength and position of formants with these stream levels. This difference is not captured by the MFCC and formant feature whereas the CQCC feature due to its high spectral resolution is able to differentiate between three levels of hypernasality.

6. Summary and Future scope

In this work hypernasality severity detection of /i/ and /u/ vowels corresponding to normal, mild and moderate-severe hypernasal speech is performed using CQCC feature. The severity detection is main concern of doctors and SLPs for the treatment of children with CP because it gives information about the velopharyngeal gap size. The spectral analysis of hypernasal speech shows the presence of nasal formants and anti-formants in the vicinity of F_1. The strength and position of nasal formants and anti-formants vary as the severity increases from mild to moderate-severe. The CQCC feature is extracted from perceptually motivated constant Q transform (CQT) spectrum which employs geometrically spaced frequency bins. This ensures a constant Q factor across the entire spectrum, and hence high resolution at lower frequencies is obtained along with high temporal resolution at higher frequencies. The high resolution at lower frequencies of CQT based spectrum resolves the nasal and oral formants, and captures the spectral changes present in mild or moderate-severe hypernasal speech. The results show that the CQCC feature captures the nasality severity information in a better way and hence the accuracy of hypernasality severity detection increases compared to MFCC and formant features. The confusion matrix of the result shows that mainly mild hypernasal speech is misclassified, which can be improved as a part of future work. Further, the feature can be explored on sentence database for the detection and severity grading of hypernasal speech.

7. Acknowledgements

The authors would like to thank Prof. M. Pushpavathi and Prof. Ajish K. Abraham of AIISH Mysore for the mentoring and also sharing speech of CLP cases. This work is in part supported by the project grants, for the projects entitled NASOSPEECH: Development of Diagnostic system for Severity Assessment of the Disordered Speech funded by the Department of Biotechnology (DBT), Govt. of India and ARTICULATE +: A system for automated assessment and rehabilitation of persons with articulation disorders funded by the Ministry of Human Resource Development (MHRD), Govt. of India.

8. References

[1] D. Sell, A. Harding, and P. Grunwell, "A screening assessment of cleft palate speech (great ormond street speech assessment)," *International Journal of Language & Communication Disorders*, vol. 29, no. 1, pp. 1–15, 1994.

[2] G. Henningsson, D. P. Kuehn, D. Sell, T. Sweeney, J. E. Trost-Cardamone, and T. L. Whitehill, "Universal parameters for reporting speech outcomes in individuals with cleft palate," *The Cleft Palate-Craniofacial Journal*, vol. 45, no. 1, pp. 1–17, 2008.

[3] A. W. Kummer and L. Lee, "Evaluation and treatment of resonance disorders," *Language, Speech, and Hearing Services in Schools*, vol. 27, no. 3, pp. 271–281, 1996.

[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] D. C. Spriestersbach, "Assessing nasal quality in cleft palate speech of children," *Journal of Speech and Hearing Disorders*, vol. 20, no. 3, pp. 266–270, 1955.

[6] 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.

[7] K. Bettens, F. L. Wuyts, and K. M. Van Lierde, "Instrumental assessment of velopharyngeal function and resonance: A review," *Journal of communication disorders*, vol. 52, pp. 170–183, 2014.

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

[9] 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.

[10] T. Pruthi and C. Y. Espy-Wilson, "Acoustic parameters for the automatic detection of vowel nasalization," in *Eighth Annual Conference of the International Speech Communication Association*, 2007.

[11] 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.

[12] 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.

[13] A. K. Dubey, S. M. Prasanna, and S. Dandapat, "Pitch-adaptive front-end feature for hypernasality detection," *Proc. Interspeech 2018*, pp. 372–376, 2018.

[14] 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.

[15] A. K. Dubey, S. M. Prasanna, and S. Dandapat, "Zero time windowing analysis of hypernasality in speech of cleft lip and palate children," in *IEEE Twenty Second National Conference on communication (NCC)*, 2016, pp. 1–6.

[16] A. K. Dubey, S. Prasanna, and S. Dandapat, "Hypernasality detection using zero time windowing," in *2018 International Conference on Signal Processing and Communications (SPCOM)*. IEEE, 2018, pp. 105–109.

[17] J. R. Orozco-Arroyave, S. M. Rendón, A. M. Álvarez-Meza, J. D. Arias-Londoño, E. Delgado-Trejos, J. F. V. Bonilla, and C. G. Castellanos-Domínguez, "Automatic selection of acoustic and non-linear dynamic features in voice signals for hypernasality detection." in *Interspeech*. Citeseer, 2011, pp. 529–532.

[18] S. M. Rendón, J. O. Arroyave, J. V. Bonilla, J. A. Londoño, and C. C. Domínguez, "Automatic detection of hypernasality in children," in *International Work-Conference on the Interplay Between Natural and Artificial Computation*. Springer, 2011, pp. 167–174.

[19] J. R. Orozco-Arroyave, J. D. Arias-Londoño, J. F. V. Bonilla, and E. Nöth, "Automatic detection of hypernasal speech signals using nonlinear and entropy measurements." in *INTERSPEECH*, 2012, pp. 2029–2032.

[20] G.-S. Lee, C.-P. Wang, C. C. Yang, and T. B. Kuo, "Voice low tone to high tone ratio: a potential quantitative index for vowel [a:] and its nasalization," *IEEE transactions on biomedical engineering*, vol. 53, no. 7, pp. 1437–1439, 2006.

[21] 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.

[22] A. K. Dubey, A. Tripathi, S. Prasanna, and S. Dandapat, "Detection of hypernasality based on vowel space area," *The Journal of the Acoustical Society of America*, vol. 143, no. 5, pp. EL412–EL417, 2018.

[23] M. Golabbakhsh, F. Abnavi, M. Kadkhodaei Elyaderani, F. Derakhshandeh, F. Khanlar, P. Rong, and D. P. Kuehn, "Automatic identification of hypernasality in normal and cleft lip and palate patients with acoustic analysis of speech," *The Journal of the Acoustical Society of America*, vol. 141, no. 2, pp. 929–935, 2017.

[24] R. Kataoka, D. W. Warren, D. J. Zajac, R. Mayo, and R. W. Lutz, "The relationship between spectral characteristics and perceived hypernasality in children," *The Journal of the Acoustical Society of America*, vol. 109, no. 5, pp. 2181–2189, 2001.

[25] A. Dubey, S. M. Prasanna, and S. Dandapat, "Zero time windowing based severity analysis of hypernasal speech," in *IEEE Region 10 Conference (TENCON)*, 2016, pp. 970–974.

[26] AIISH, "All india institute of speech and hearing, mysore, india." [Online]. Available: web- site: http : //www.aiishmysore.in

[27] K. Sjölander and J. Beskow, "Wavesurfer-an open source speech tool," in *Sixth International Conference on Spoken Language Processing*, 2000.

[28] K. N. Stevens, *Acoustic phonetics*. MIT press, 2000, vol. 30.

[29] S. Maeda, "The role of the sinus cavities in the production of nasal vowels," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 7, 1982, pp. 911–914.

[30] J. Youngberg and S. Boll, "Constant-q signal analysis and synthesis," in *ICASSP'78. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3. IEEE, 1978, pp. 375–378.

[31] J. C. Brown, "Calculation of a constant q spectral transform," *The Journal of the Acoustical Society of America*, vol. 89, no. 1, pp. 425–434, 1991.

[32] C. Schörkhuber, A. Klapuri, and A. Sontacchi, "Audio pitch shifting using the constant-q transform," *Journal of the Audio Engineering Society*, vol. 61, no. 7/8, pp. 562–572, 2013.

[33] R. Jaiswal, D. Fitzgerald, E. Coyle, and S. Rickard, "Towards shifted nmf for improved monaural separation," 2013.

[34] M. Todisco, H. Delgado, and N. Evans, "A new feature for automatic speaker verification anti-spoofing: Constant q cepstral coefficients," in *Speaker Odyssey Workshop, Bilbao, Spain*, vol. 25, 2016, pp. 249–252.