Detection of hypernasality based on vowel space area

Akhilesh Kumar Dubey, Ayush Tripathi, S. R. M. Prasanna, and S. Dandapat

Citation: The Journal of the Acoustical Society of America **143**, EL412 (2018); doi: 10.1121/1.5039718
View online: https://doi.org/10.1121/1.5039718
View Table of Contents: https://asa.scitation.org/toc/jas/143/5
Published by the Acoustical Society of America

Detection of hypernasality based on vowel space area

Akhilesh Kumar Dubey,[1,a)] **Ayush Tripathi,**[2] **S. R. M. Prasanna,**[1] **and S. Dandapat**[1]
[1]*Department of Electronics and Electrical Engineering, Indian Institute of Technology Guwahati, Guwahati 781039, India*
[2]*Department of Electrical and Electronics Engineering, Visvesvaraya National Institute of Technology (VNIT), Nagpur, Maharashtra 440010, India*
d.akhilesh@iitg.ernet.in, ayushtripathi1811@gmail.com, prasanna@iitg.ernet.in, samaren@iitg.ernet.in

Abstract: This study proposes a method for differentiating hypernasal-speech from normal speech using the vowel space area (VSA). Hypernasality introduces extra formant and anti-formant pairs in vowel spectrum, which results in shifting of formants. This shifting affects the size of the VSA. The results show that VSA is reduced in hypernasal-speech compared to normal speech. The VSA feature plus Mel-frequency cepstral coefficient feature for support vector machine based hypernasality detection leads to an accuracy of 86.89% for sustained vowels and 89.47%, 90.57%, and 91.70% for vowels in contexts of high pressure consonants /k/, /p/, and /t/, respectively.

[DDO]
Date Received: December 10, 2017 **Date Accepted:** May 8, 2018

1. Introduction

The cleft palate (CP) is a craniofacial birth defect in children. Five universally accepted speech disorders associated with CP are hypernasality, hyponasality, nasal air emission and/or turbulence, consonant production errors, and voice disorder.[1] Among these disorders, hypernasality is the most frequently occurring disorder, which is defined as excess nasality heard on the voiced sounds, especially vowels.[1] Hypernasality is due to leakage of air through the nose during the production of vowels. It is different from nasal air emission which is a term defined for leakage of air through the nose during the production of pressure-sensitive sounds (plosives, nasal stop, fricatives, and affricates).[2] Hypernasality leads to reduced intelligibility for many speakers with CP. In order to block the air leakage from the nose, multiple surgeries may be performed to repair the cleft and create a normal resonance balance during speech production. Despite the surgical repair, hypernasality still persists in many individuals with CP. This is due to the inability of velum to completely close the velopharyngeal port during the production of speech sounds. The evaluation of hypernasality helps plastic surgeons and speech-language pathologists (SLPs) to decide whether additional surgery or speech therapy is needed for the treatment of hypernasality.

In the clinical environment, hypernasality evaluation is done perceptually by experienced SLPs, and the judgment is verified using instrumental techniques. Verification is done because the perceptual judgment may vary among the SLPs.[3] This happens because hypernasality occurs in conjunction with abnormalities in pitch, loudness, voice quality, and/or articulation in CP speech. The co-existence of these abnormalities affects the nasality perception.[4] The instrumental techniques may be invasive or non-invasive. The invasive instrumental techniques such as x ray (cephalometry), videofluoroscopy, and nasendoscopy are used to observe the movement pattern of the velopharyngeal port. These techniques may cause discomfort and have a radiation risk to the subject. The non-invasive instruments like Nasometer (PENTAX Medical, Montvale, NJ),[5] accelerometer, and PERCI-SARS (PERCI-SARS MicroTonics Corp., Chapel Hill, NC) measure sound pressure, nasal vibrations, and the oral and the nasal airflow, respectively. These instruments are portable, comfortable, and safe for the subject.

Researchers have used acoustic cues as an alternative objective measure for the detection of hypernasality. The acoustic cues are obtained from spectral analysis of hypernasal speech using digital signal processing methods. In this method, isolated phoneme vowels /ɑ/, /i/, and /u/ or vowels taken from consonant-vowel syllables are

[a)]Author to whom correspondence should be addressed.

analyzed for hypernasality detection. The isolation of vowels is done to avoid consonants because they may add nasal air emission and consonant production errors in the data. The important acoustic cues for nasalized vowels such as shift in formants location, especially first formant (F_1), reduction in F_1 strength, and addition of extra-nasal formant and anti-formant pairs in vowel spectrum are proposed in literature for nasality detection.[4] Hence, features are extracted based on these cues for hypernasality detection. The features such as difference between the low-pass and bandpass profile of the Teager energy operator (TEO),[6] voice low tone to high tone ratio,[7] and a combination of Mel-frequency cepstral coefficient (MFCC) and TEO profile[8] are used to detect hypernasality. In Ref. 4, a feature extracted from a high-resolution group delay spectrum is used for hypernasality detection. The features based on acoustic and noise, cepstral analysis, and nonlinear dynamic features and a combination of nonlinear dynamic features plus entropy measurements,[9] are also used for hypernasality detection. The automatic detection and severity grading of hypernasality are done in Ref. 10 using an energy distribution ratio feature and a Gaussian mixture model classifier.

The above features perform well for hypernasality detection with detection accuracies ranging from 70% to 90%. These accuracies, however, vary from vowel to vowel even for the same feature. For example, the hypernasality detection works in Refs. 6 and 7 are performed for specific vowels only. This variation in accuracy happens because hypernasality is more evident in high vowels compared to low vowels.[1] In literature, no attempt has been made to detect hypernasality using a feature which is extracted simultaneously from the three vowels /ɑ/, /i/, /u/. The vowel space area (VSA) is a feature which is extracted simultaneously from the three vowels /ɑ/, /i/, /u/, hence, it is used for hypernasality detection in this work.

The VSA is a two-dimensional (2D) area bounded by lines joining the first formant F_1 and second formant F_2 coordinates of vowels. Recently VSA is used to study the characteristics of various pathology like psychological distress,[11] hearing impairment,[12] and for analyzing regional dialect variation and sound change.[13] A recent study of VSA based on hypernasal speech produced by individuals with repaired CP has shown a reduction in the vowel space.[14] Another study analyzing the effect of severity of hypernasality on VSA has reported in Ref. 15. However, the application of VSA as a feature for automatic detection of hypernasality has not been explored systematically. Motivated by the ability of VSA to characterize the pathological speech, dialect variation, and hypernasality severity analysis by capturing the shift in positions of F_1 and F_2 frequencies in vowels, this work uses the VSA as an additional feature along with the 39-dimensional MFCC feature for the hypernasality detection using the support vector machine (SVM) classifier. The detection is done for sustained vowels /ɑ/, /i/, /u/ and the same vowels in context of pressure consonants /k/, /p/, /t/, respectively.

The organization of the paper is as follows: Section 2 discusses the spectral analysis of hypernasal speech. Section 3 discusses the CP speech database. Section 4 discusses the hypernasality analysis using VSA. Section 5 discusses the result of hypernasality detection using VSA. The conclusion and future work are given in Sec. 6.

2. Spectral analysis of hypernasal speech

The vowel spectrum is affected by hypernasality with the addition of extra formant and anti-formant pairs.[4] This addition happens at natural frequencies of the nasal tract and sinuses present. The natural frequency of the nasal tract lies in the frequency range of 450 to 650 Hz and 1800 to 2400 Hz[16] and for sinuses, it lies around 400 and 1300 Hz.[17] These extra pairs reduce the F_1 strength and shift the locations of F_1 and F_2 frequencies. Figures 1(a)–1(c) show the linear prediction (LP) spectrum of normal and hypernasal vowels /ɑ/, /i/, and /u/, respectively. From Fig. 1 it can be observed that

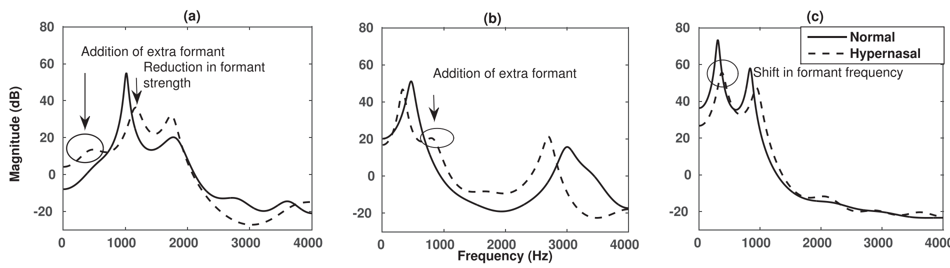

Fig. 1. LP magnitude spectrum of the normal and hypernasal vowel. (a) For /a/ vowel, (b) for /i/ vowel, and (c) for /u/ vowel. The figure shows the addition of extra formants, reduction in formant strength, and shift in formant frequency for hypernasal vowels.

there is an addition of extra formant either in frequency below F_1 or above F_1. Further, the reduction in first formant strength and the shifting of F_1 frequency for all three vowels in the case of hypernasal speech can also be observed from Fig. 1. This shifting of formants affects the VSA of speech perceived as hypernasal.

3. Speech database

The speech database used in this work is collected in All Indian Institute of Speech and Hearing, Mysore, India.[18] Two group of speakers participated in data collection: The CP group containing 15 children with repaired CP and hypernasality present in their speech and the control normal (CN) group containing 15 children with normal speech production. Each group contains 9 boys and 6 girls in the age range of 7–12 yrs. Hence in both groups the distribution of age and gender is matched. None of the children who participated for data collection have any history of hearing impairment disorder. Speech is recorded in the Kannada language, which is a language spoken in Karnataka state located in the southern part of India. The dataset consisted of 62 sustained vowel phonations (/ɑ/, /i/, and /u/) for the CN group and 56 for the CP group. Similarly, for the CN group 122 productions of each /k/ context words (/kaka/, /kiki/, and /kuku/) were recorded while 102 were recorded for the CP group. The number of /t/ context words (/tata/, /titi/, and /tutu/) recorded were 80 in the CN group and 83 for the CP group and the number of /p/ context words (/papa/, /pipi/, and /pupu/) recorded included were 115 for the CN group and 105 for the CP group. The investigator produced each target stimulus word and the subject was asked to repeat the word. The repeated response was recorded. The speech recording was done in a sound-treated room using a sound level meter microphone at a sampling frequency 44.1 kHz, with a quantization of 16 bits in .WAV format. Each recorded sound was perceptually judged for the normal and hypernasal speech by three experienced SLPs using a binary scale. The manual isolation of vowels from words are done by using the Wavesurfer tool.[19] The speech samples were down-sampled at 8 kHz for further analysis.

4. Analysis of hypernasality using VSA

4.1 VSA

In this work, VSA is defined as the area of a triangle formed by lines joining the F_1 and F_2 coordinates of three vowels /ɑ/, /i/, and /u/. To compute VSA, preprocessing of speech was done to detect sustained region of vowels. The next step was to find the formants F_1 and F_2 from each frame of the vowels. The formant frequencies (F_1, F_2) of vowels form coordinates of three vertices of the vowel triangle. Euclidean distance between each pair of vertices gives the length of three sides of the triangle. Area of the triangle was calculated using the Heron's formula $\text{Area} = \sqrt{s(s-a)(s-b)(s-c)}$, where $s = a + b + c/2$ and a, b, and c are the lengths of three sides of the triangle. The formant tracking method was used to find formant frequencies F_1 and F_2 for all samples. The LP analysis with order 12 was used for frame by frame formant tracking based on the sustained region of the vowels. The frame size of 20 ms and frame shift of 10 ms is used. An equal number of frames of normal and hypernasal vowels /ɑ/, /i/, and /u/ are considered for this analysis.

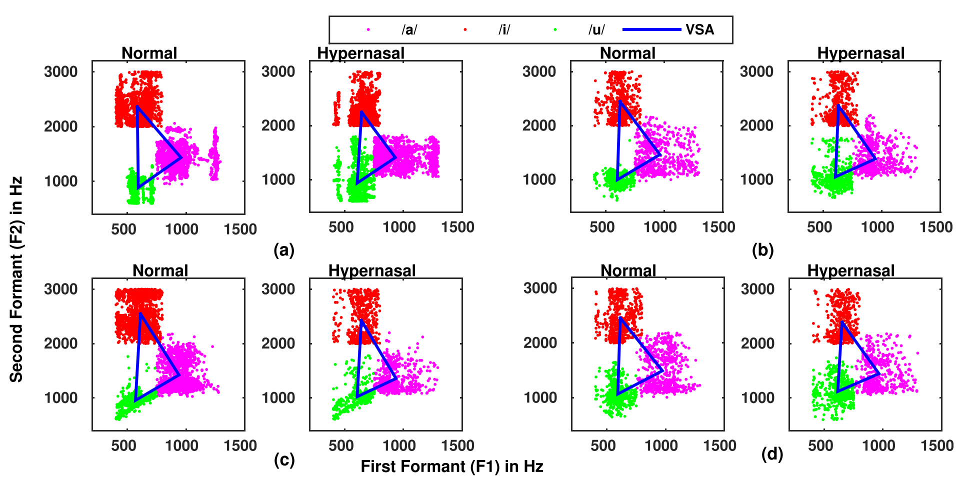

Fig. 2. (Color online) Vowel space plots of normal and hypernasal speech in (a) sustained vowels, (b) vowels in /k/ context, (c) vowels in /p/ context, and (d) vowels in /t/ context.

Table 1. Value of mean VSA for normal and hypernasal vowels and their difference.

	VSA in 10^4 Hz^2		
Vowel type	Normal	Hypernasal	Difference
Sustained	30.62	19.29	11.33
/k/ context	28.88	20.82	8.06
/p/ context	31.65	21.63	10.02
/t/ context	28.92	20.45	8.47

4.2 VSA in hypernasal speech

To compare the VSA for two groups, the sustained vowels /ɑ/, /i/, /u/ and the same vowels in context of pressure consonant /k/, /p/, /t/, respectively, were analyzed. Figures 2(a)–2(d) show F_1, F_2 frequencies plotted in 2D space as a coordinate for the isolated sustained vowels and the vowels in the contexts of /k/, /p/, and /t/, respectively. From Fig. 2 it can be observed that the regions of formant frequencies coordinates are smaller for the samples with hypernasality present compared to normal vowels. This may be due to the addition of extra-nasal formants in the spectrum, which causes the shifting in F_1, F_2 frequencies. Figure 2 also shows vowel triangles formed by the mean values of all formant frequency coordinates which is called VSA mean. The calculated VSA mean for the hypernasal speech was smaller than that calculated for the CN group. For more clearance, VSA mean values and their difference (normal-hypernasal) are listed in Table 1, which also confirms the reduction in VSA mean for the CP group compared to the CN group. The reduction is highest for sustained vowels compared to vowels produced in the context of high pressure consonants. To demonstrate differences across the speech samples, boxplots are shown in Fig. 3. A one-way analysis of variance test was also conducted to check the significance of discrimination in VSA on the entire database. The test shows significant discrimination ($p < 0.001$) for sustained vowels as well as for the vowels in context of high pressure consonants.

5. Detection of hypernasal speech using VSA

In this section, the usefulness of the VSA feature for hypernasality detection using the 40-dimensional (39-MFCC+VSA) feature is discussed. The VSA feature captures the formants location, and the MFCC feature captures the vocal tract spectrum shape to differentiate the speech from the two groups. An equal number of frames from three vowels are taken for the detection task. From each single frame of vowels /ɑ/, /i/, and /u/, formants F_1, F_2, and the 13-dimensional MFCC feature are computed. Then, each 13-dimensional MFCC feature in a sequence of /ɑ/, /i/, and /u/ are appended to find the 39-dimensional MFCC feature. Further, the VSA computed using F_1, F_2 formants of three vowels is appended to make it a 40-dimensional feature. The detection was done using the SVM classifier with polynomial kernel. The two-class SVM model is trained with 80% of the data and tested with the remaining 20% data. The detection is done for four types of vowels, namely, sustained vowels /ɑ/, /i/, /u/ and the same vowels in context of pressure consonant /k/, /p/, /t/, respectively.

5.1 Experiment results

Table 2 shows the hypernasality detection result using the VSA feature, the MFCC feature, and the VSA feature appended with the MFCC feature (MFCC+VSA) for all four types of vowels. It can be observed from Table 2 that the accuracy of the VSA feature alone is significant for all four types of vowels but it is less than the baseline MFCC feature in each case. When the VSA feature is appended with the MFCC

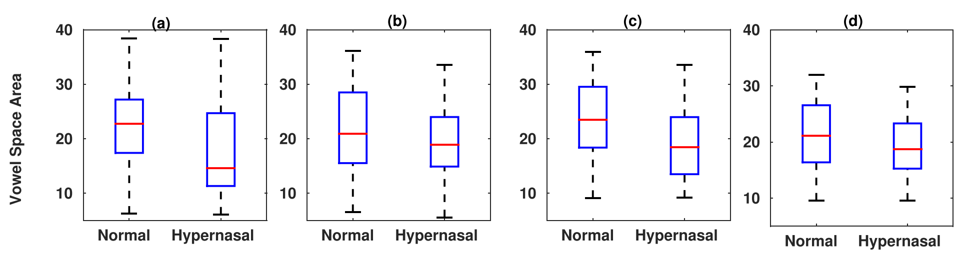

Fig. 3. (Color online) Boxplot showing the value of VSA for normal and hypernasal (a) sustained vowel, (b) /k/ context vowel, (c) /p/ context vowel, and (d) /t/ context vowel.

Table 2. Hypernasality detection accuracy (in %) for sustained vowels and the vowels in context of consonants /k/, /p/, /t/. (Acc = Accuracy, Sen = sensitivity, Sel = Selectivity).

Vowel type	Feature	Acc	Sen	Sel
Sustained	VSA	65.77	70.14	61.40
	MFCC	80.03	86.28	73.78
	MFCC+VSA	86.89	95.85	77.93
/k/ context	VSA	80.62	79.13	82.11
	MFCC	84.65	87.04	82.26
	MFCC+VSA	89.47	89.57	89.37
/p/ context	VSA	68.76	68.76	68.77
	MFCC	89.52	92.64	86.40
	MFCC+VSA	90.57	92.64	88.51
/t/ context	VSA	69.31	77.86	60.75
	MFCC	88.65	82.11	95.19
	MFCC+VSA	91.70	88.21	95.19

feature, the accuracy goes higher than the accuracy of the MFCC feature alone. There is an increment of 6.86%, 4.82%, 1.05%, and 3.05% for the sustained vowels and vowels in the context of /k/, /p/, and /t/, respectively, for the appended (39-MFCC+VSA) feature compared to the baseline MFCC feature.

6. Conclusions and future work

This paper presents the VSA feature for the detection of hypernasality during sustained vowel production. The addition of extra-nasal formants in the vicinity of F_1 in hypernasal vowels causes movement of F_1 and F_2, which results in a reduced VSA for speech judged as hypernasal compared to normal speech production. The addition of the MFCC feature to the VSA measure enhances detection accuracy compared to the baseline MFCC feature. The VSA feature is evidence derived using three vowels /ɑ/, /i/, and /u/, hence, providing a combined evidence which can be used by the researcher for better identification of hypernasality in speech. This work can be further extended as a study of the VSA feature for different severity level classification of hypernasal speech. Further, the compensatory behaviors like less mouth opening and more breathy phonation, which are often associated with hypernasality, may also affect the VSA feature; this can be explored in future work.

Acknowledgments

This work is in part supported by the project grants, for the projects entitled NASOSPEECH funded by the Department of Biotechnology (DBT), Government of India and ARTICULATE + funded by the Ministry of Human Resource Development (MHRD), Government of India.

References and links

[1]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," Cleft Palate-Craniofacial J. **45**(1), 1–17 (2008).

[2]A. W. Kummer and L. Lee, "Evaluation and treatment of resonance disorders," Lang., Speech, Hear. Svc. Schools **27**(3), 271–281 (1996).

[3]R. I. Zraick and J. M. Liss, "A comparison of equal-appearing interval scaling and direct magnitude estimation of nasal voice quality," J. Speech, Lang., Hear. Res. **43**(4), 979–988 (2000).

[4]P. Vijayalakshmi, M. R. Reddy, and D. O'Shaughnessy, "Acoustic analysis and detection of hypernasality using a group delay function," IEEE Trans. Biomed. Eng. **54**(4), 621–629 (2007).

[5]C. M. Mayo and R. Mayo, "Normative nasalance values across languages," Echo **6**(1), 22–32 (2011).

[6]D. Cairns, J. H. Hansen, and J. E. Riski, "A noninvasive technique for detecting hypernasal speech using a nonlinear operator," IEEE Trans. Biomed. Eng. **43**(1), 35–45 (1996).

[7]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 Trans. Biomed. Eng. **53**(7), 1437–1439 (2006).

[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," J. Acoust. Soc. Am. **126**(5), 2589–2602 (2009).

[9]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.

[10]L. He, J. Zhang, Q. Liu, H. Yin, and M. Lech, "Automatic evaluation of hypernasality and consonant misarticulation in cleft palate speech," IEEE Signal Process. Lett. **21**(10), 1298–1301 (2014).

[11]S. Scherer, L.-P. Morency, J. Gratch, and J. Pestian, "Reduced vowel space is a robust indicator of psychological distress: A cross-corpus analysis," in *IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*) (2015), pp. 4789–4793.
[12]E. A. Wieland, E. B. Burnham, M. Kondaurova, T. R. Bergeson, and L. C. Dilley, "Vowel space characteristics of speech directed to children with and without hearing loss," J. Speech, Lang., Hear. Res. **58**(2), 254–267 (2015).
[13]R. A. Fox and E. Jacewicz, "Reconceptualizing the vowel space in analyzing regional dialect variation and sound change in American English," J. Acoust. Soc. Am. **142**(1), 444–459 (2017).
[14]S. Haque, M. H. Ali, and A. F. Haque, "Variability of acoustic features of hypernasality and its assessment," Int. J. Appl. Eng. Res. **11**(18), 9262–9268 (2016).
[15]K. Nikitha, S. Kalita, C. M. Vikram, M. Pushpavathi, and S. R. M. Prasanna, "Hypernasality severity analysis in cleft lip and palate speech using vowel space area," in *INTERSPEECH* (2017).
[16]K. N. Stevens, *Acoustic Phonetics* (MIT Press, Cambridge, MA, 2000), Vol. 30.
[17]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*) (1982), Vol. 7, pp. 911–914.
[18]AIISH, "All India institute of speech and hearing, mysore, India," web-site: http://www.aiishmysore.in (Last viewed May 20, 2018).
[19]K. Sjölander and J. Beskow, "Wavesurfer-an open source speech tool," in *Sixth International Conference on Spoken Language Processing* (2000).