Effect of Inter-Speaker Hypernasality Variation in Cleft Palate on Acoustic Vowel Space

Shahina Haque[1], Md. Hanif Ali[2] and A. K. M. Fazlul Haque[3]

[1,2]Department of CSE, Jahangirnagar University, Savar, Dhaka, Bangladesh.
[1]Department of GED, Daffodil International University, Dhaka, Bangladesh
[3]Department of ETE, Daffodil International University, Dhaka, Bangladesh

Abstract— **Vocal tract transfer function of voiced hypernasal speech of cleft palate (CP) speakers contains information related to hypernasality (HP). The acoustic sign of HP is profoundly variable making HP detection more challenging for human trained professionals. Vowels /i/, /a/ and /u/ uttered by adults (male, female) and children with increasingly HP in read speech sentences are used in this study. In the first half part, variability in acoustic parameters among and across vowel class are made. Test results show significant fluctuation in acoustic parameters among CP speakers. Variability of spectral parameters came out to be maximum from the experimental result for /u/ showing that it is the most affected vowel for HP. For children it is least variable and for females it shows most variable value. Variability of /i/, /a/ are seen to be greatest for children. Among the vowels, /a/ which is mid low vowel, shows minimum variation for females and high front vowel /i/ shows least variation for males. In the later part of our study, a threshold-based parameter is used for HP detection. Vowel space area (VSA) of healthy speech and hypernasal speech is used as a measure to detect HP. It is observed from the result that VSA traversed by CP children is reduced by 4.15 times and CP males is shrunken by 3.7 times with respect to VSA of CP females. VSA of CP females are observed to have the greatest area indicating that speech of CP female has better intelligibility than male or children. VSA of CP children producing HP are observed to have the lowest value indicating that speech clarity of CP children decreases more than the male and female group.**

Keywords—Hypernasality; cleft palate; acoustic feature; continuous read speech; vowel space area

I. INTRODUCTION

Hypernasality (HP) is a perceptual quality that may result from structural abnormalities of defective velopharyngeal (VP) mechanism which produces excessive amount of nasal resonance during speech production reducing speech intelligibility [1]. Worldwide, HP produced by CP speaker is the second highest abnormality [2]. HP has similarity with vowel nasality and is reflected in the vocal tract transfer function of the spectrum of vowels as shown in Fig. 1. HP detection assists physicians to take decision whether speech therapy or surgery is to be given to the CP speakers. Usually, speech language pathologists make the decision for HP detection which is subjective and is the standard technique for assessing hypernasality [3]. Detection of HP using speech signal processing techniques is becoming popular due to non-invasive nature, simplicity, objectivity, and precisions for the investigation of CP speakers.

HP detection has been the topic of interest of many researchers [4-14, 29]. It is being studied by analysis of hypernasal speech,

synthesizing hypernasal speech and comparing it with nasal or nasalized vowel. Cairns used Teager energy operator for studying HP [15]. Rah used spectral zero detection for detecting HP [16]. From previous study it is observed that acoustic symptoms of hypernasality may be several, the formant bandwidth may broaden, additional multiple poles and zeros may be present in the spectrum, vowel formants may shift. Perceptual experiment using formants in the spectrum was carried out and was observed that formant at 250Hz plays the most effective role for nasalization [17-18]. In [19] group delay function was used for detecting HP successfully.

Shifts of oral vowel formants due to HP affects VSA and HP related information can be captured in VSA. Using continuous read speech, in study for CP males and female subjects, VSA was used for HP detection and to study the
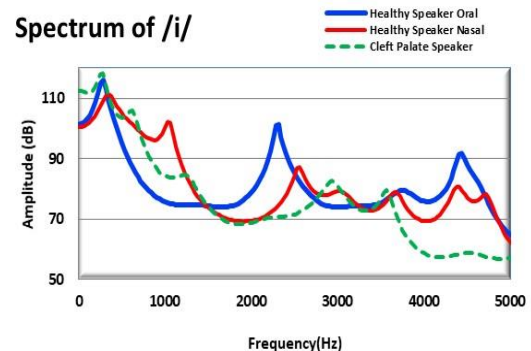

Fig. 1. Difference in Vocal tract transfer function of vowel /i/ for healthy oral, healthy nasal and Hypernasal speech

variation of vocal tract parameters [9-10]. VSA is a space that can map articulatory movement and acoustic properties while vowels are produced. It is formed by combining first formant frequency, F1 and second formant frequency, F2 for all vowels [20]. Mouth opening is equivalent to F1 value and tongue height to F2 value. Therefore, VSA can track the development of speech articulators. For clear intelligible speech, VSA has a larger value than speech with less intelligibility. VSA is also used for studying the characteristics of the speaker, peculiarity and personality of vowels [21-23]. Computing VSA can also act as an indicator for abnormal speech production, development of speech. In [24], VSA was automatically detected and shown to have better performance than usual method. In [25], it is showed that VSA reduces due to mental stress. An objective measure was proposed in [26] by analyzing hypernasal vowel near plosive. Avaregae falling and rising amplitudes were found to be less in nasal vowel than oral

vowels. VSA was found to be drastically reduced for cerebral paralysis in [27] which studied intelligibility and speech production variability. [28] studied the repaired surgery speech and compared it with spectral enhancement hypernasal speech. Spectrally enhanced speech was found to have better intelligibility. Recently, an objective HP measure method is proposed by training a deep neural network (DNN) with normal speech. Without using any clinical data this method can detect HP [29].

As there are differences in spectral characteristics between children and adults, there are not many studies on the detection of HP using a mixed dataset and to compare difference in speech parameters for HP assessment induced by abnormal VP opening in continuous read speech expression (which shows greater complexities) of males, females and children. This formed the inspiration for this research.

Deviation in speech articulation due to HP as compared to the healthy speech can be measured by variability of acoustic features of speech. Variability gives a measure of how dispersed a measurement is around the average value. Studying variability of speech features of CP speech is important for applications in many fields of speech technology such as automatic speech recognition (ASR) systems. Challenges in any speech, speaker or speech processing system is the ability to normalize and minimize the inter-speaker variation for training, classification [30]. In this study, we observed how useful speech parameters extracted from continuous read speech of children and adults can be to give us various information regarding HP. In the first part of this study, we are concerned with a comparative study of inter-speaker variability measurement in CP male, female, and children speech with varying HP. Second half of this study is concerned with comparing VSA of CP male, female and children speech and assessment of HP using VSA.

Resemblance of nasality and HP in production can be observed through acoustic features. Most studies compare coarticulatory nasalization and compares with HP. But we used Bangla nasal speech data which is phonemic to compare with HP.

This paper is organized as given below:

Description of speech materials used is given in section II, acoustic analysis procedure is given in section III, results and discussion are given in section IV. Conclusion is given in section V.

II. SPEECH MATERIALS

From the literature it is observed that the corner vowels /i/, /a/ and /u/ are HP sensitive vowels. Therefore, these vowels are selected for this study. Non-CP healthy speech corpus of male, female and children for Bangla and English consists of recording selected isolated Bangla oral and nasal vowels pronounced three times in a quiet room and recording in a DAT tape at a sampling rate of 48 kHz and 16-bit value. Best speech sample is used in this work. Continuous read speech data for three vowels /i/, /a/ and /u/ of speakers (male, female, and children) with healthy to gradually increasing severity of HP are obtained from American Cleft Palate Craniofacial Association speech database. From the middle of the speech waveform a stable portion is cut for the purpose of analysis. For analysis, speech samples are resampled to 16kHz and then

normalized. Speech sample of non-CP healthy speakers are used as reference.

Two types of speech stimuli used are Healthy and CP speech. Total 90 data used are:

- **27 Non-CP speech stimuli from 6 Healthy speakers:** 18 Isolated uttered Bangla data (3 oral (BO) vowel and 3 nasal (BN) vowel speech stimuli from each of 1 male, 1 female and 1 child). 9 read speech oral (EO) vowel speech stimuli (3 vowel speech stimuli from each of 1 male, 1 female and 1 child).

- **63 speech stimuli from 21 CP speakers:** 63 CP data (3 vowel data from each of 7 males, 7 females and 7 children) segmented from continuous English read speech sentences.

III. ACOUSTIC ANALYSIS OF NON-CP HEALTHY AND HYPERNASAL SPEECH

Speech samples are first normalized. As speech signal is not stationary. it is divided into frames of short duration by windowing to make it stationary. Hamming windowing function is used for making frames. A frame-rate of 50-100 frames/sec is used for analysis. Duration of speech segment is taken to be 20-30 msec for each frame. Hamming windowing function is shifted by 10msec to make a frame. These speech samples are then used for analysis.

Linear Predictive Coding (LPC) analysis is a method of speech analysis based on source filter model of speech production. LPC analysis method breaks down speech signal into pitch frequency, it's amplitude. Vocal tract is modelled by all pole coefficients called LPC order. For voiced speech, excitation source is periodic impulse train. For unvoiced speech vocal tract is excited by random polarity noise. LPC model has voiced or unvoiced decision parameter, for voiced pitch pitch frequency, G which is gain parameter, a_k are the linear prediction coefficient parameter. In z domain the vocal tract transfer function V(z) is given by Eq.1. order of all pole filter is given by p.

$$V(z) = \frac{G}{1 - \sum_{k=1}^{p} a_k z^{-k}} \qquad (1)$$

LPC filter of order 28 and Hamming window of length 20msec at 10ms shift is used for analysis. The extracted speech parameters are the power, pitch period and formants,

IV. RESULTS AND DISCUSSION

For comparing the inter-speaker variability of acoustic features of HP across male, female and children speech with varying HP we used scatter plot. For detecting and comparing HP across the selected speakers, VSA of healthy and hypernasal speech is employed. To measure the inter speaker variability parameters representing formants are plotted in two dimension scatter plot. Dispersion about the mean of the scatter plot gives the inter-speaker variability.

It is observed from analysis that, spectrum of female and children is not as clear and concise for formant extraction for some CP speakers as compared to male spectrum. Therefore, it was comparatively easier to extract formants from male spectrum than female and children spectrum.

Fig.2 shows the working procedure. First, the values of the vocal tract parameters (first and second formants, F1 and F2) of the vowels are analyzed by LPC technique and are documented. Second, a comparison on the variation of the measured vocal tract parameters of males, females and children is conducted. Finally, VSA of CP speakers is compared to healthy normal VSA to assess HP. HP detection variation among males, females and children are investigated to make a comparison.

A. Comparative Study of Acoustic Feature Variation

Fig. 3. shows the vocal tract transfer function for healthy speaker and CP8 (cleft palate speaker number 8) male, female and children speakers respectively. For CP2, spectrum of /i/ of male and female of CP2 is plotted in Fig. 4. With increasing hypernasality, spectrum with HP shows significant change in spectrum. Extra poles start to appear in the spectrum. For female and children spectrum high frequency formants starts to appear. Fig. 5 shows the scattered plot of the F1 and F2 values in Bark. In scatter plot diagram, F1 and F2 coordinate represents each vowel. Interspeaker variation within and across vowels is displayed in the scatter plot. Each vowel is clearly identified and can be differentiated in scatter plot. Average value of three formants is represented by each point for every subject.

Standard deviation from the average value is calculated which is a measure of various inter-speaker variability. Across male, female and children, inter-speaker variability of acoustic features among CP speakers are measured to be different within and across vowels. For pronouncing normal /i/, articulator is
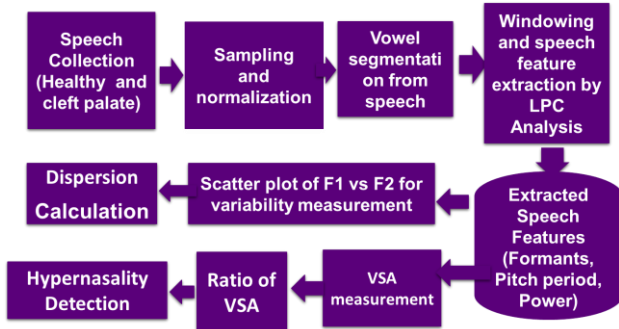

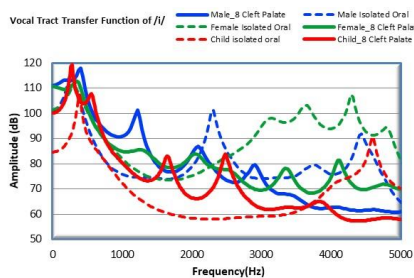Fig. 2. Block diagram of working procedure for acoustic variability measurement and HP detection.

Fig. 3. Vocal tract transfer function for /ɪ®/ and /i/ for healthy and CP speaker 8 of male (M), female (F) and Child (C) speakers.
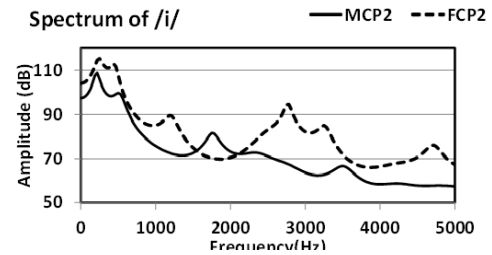

Fig. 4. Vocal tract transfer function of /i/ of Male and Female CP speaker 2(MCP2 and FCP2)

characterized by semi-openness, and has the highest front position among the vowels. /i/ and /u/ has the lowest F1 among the vowels. /u/ has the highest back position with lowest F2. During the production of /u/, articulators are characterized by lip-rounding, closeness, backness.

The variability is measured by dispersion of the scatter plot of the formant values. The inter-speaker variability among CP subjects for /i/ is 0.96 with mean (4.9, 13.82) for male, 1.23 with mean (3.7, 14.2) for female and 1.35 for children with mean (5.5, 13.7). The variability for /a/ is 1.13 with mean (5.65, 10.46) for male, for female 0.71 with mean (7.2, 10.9) and for children 1.39 with mean (6.1, 10.4). Variability for /u/ is 2.05 with mean (4.58, 9.64) for male, it gives 3.06 with mean (3.6, 8.6), for female and for children 2.03 with mean (5.2, 10.1).

From Fig. 6, inter-speaker variability measurement shows that high back vowel /u/ is mostly affected and has the highest variability across speakers in the concerned speech data. Female speech has the highest variability for /u/ among the selected speaker group. For some speakers it reflects the differences in articulatory openness by showing large value.

Vowel /i/ is least affected and has the lowest variability with HP in males and is most affected in children speech. Vowel /a/ is least affected in female read speech, has intermediate value in male speech and shows highest value in children speech. The amount of inter-speaker variability of CP speakers in the high front vowels /i/ is less than open vowel /a/ for males.
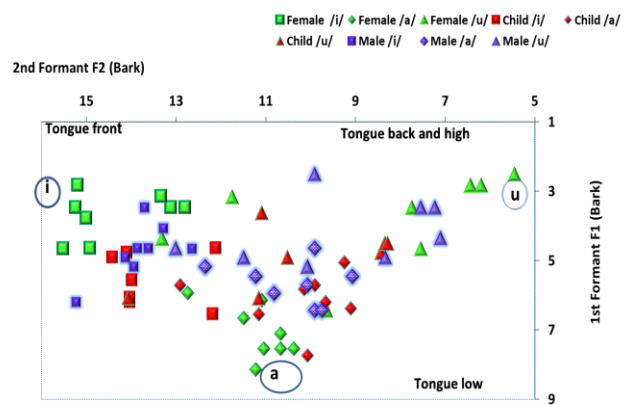

Fig. 5. Scatter plot of /i/, /a/, /u/ for CP Male, Female and Children speakers.
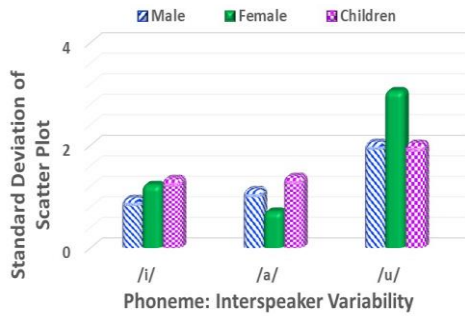
Fig. 6. Interspeaker variability measurement of /i a u/ for CP Male, Female and Children speakers.

B. Variability of VSA and HP Detection

Second part of our study is concerned with VSA calculation for all 21 CP speakers. VSA obtained from healthy Non-CP speakers speech is taken as reference. Ratio of VSA of hypernasal speech and normal speech is calculated and used as a measure for HP detection. VSA of all selected speech data and their graphical plot is shown in Fig. 7.

From the result, four types of VSA are obtained. VSA of isolated oral vowel has the highest area. Isolated uttered nasal vowel of healthy speakers has the second highest area. Read speech vowel of healthy speakers has the third highest area. The lowest VSA is obtained for average of CP speakers read speech. For males, females and children, the acoustic vowel spaces show how the isolated VSA distanced from read speech VSA and the hypernasal speech VSA. Results show that,

$$VSA_{isolated.oral} > VSA_{isolated.nasal} > VSA_{read.oral} > VSA_{HP}$$

Fig 7(a) shows the VSA for male, Fig 7(b) shows the VSA for female and 7(c) shows VSA for children. Average value of VSA of CP male, CP female, and CP children speaker along with the healthy speaker is indicated by blue, green and red triangle in Fig. 7(d). Solid triangle and dotted triangle shows the value for CP and healthy speakers respectively. Fig 7(e) shows how the VSA decreases with the increase of severity of HP indicating that VSA can perform as HP indicator. Fig. 7(f) shows the summary of result obtained for VSA of males, females and children. From Fig. 7(f) it is clear that VSA of female has the highest value for all type of speech stimuli indicating that speech clarity of CP female speaker is highest and CP children speech VSA is lowest among all. Therefore from the experimental result it is observed that speech clarity of children reduces drastically.
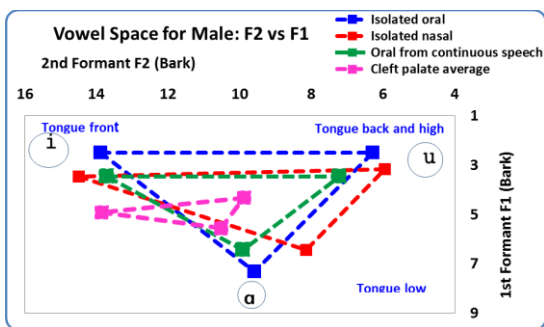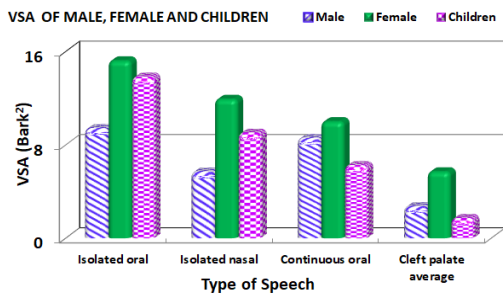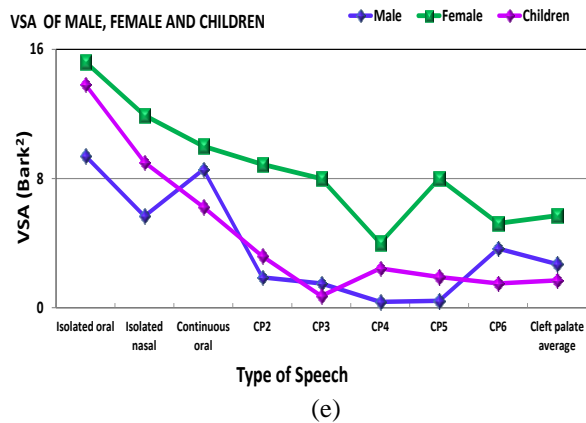

(a)

(b)

(c)

(d)
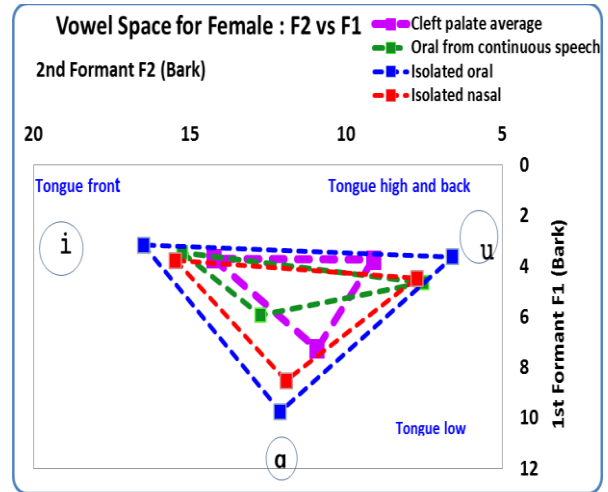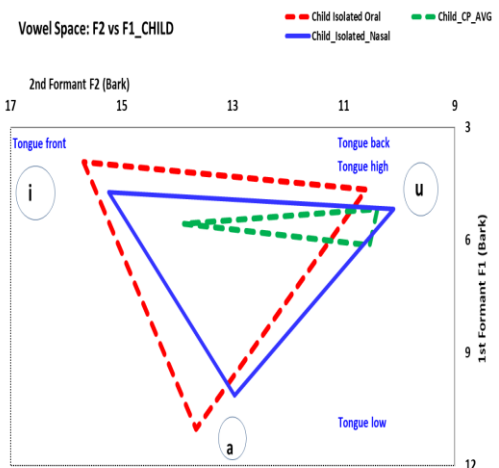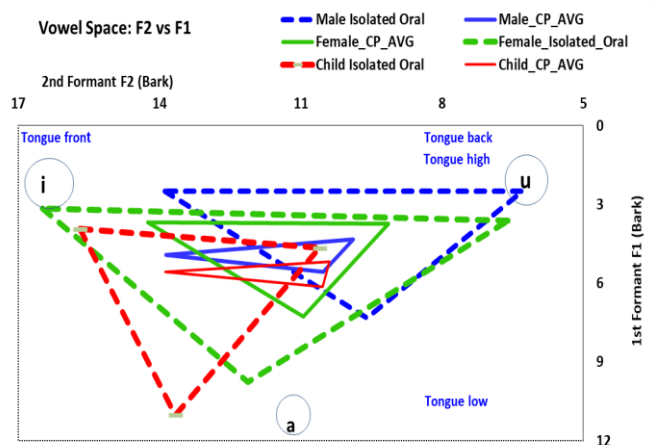
Fig. 7. Various VSA in F1xF2 plane for (a) males, (b) females (c) children (d) male, female and children (e) Gradual change and decrease of VSA with increasing VP opening resulting in gradual increasing severity of HP (f) Summary of VSA result
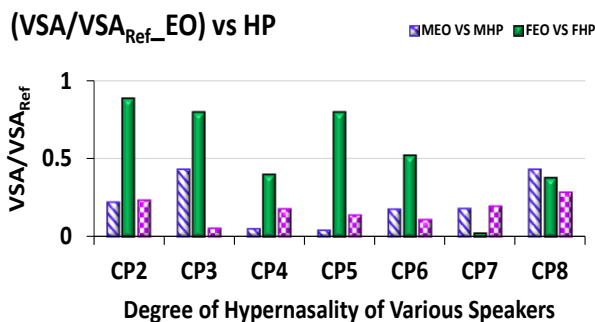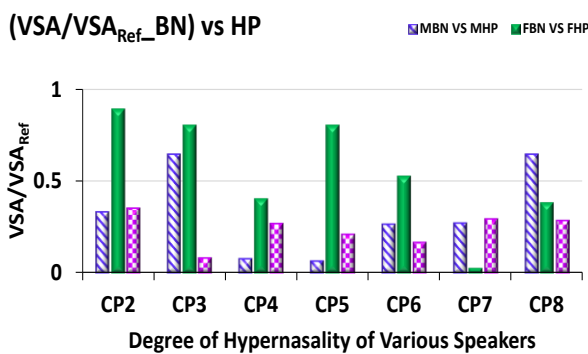

Fig. 8. Vowel space ratio across various VP opening conditions taking healthy Isolated oral vowel as reference.

Fig. 8 shows the plotted result of the vowel space ratio obtained across various VP opening conditions by taking healthy Isolated oral vowel as reference for male, females and children. While taking Oral vowel EO as reference the maximum value of ratio came out to be 0.45 for children, 0.4 for male and 0.84 for females. While taking BN as reference the maximum value of ratio came out to be 0.67 for children, 0.65 for male and 0.74 for females.

These measures may be used as threshold for determining HP. Female CP average VSA is observed to be 3.7 times more than average VSA obtained for male and 4.1 times for children. Euclidean distance measure between male and female VSA in /i/ is 1.29, /a/ is 1.75 and /u/ is 0.75. This reduction in VSA appropriately reflects the effect of HP of CP female speakers across various conditions of severity.

It is understood that due to the effect of HP speech clarity is reduced in males, females and children. Therefore as observed, VSA can differentiate hypernasal speech from normal vowel articulation indicating it's relation to the sensitivity to interspeaker variability. VSA of connected read speech of CP speakers may be used for detecting HP of CP speakers across various conditions of severity.

V. CONCLUSION

Continuous read speech is articulated differently from isolated vowel. This study reports the outcomes of the experimental observations obtained by comparing acoustic characteristics of males and females with increasing severity of HP and brings together the isolated Bangla oral and nasal vowels and continuous read speech English vowels of normal and CP speakers. The main objective of this work is to make a comparative study for the variation of acoustic features for CP male, female and children speakers due to various VP opening and detection of HP using VSA. For this purpose, VSA is estimated for each of the selected speakers utilizing extracted acoustic features. The evolution of VSA with the 8 degrees of hypernasal articulation is analyzed. This triangle consists of the first two formants of three vowels /i/, /a/ and /u/ represented in the vowel space. Interspeaker variability of HP among CP male, female and children speakers is measured by calculating mean and standard deviation of the selected vowels. /u/ shows the most variability in all speaker groups. Second conclusion is the significant reduction of the VSA in both CP male, female and children speakers as speech becomes less articulated and disordered. VSA of females shows greater excursion than children and males.

REFERENCES

[1] R. Rullo, D. Di Maggi o, V.M. Festa, and N. Mazzarella, "Speech assessment in cleft palate patients: a descriptive study," Int. J. of Pediatric otorhinolaryngology, 73(5):641-644, 2009.

[2] "Congenital malformations worldwide," International Clearing house for Birth Defects Monitoring Systems, Amsterdam, Holland. Tech. Rep., 1991.

[3] H. Extence and S. Cassidy, "The role of the speech pathologist in the care of the patient with cleft palate," in Maxillofacial Surgery (Third Edition), third edition ed., P. A. Brennan, H. Schliephake, G. Ghali, and L. Cascarini, Eds. Churchill Livingstone, 2017, pp. 1014 – 1023.

[4] E. Akafi, M. Vali, N. Moradi, Detection of hypernasal speech in children with cleft palate, in 19th Iranian Conference of Biomedical Engineering (ICBME) (IEEE, 2013), pp. 237–241

[5] J.R.O. Arroyave, J.F.V. Bonilla, Automatic detection of hypernasality in children, in International Work-Conference on the Interplay Between Natural and Artificial Computation (IWINAC) (Springer, 2011), pp. 167–174.

[6] T. Bocklet, K. Riedhammer, U. Eysholdt, E. Nöth, Automatic phoneme analysis in children with Cleft Lip and Palate, in IEEE International Conference on Acoustics, Speech and Signal Processing (IEEE, 2013), pp. 7572–7576

[7] T. Dodderi, M. Narra, S.M. Varghese, D.T. Deepak, Spectral analysis of hypernasality in cleft palate children: a pre-post surgery comparison. J. Clin. Diagn. Res. 10(1), 1–3 (2016)

[8] R. Kataoka, D.W. Warren, D.J. Zajac, R. Mayo, R.W. Lutz, The relationship between spectral characteristics and perceived hypernasality in children. J. Acoust. Soc. Am. 109(1), 2181–2189 (2001)

[9] S. Haque, M.H. Ali, A.K.M.F. Haque, Cross-gender acoustic differences in hypernasal speech and detection of hypernasality, in International Workshop on Computational Intelligence (IWCI) (IEEE, 2017), pp. 187–191

[10] S. Haque, M.Hanif, A.K.M. Fazlul,Variability of acoustic features of hypernasality and it's assessment.Int. J. Adv. Comput. Sci. Appl. 7(9), 195–201 (2016). http://dx.doi.org/10.14569/IJACSA.2016.070928.

[11] G.S. Lee, C.P. Wang, C.C. Yang, T.B. Kuo, Voice low tone to high tone ratio: a potential quantitative index for vowel [a:] and its nasalization. IEEE Trans. Biomed. Eng. 53(7), 1437–1439 (2006)

[12] G.S. Lee, C.P. Wang, S. Fu, Evaluation of hypernasality in vowels using voice low tone to high tone ratio. Cleft Palate Craniofac. J. 46(1), 47–52 (2009)

[13] K. Nikitha, S. Kalita, C. Vikram, M. Pushpavathi, S.M. Prasanna, Hypernasality severity analysis in cleft lip and palate speech using vowel space area, in Interspeech, 2017, pp. 1829–1833

[14] M Saxon, A Tripathi, Y Jiao, JM Liss, V Berisha, Robust estimation of hypernasality in dysarthria with acoustic model likelihood features - IEEE/ACM Transactions on Audio, Speech, and Language Processing 28, 2511-25222020

[15] D.A. Cairns, J.H.L. Hansen and J.E. Riski, "Detection of hypernasal speech using a nonlinear operator," in Proc. of IEEE Conf. on Engineering in Medicine and Biology Society, pp. 253-4, 1994.

[16] D. K. Rah, Y. I. ko, C. Lee, and D. W. Kim, "A noninvasive estimation of hypernasality using a linear predictive model," Ann. Biomed. Eng., vol. 29, pp. 587–594, 2001.

[17] P. Vijayalakshmi and M. R. Reddy, "Analysis of hypernasality by synthesis," in Proc. of Int. Conf. Spoken Language Processing, Jeju island, South Korea, Oct. 2004, pp. 525–528.

[18] P. Vijayalakshmi, M. R. Reddy and Douglas O'Shaughnessy, "Acoustic analysis and detection of hypernasality using group delay function," IEEE Trans. Biomedical Engineering, vol. 54, no. 4, pp. 621 – 629, Apr. 2007.

[19] P. Vijayalakshmi and M. R. Reddy, "The analysis of band-limited hypernasal speech using group delay based formant extraction technique," in INTERSPEECH, Eurospeech, Lisbon, Portugal, Sep.2005, pp. 665 - 668.

[20] A. Bladon, "Two-formant models of vowel perception: Shortcomings and enhancement," Speech Commun. 2(4), 305–313 (1983).

[21] A. T. Neel, "Vowel space characteristics and vowel identification accuracy," J. Speech Lang. Hear. Res. 51(3), 574–585 (2008).

[22] S. Skodda, W. Gronheit, and U. Schlegel, "Impairment of vowel articulation as a possible marker of disease progression in Parkinson's disease," PloS ONE 7(2), e32132 (2012).

[23] L. B. Leonard, S. E. Weismer, C. A. Miller, D. J. Francis, J. B. Tomblin, and R. V. Kail, "Speed of processing, working memory, and language impairment in children," J. Speech Lang. Hear. Res. 50(2), 408 (2007).

[24] S. Sandoval, V. Berisha, R. L. Utianski, and J. M. Liss, "Automatic assessment of vowel space area," J. Acoust. Soc. Am. **134**, EL477 (2013).

[25] S. Scherer, L. Morency, J. Gratch, and J.P. Pestian, "Reduced vowel space is a robust indicator of psychological distress: a cross-corpus analysis," in Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pages 4789-4793, 2015.

[26] M. Eshghi, M. M. Alemi and M. Eshghi, "Vowel nasalization might affect the envelop of the vowel signal by reducing the magnitude of the rising and falling slope amplitude," J. Acoust. Soc. Am. **137**, 2304 (2015).

[27] L. Chen, Y. C. Lin, C.H. Katherine and D. K. Raymond,"Perceptual speech intelligibility and speech production variability in Mandarin-speaking children with cerebral palsy," J. Acoust. Soc. Am. **139**, 2045 (2016); http://dx.doi.org/10.1121/1.4950051).

[28] C.M. Vikram, A. Nagaraj, and S. R. M. Prasanna, "Spectral Enhancement of Cleft Lip and Palate Speech," in INTERSPEECH 2016, San Francisco, USA, September, 2016. DOI:10.21437.

[29] A Deep Learning Algorithm for Objective Assessment of Hypernasality in Children with Cleft Palate, Vikram Cmathad, Nancy Scherer, Kathy Chapman, Julie Liss, Visar Berisha, IEEE Trans Biomed Eng. 2021 Feb 10;PP. doi: 10.1109/TBME.2021.3058424.

[30] S. Umesh, Studies on inter-speaker variability in speech and its application in automatic speech recognition. Sadhana 36, 853–883 (2011). https://doi.org/10.1007/s12046-011-0049-x.