Interspeech 2018
2-6 September 2018, Hyderabad

Analysis of Breathiness in Contextual Vowel of Voiceless Nasals in Mizo

Pamir Gogoi[1], *Sishir Kalita*[2], *Parismita Gogoi*[2], *Ratree Wayland*[1], *Priyankoo Sarmah*[2], *S. R. M. Prasanna*[2,3]

[1]University of Florida, Gainesville, USA
[2]Indian Institute of Technology Guwahati, Guwahati, India
[2]Indian Institute of Technology Dharwad, Dharwad, India

(pgogoi,ratree)@ufl.edu, (sishir, parismitagogoi, priyankoo, prasanna)@iitg.ernet.in

Abstract

This study analyses the source characteristics of voiced and voiceless nasals in Mizo, a Tibeto-Burman language spoken in North-East India. Mizo is one of the few languages that has voiced and voiceless nasals in its phoneme inventory. This analysis is motivated by the interaction between breathiness and nasality reported in a number of speech perception studies using synthetic stimuli. However, there are no studies examining this interaction in vowels after voiced and voiceless nasals. Existing research has also documented the interaction between breathy phonation and vowel height. The current study is an acoustic analysis of breathiness in high and low vowels following voiced and voiceless nasals in Mizo. The acoustic parameter measures are: H1H2 ratio, spectral balance (SB), strength of excitation (SoE), and waveform peak factor (WPF). The values obtained for all the four acoustic measures suggest that vowels following voiceless nasals exhibit stronger acoustic characteristics associated with breathy phonation than vowels following voiced nasals. In addition, the degree of acoustic breathiness is affected by vowel height.

Index Terms: Phonation, Voicing, Nasality, Breathiness.

1. Introduction

Nasal sounds are present in almost all the languages of the world and cross-linguistically, nasals are generally voiced segments. However, a few languages exhibit a voicing contrast in their nasal sounds; that is, they have both voiced and voiceless nasals in their phoneme inventory. As implied, the characteristic feature of a voiceless nasal is that there is no vocal fold activity while the air is released through the nasal cavity. However, existing literature shows that there can be different types of voiceless nasals in terms of when and how the voicing begins. For instance, in Mizo, a Tibeto-Burman language investigated in this study, the voiceless portion of the nasal is followed by a short period of voicing as it transitions into the following vowel. The effect of nasal consonants on its contextual vowels has also been of interest to many researchers and a number of research shows that the effect of contextual nasalization varies across languages [1]. In addition, the interaction between nasalization and other features including manner and voicing has also been investigated [2],[3] [4],[5], [6],[7]. Specifically, Ohala and Ohala (1993) provide an aerodynamic and phonetic account of how nasality interacts with place and manner features [4], and a similar account on nasality and voicing interaction is given by Sole [7]. According to Sole, the degree to which different segment type interacts with nasalization forms a hierarchy resembling the one presented by Schourup [8], with obstruents being the least likely to nasalize and vowels the most.

However, a phenomenon not as widely studied is the interaction between nasality and phonation type (e.g., breathiness) across languages. To explain the so-called spontaneous nasalization, an emergence of a contrastive nasal vowel in words that do not contain a nasal consonant etymologically, such as the development on the nasal vowel after a voiceless fricative [s] in Hindi [sãp] from Sanskrit [sarpa] snake, Ohala and Amador speculates that the need for a wider-than-normal glottal opening required for high airflow rates in voiceless segments such as voiceless fricatives or voices aspirated plosives may be partially assimilated by the adjacent (voiced) vowels [9]. In turn, this slightly larger glottal opening during voicing creates acoustic effects such as broadening first formant bandwidth mimicking nasalization without being nasal physiologically. The hypothesis is tested and confirmed [3].

However, nothing is known about how the voice quality of the vowel can vary as a function of the voicing of the preceding nasal. We hypothesize that the following vowels may assimilate acoustic effects of voiceless nasals. Nasality and breathiness shared a number of acoustic cues including low-frequency peak and wide F1 bandwidth, as mentioned by Arai in [10]. In a breathiness and nasality rating experiment using /a/ vowels synthesized by a software packaged developed by [11] and [12], [10] also found that perceived nasality increases when the frequency of the nasal zero (FNZ) values are further apart from the nasal pole (FNZ) values, and perceived breathiness increases when aspiration (AH) increases. More interestingly, however, there is an interaction between perceived nasality and breathiness. Specifically, [10] found that perceived breathiness increases when the pole-zero is slightly part (FNZ=600), but when nasality is strong (FNZ =700), perceived breathiness decreases. Similar finding is reported in [13]. In this study, breathy sources are found to increase perceived hypernasality for the mild hyper-nasal filter, but to reduce it for moderate or severe hyper-nasal filters. Both of these studies demonstrate mutual acoustic and perceptual effects of spreading glottis (breathiness) and velopharyngeal port opening (nasalization).

Studies have also shown an interaction between breathy phonation and vowel height. For instance, [14] found that more vowel tokens are identified as the higher vowel (i.e., /i/ vs /ɪ/; /u/ vs /ʊ/) when modeled with a breathy phonation particularly after female voices. This finding is congruent with the report that breathy phonation frequency occurs with vowels produced with a raised tongue body and/or advancement of the tongue root [15]. According to [15], in approximately 50 languages studied, breathy phonation is invariably found to associate with higher vowels. According to [14], acoustic effects of breathy phonation and a raised tongue body and/or an advanced tongue root effectively enhance low-frequencies and thus the perceptual distinctiveness of vowels differentiated by low-frequency

 10.21437/Interspeech.2018-1899

prominence. Motivated by the above discussion, this work investigates if the voiceless portion of the nasal has an effect on the voice source characteristics of its contextual vowels. In case of Mizo, the voicing beings towards the end of the nasal before it transitions into the vowel. However, we analyze only the transition region of the following vowel. Three different vowels are analyzed; two high vowels /i/ and /u/ and the low vowel /a/. The effect of breathiness in vowels following the voiced and voiceless nasals are studied using excitation source analysis.

The rest of the paper is organized as follows: Section 2 describes the experiment and the features used in this study as a measure of breathiness. Section 3 presents the results and the discussion. Finally, Section 4 presents the conclusion and future directions of the study.

2. Method

2.1. Dataset

Speech samples of two female native Mizo speakers were analyzed. The speakers read a list of 30 monosyllabic words of CV and CVC structure, where the onset is either voiced or voiceless nasal, followed by /a/, /i/ or /u/. The final C of the CVC syllables were non-nasal obstruents. The nasal consonants included in the data set were /m, n, ŋ/ and their voiceless counterparts. The word list consisted of 5 words of each of the 3 vowel type, in both voiced and voiceless nasal context, read by two speakers (3 vowels $\times$ 2 nasal type $\times$ 2 speakers = 30 tokens). The recording was conducted in a sound-attenuated booth with a sampling frequency was 44.1 kHz, 24 bits in WAV. format. The data collected was manually annotated and the onset of the vowels was marked at the onset of voicing. The features extracted from the speech samples are described in the subsections below. Statistical analyses were performed on the extracted values to examine the significance of the results.

2.2. Preprocessing and feature extraction

This subsection explains the features extracted from the region of interest, i.e. transition region of vowels. All the features explored in this work characterize the excitation source signal. Before extracting the features, speech signals are downsampled to 8000 Hz, and the amplitudes are normalized by $\ell_2\text{-}norm$ of the speech signal. Source features are extracted from the Linear prediction residual (LPR) based approximation of excitation source signal. The number of LP coefficients considered are $fs/1000+4$, i.e. 12. All the features are extracted Glottal Closure Instant (GCI) synchronized. Zero Frequency Filter (ZFF) based GCI estimation method is used in this work [16]. In this case, differenced version of the speech signal is passed through a zero frequency resonator. The output of the zero frequency resonator is exponentially growing or decaying depending on the polarity of the speech signal, which is further processed by a trend removal moving average filter. The positive to negative zero crossing locations are considered as the estimated GCIs of the speech signal.

2.2.1. H1-H2:

This measure describes the relative magnitude of the first and the second harmonics. Breathy glottal source signals obtained through inverse filtering typically show more symmetrical opening and closing phases with little or no complete closed phase. The round near-sinusoidal shape of the breathy glottal waveform is responsible for a relatively high amplitude of the first harmonic (H1) and relatively weak upper harmonics. However, to assess whether there is an increase in the H1 amplitude or not, H1 amplitude must be compared with some reference such as amplitude of the second harmonic (H2) [17], [18]; or amplitude of F1 [19],[20],[21]. In this study, H1 amplitude is compared to that of H2. H1-H2 is an indicative measure of open quotient [22]. Thus, breathy voice has higher H1H2 ratio compared to modal phonation [23, 24]

2.2.2. Strength of Excitation (SoE):

This measure quantifies the abruptness of the glottal closing [25, 16]. In breathy phonation, glottal closing is not abrupt which leads to lower SoE value compare to modal voice [26]. This feature is computed from the Hilbert Envelope (HE) of the LP residual with GCI synchronously. It is expected that as the breathiness increases, the peak strength of the HE of LP residual around GCI will be decreased. To compute the SoE feature, a 3 ms segment of HE of LP residual is considered around each GCI. Then, SoE is defined as the ratio of standard deviation to the mean value of the segment [25].

2.2.3. Waveform peak factor (WPF):

Waveform peak factor is used to measure the decay characteristic the LP residual signal around the GCIs. It is expected that breathiness may affect the peak strength of the LP residual, as the glottal closure is less sharp. The energy distribution of excitation source around the glottal closure may be a better attribute to capture this information. WPF is defined as follows [24],

$$WPF = \frac{max(|s_i|)}{[\frac{1}{N}\sum_{i=1}^{N} s_i^2]^{\frac{1}{2}}},$$

where s_i is the amplitude in the i^{th} sample and N is the total number of sample points considered in one glottal cycle. This factor is defined as when the waveform is flat the value of this factor will be minimum, i.e. 1. For the case of impulse function, the value of this factor will be maximum, i.e. $N^{1/2}$.

2.2.4. Spectral balance (SB):

Spectral balance feature is used to describe the spread of energy in the voice source spectrum. In this work, this feature is computed by passing the voice source spectrum through 24 triangular Mel-filterbanks, which are equally spaced in the Mel scale. The resultant 24 Mel-energies provide a smoothed representation of the excitation source spectrum [27]. Then, spectral balance (SB) feature is derived to characterize the spectral distribution as follows,

$$SB = \sum_{i=1}^{3} PE(i) / \sum_{i=1}^{24} PE(i), \quad (1)$$

$$(2)$$

where $PE(i)$ denotes the perceptive energy computed in the i^{th} Mel-frequency band.

To illustrate the significance, the LP residual and the features are computed from speech segments of around 4-5 glottal cycles of the transition regions of both voiced and voiceless nasals, which is shown in Fig. 1. From the figure, it can be clearly seen that the peak strength around the GCI is decreased in the case of voiceless nasal but not for the voiced nasal. There is an increase in the value of H1H2 in the vowels after voiceless nasals, while the values for the SB feature is higher after voiced

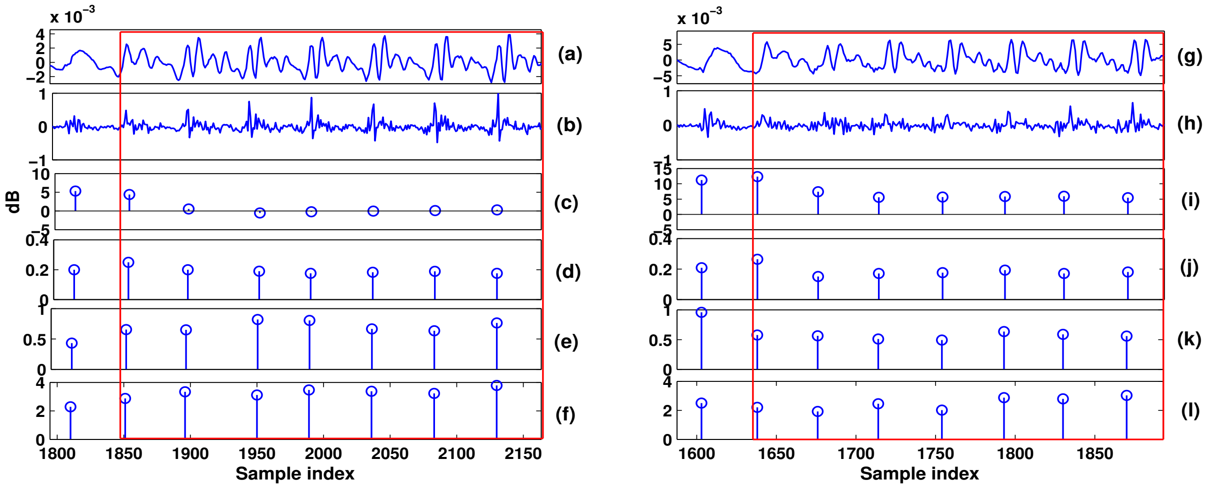

Figure 1: *Excitation source features around the transition region of the following vowel of nasal sounds. Red rectangle represents the transition region of the vowel. (a) - (f) represent the speech waveform, LP residual, H1H2 ratio, SB, SoE, and WPF for voiced nasal context, whereas (g)-(l) for that of unvoiced nasals.*

nasals. The SoE and WPF values are found to be higher when the vowel follows a voiced nasal.

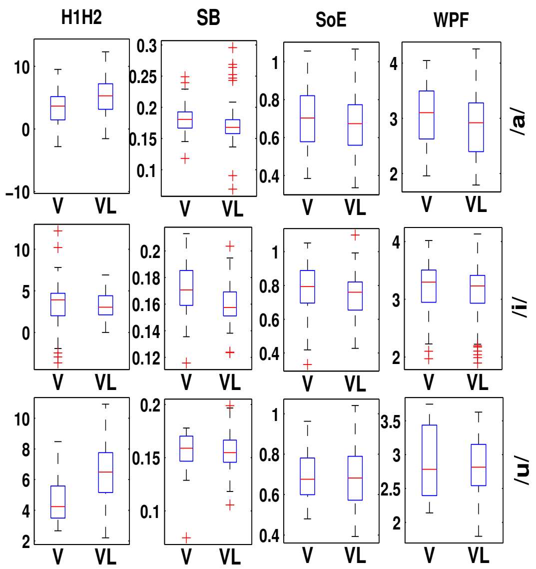

Figure 2: . *Distributions of all features for voiced (V) and voiceless (VL) nasals, respectively. Top, middle, and bottom row signifies the /a/, /i/, and /u/ vowel contexts, respectively*

3. Experimental results and discussion

This section presents the experimental results and the discussion of the analysis. The significance of the extracted features from the transition regions of preceding vowels of both voiced and voiceless nasals are studied using qualitative and quantitative procedure. Qualitative analysis is performed by box plot in Fig. 2, to show the distributions the four features. From the figure, it can be seen that the discrimination is comparatively greater in the case of /a/ than /i/ and /u/. For a better understanding of the significance, a qualitative procedure using statistical analysis is performed. The mean and the standard deviation of

Table 1: *Mean and SD of the 4 features for all the vowels (/a/, /i/, and /u/*

Context:	voiced		voiceless	
	mean	sd	mean	sd
H1H2	3.54	3.61	5.06	2.59
SB	0.17	0.02	0.16	0.02
SoE	0.73	0.15	0.69	0.14
WPF	3.08	0.52	2.90	0.52

Table 2: *Mean and SD of the 4 features for /a/*

Context:	voiced		voiceless	
	mean	**sd**	**mean**	**sd**
H1H2	3.25	4.51	5.21	2.83
SB	0.18	0.02	0.17	0.02
SoE	0.70	0.16	0.67	0.15
WPF	3.05	0.54	2.85	0.46

Table 3: *Mean and SD of the 4 features for /i/*

Context:	voiced		voiceless	
	mean	**sd**	**mean**	**sd**
H1H2	3.45	2.99	3.39	1.56
SB	0.17	0.01	0.15	0.01
SoE	0.78	0.15	0.74	0.13
WPF	3.19	0.46	3.11	0.50

Table 4: *Mean and SD of the 4 features for /u/*

Context:	voiced		voiceless	
	mean	**sd**	**mean**	**sd**
H1H2	4.52	1.42	6.32	2.08
SB	0.15	0.01	0.15	0.01
SoE	0.68	0.12	0.68	0.14
WPF	2.89	0.54	2.80	0.41

the four features are presented in Table 1. As hypothesized, the mean of H1H2 ratio in the context of voiceless nasal is higher than that of the voiced nasals and the mean of SoD, SB and WPF are lower for voiceless nasals than voiced nasals. The results of all the four features imply that breathiness in the vowels following a voiceless nasal is higher than the ones after voiced nasals. T-tests performed on the extracted values confirm that the differences in the means between the voiced and voiceless context in all the four features are statistically significant. H1H2 [$t(299)=-5.01$, $P < 0.001$], SB [$t(399) = 4.12$, $P<0.001$], SOD [$t(368)=2.46$, $P<0.05$], WPF [$t(387)=3.55$, $P<0.001$].

Table 2,3 and 4 show the mean values and standard deviations of the four acoustic parameters measured for the vowels /a/, /i/ and /u/ separately. To examine the significance of vowel quality on breathiness, a factorial ANOVA is performed with the values obtained for the four acoustic parameters as the dependent variable and the nasal context (voiced/voiceless) and vowel type (a/i/u) as the independent variables. The results show a significant main effect of nasal context and vowel height in all the four features. However, the interaction between nasal context and vowel type is significant only for the H1H2 [$F(2,489)=5.76$, $P<0.01$] and SB [$F(2,443)=3.14$, $P<0.01$] features.

H1H2 ratio indicates that /a/ and /u/, but not /i/ are acoustically breathier following voiceless nasals than voiced nasals. On the other hand, spectral balance values for /a/ and /i/, but /u/ are lower in the context of voiceless nasals than voiced nasals.

4. Conclusions

No earlier studies have explore the interaction between nasality and breathiness in vowels following voiced and voiceless nasals. It is found that four acoustic parameters extracted from the onset portion of vowels following voiceless nasals exhibit acoustic characteristics of being breathier than those following voiced nasals in Mizo. This finding is consistent with the hypothesis that wider glottal opening during voiceless nasals affects the source (glottal) spectrum of the onset portion of its following vowels. Specifically, H1H2 ratio, SB, SoE, and WPF measures suggest that the glottal configuration of this portion of the vowel is similar to that of a breathy phonation. Specifically, in breathy phonation, there is a considerable amount of low-frequency flow because the glottis is never completely closed at the back over the vibratory period. Consequently, the source signal shows increased open quotient (H1-H2) and a very strong fundamental component (H1). However, though not consistently observed across four acoustic measures, high vowels seem to be breathier than low vowel overall. In addition, for some parameters, an interaction between vowel height and nasal context is observed.

The future work planned is to explore the glottal opening and closing behavior using the estimated glottal flow waveform in the preceding vowel context of voiceless nasals. The effect of nasalization on the LP residual in the nasal vowel transition due to the addition of extra zeros in the vocal tract spectrum might show characteristics associated with breathy phonation. Our future work shall try to analyze more into this interaction.

5. References

[1] P. S. Beddor and W. Strange, "Cross-language study of perception of the oral–nasal distinction," *The Journal of the Acoustical Society of America*, vol. 71, no. 6, pp. 1551–1561, 1982.

[2] M. Ohala, "Nasals and nasalization in hindi," *Language Universals Project, Stanford: Stanford University*, pp. 317–332, 1975.

[3] M. Ohala and J. J. Ohala, "Nasal epenthesis in hindi," *Phonetica*, vol. 48, no. 2-4, pp. 207–220, 1991.

[4] J. J. Ohala and M. Ohala, "The phonetics of nasal phonology: Theorems and data," in *Nasals, nasalization, and the velum*. Elsevier, 1993, pp. 225–249.

[5] J. Ohala, M.-J. Solé, and G. Ying, "The controversy of nasalized fricatives," in *ICA-ASA Joint Meeting, Seattle, Washington*, 1998.

[6] J. J. Ohala, "Phonetic explanations for sound patterns," *A figure of speech: A festschrift for John Laver*, p. 23, 2005.

[7] M.-J. Solé, "Compatibility of features and phonetic content. the case of nasalization," in *Proceedings of ICPhS*, 2007, pp. 261–266.

[8] L. Schourup, "Characteristics of vowel nasalization," *Research on Language & Social Interaction*, vol. 5, no. 4, pp. 530–548, 1972.

[9] J. I. Ohala and M. Amador, "Spontaneous nasalization," *The Journal of the Acoustical Society of America*, vol. 69, no. S1, pp. S54–S54, 1981.

[10] T. Arai, "Cue parsing between nasality and breathiness in speech perception," *Acoustical science and technology*, vol. 27, no. 5, pp. 298–301, 2006.

[11] D. Klatt, "The new mit speech vax computer facility," *Speech Communication Group Working Papers IV*, 1984.

[12] D. H. Klatt and L. C. Klatt, "Analysis, synthesis, and perception of voice quality variations among female and male talkers," *the Journal of the Acoustical Society of America*, vol. 87, no. 2, pp. 820–857, 1990.

[13] S. Imatomi, "Effects of breathy voice source on ratings of hypernasality," *The Cleft palate-craniofacial journal*, vol. 42, no. 6, pp. 641–648, 2005.

[14] A. J. Lotto, L. L. Holt, and K. R. Kluender, "Effect of voice quality on perceived height of english vowels," *Phonetica*, vol. 54, no. 2, pp. 76–93, 1997.

[15] K. Denning, "The diachronic development of voice quality," Ph.D. dissertation, Ph. D. Disseration. Stanford University, 1989.

[16] K. S. R. Murty and B. Yegnanarayana, "Epoch extraction from speech signals," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 8, pp. 1602–1613, 2008.

[17] J. Hillenbrand, R. A. Cleveland, and R. L. Erickson, "Acoustic correlates of breathy vocal quality," *Journal of Speech, Language, and Hearing Research*, vol. 37, no. 4, pp. 769–778, 1994.

[18] J. Hillenbrand and R. A. Houde, "Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech," *Journal of Speech, Language, and Hearing Research*, vol. 39, no. 2, pp. 311–321, 1996.

[19] P. Ladefoged, "A course in phonetics, 2nd edn," 1982.

[20] P. Ladefoged and N. Antoñanzas-Barroso, "Computer measures of breathy phonation," *UCLA working papers in phonetics*, vol. 61, pp. 79–86, 1985.

[21] P. L. Kirk, J. Ladefoged, and P. Ladefoged, "Quantifying acoustic properties of modal, breathy and creaky vowels in jalapa mazatec," *American Indian linguistics and ethnography in honor of Laurence C. Thompson*, pp. 435–450, 1993.

[22] R. Wayland and A. Jongman, "Acoustic correlates of breathy and clear vowels: The case of khmer," *Journal of Phonetics*, vol. 31, no. 2, pp. 181–201, 2003.

[23] E. Castillo-Guerra* and A. Ruz, "Automatic modeling of acoustic perception of breathiness in pathological voices," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 4, pp. 932–940, April 2009.

[24] D. G. Childers and C. K. Lee, "Vocal quality factors: Analysis, synthesis, and perception," *The Journal of the Acoustical Society of America*, vol. 90, no. 5, pp. 2394–2410, 1991.

[25] G. Seshadri and B. Yegnanarayana, "Perceived loudness of speech based on the characteristics of glottal excitation source," *The Journal of the Acoustical Society of America*, vol. 126, no. 4, pp. 2061–2071, 2009.

[26] S. A. Thatia, B. B., P. Bhaskararao, and B. Yegnanarayana, "Analysis of breathy voice based on excitation characteristics of speech production," in *In Proc. International Conference on Signal Processing and Communications*, July 2012.

[27] T. Drugman, T. Dubuisson, and T. Dutoit, "On the mutual information between source and filter contributions for voice pathology detection," in *Tenth Annual Conference of the International Speech Communication Association*, 2009.