Acoustics of articulatory constraints: Vowel classification and nasalization

Ayushi Pandey, Indranil Dutta
ayuship.09@gmail.com, indranil@efluniversity.ac.in

Vowel classification under varying coarticulatory influence of nasals

Purpose: Measuring sensitivity of machine learning classifiers towards varying articulatory constraints of nasal consonants

Data: Contextually nasalized vowels in Bangla. Coarticulatory influence from nasal consonants; dental /**n**/ and labial /**m**/

Introduction

- Feature identification and retrieval of acoustic correlates of nasalized vowels [1]
- Feature independent methods such as cepstral coefficients and formant tracking [2, 3]
- Established features: F1 amplitude reduction, increased bandwidth, spectral prominence around F1 region [4, 5]
- Comparison of relative prominence of a certain class of features over others [4, 5]
- Ensemble of machine learning classifiers to classify vowels under the influence of nasals [6, 7]
- Popular classifiers: Support Vector Machines, Naïve Credal Classifier, Multiboost classifier, HMM models [6, 7]
- Reduced performance of feature based learning and classification in the case of contextual nasalization [2, 3]
- Incorporation of constraints specific to coarticulatory sensitivity of the consonant has not received attention so far

♪ Our work: Two-fold approach

1. Feature based classification of vowels under contextual influence of nasal consonants with different articulatory complexity
2. Testing the resultant classification against the predictions of The Degree of Articulatory Constraints (DAC) with respect to coarticulatory sensitivity and directionality

Corpus

Speech corpus Read-out sentences from continuous Bangla ASR corpus [8]

Nasals Bangla has contrastive oral-nasal vowels

- 2 Female and 2 Male speakers; 200 sentences

Annotation Manual annotation in Praat [9]

- Single-tier annotation
- Vowel identity with consonant and directionality information

Nasality specific feature-set

- Amplitude of the first formant (A1) and the intervening harmonics (H1, H2)
- Amplitudes of the spectral prominences below the first formant (P0) and above the first formant (P1)
- Differences in amplitude of the first formant and the spectral prominences (A1-P0, A1-P1)
- Fundamental frequency of the vowel (F0)
- Vowel duration

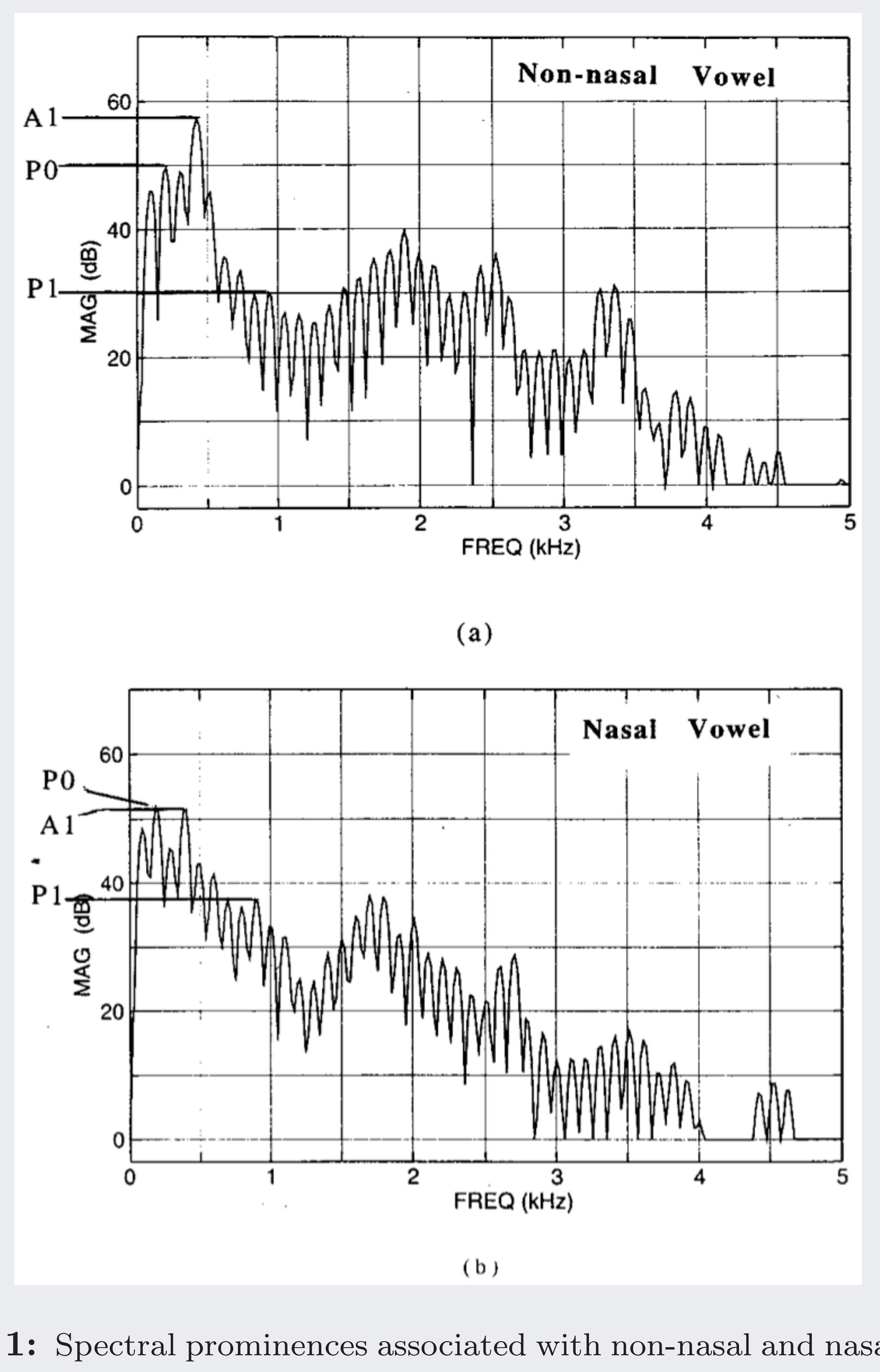

Figure 1: Spectral prominences associated with non-nasal and nasal vowels [4]

The Degree of Articulatory Constraints

- Phonetic segment assigned value based on closure/constriction formation
- Assignment based on participation of speech segment in closure/constriction formation
- Relative salience of coarticulatory effects of phonetic segments on each other
- Phonetic segment with more tongue-dorsum involvement: more articulatorily constrained
- Labial /m/; least coarticulatory resistance [10]

Predictions for vowel classification

Labials Minimal articulatory complexity should allow a more accurate retrieval of vowel targets.

Dentals Vowel targets predicted to be more variant and consequently a decrease in accuracy is expected.

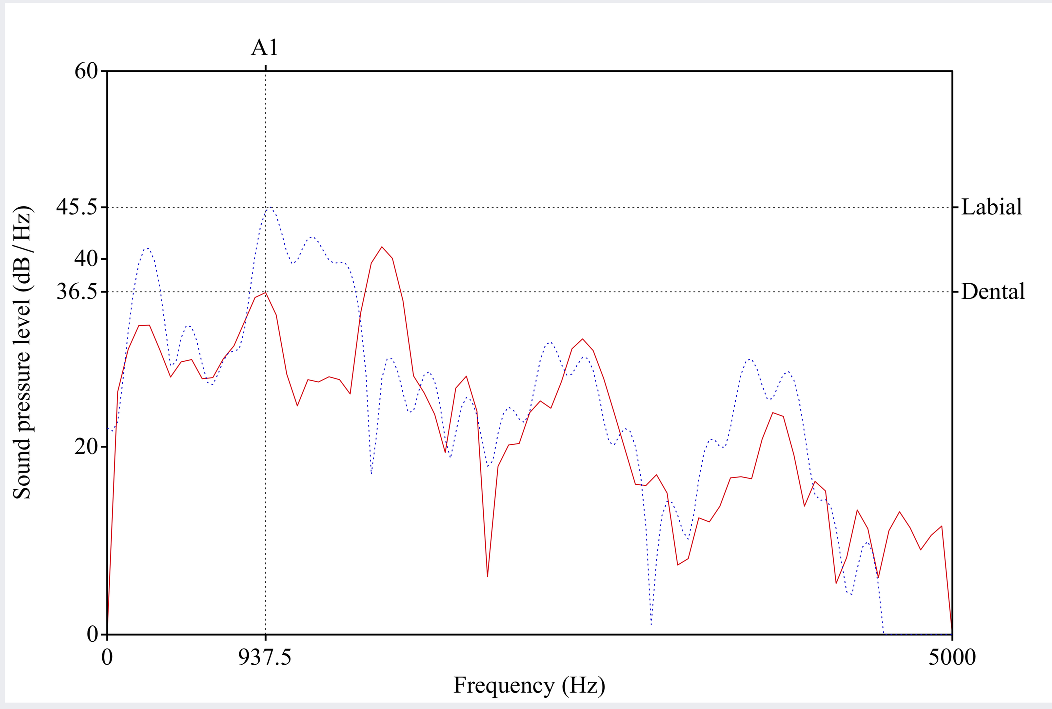

Figure 2: Coarticulatory effects for labials and dentals on vowel /a/. Amplitude of F1 for the spectrum coarticulatorily influenced by /n/ is considerably reduced. Spectral prominences around the F1 can also be observed.

Directionality and POA

- Directionality effects subject to investigation in the acoustics domain.
- Effects of nasalization reduce as vowel reaches its steady state.
- Anticipatory direction of nasalization lowers the mean amplitude in chunk 3.

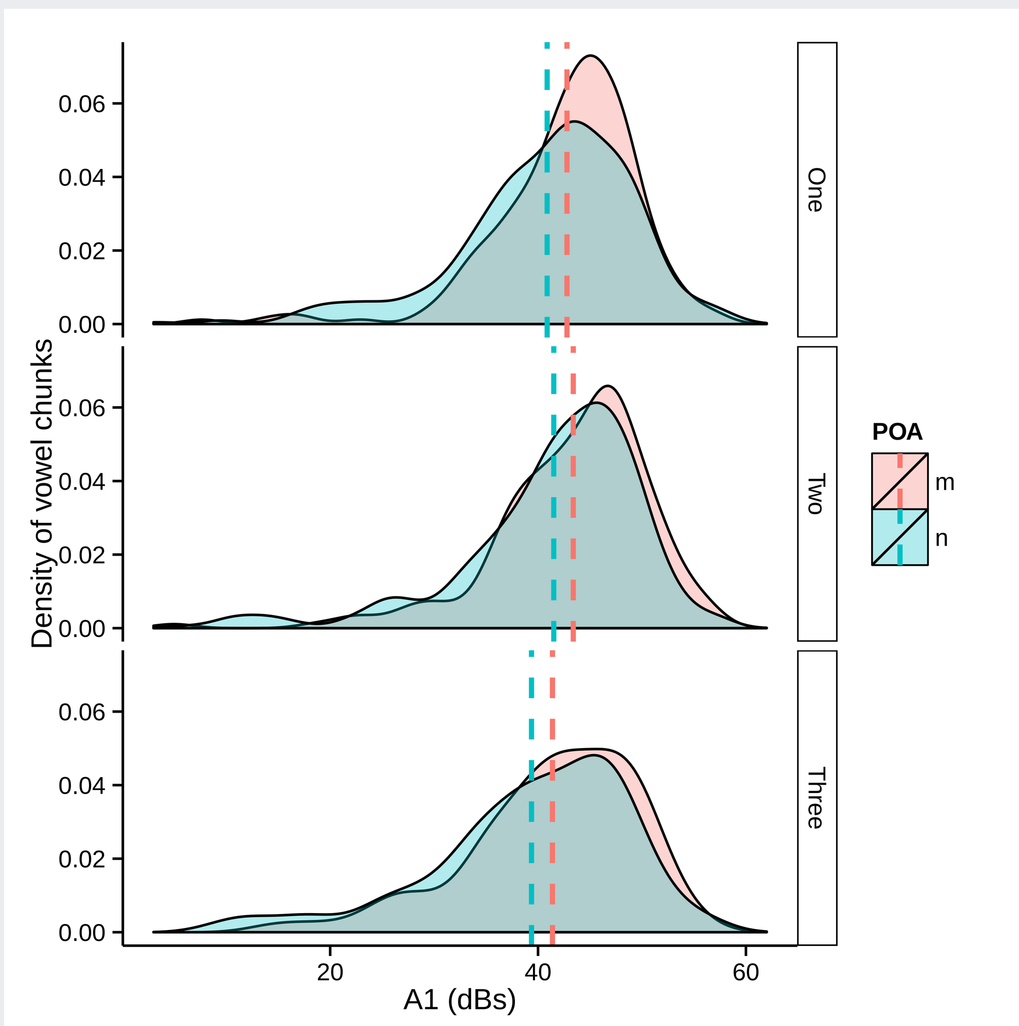

Figure 3: Nasalization introducing consonant-specific effects on the amplitude of the first formant. Labials consistently show higher mean values across the temporal chunks. The reduced amplitude of chunk 3 can be attributed to the anticipatory direction of coarticulation.

Effects of consonantal coarticulation

- The e1071 package in R is used to perform a linear SVM and Naïve Bayes classification [11].
- Modelling nasality specific features against vowel labels
- Dataset divided into 80% training and 20% test.
- Table 2 shows the classification accuracies for both SVM and Naïve Bayes classifiers.

Table 1: Comparative results of consonant context

Consonant	SVM	Naïve Bayes
Labials (L)	84.2% Cohen's κ 0.78	73.65% Cohen's κ 0.65
Dentals (D)	62.3% Cohen's κ 0.59	72.8% Cohen's κ 0.64

- Vowel classification accuracy is greater under Labial place of articulation compared to Dental with an SVM classifier
- Naïve Bayes classifier shows opposite results but the magnitude of difference in accuracy is smaller.

Effects of directionality

- Features obtained for 3 equi-durational temporal chunks of the vowel
- Training involved first 2 temporal chunks for carryover, and last 2 temporal chunks for anticipatory nasalization
- An 80-20% partition is used for training and test, respectively.

Table 2: Comparative results of directionality

Directionality	SVM	Naïve Bayes
Carryover (C)	94.8% Cohen's κ 0.93	69.2% Cohen's κ 0.6
Anticipatory (A)	87.5% Cohen's κ 0.83	85% Cohen's κ 0.81

- Vowel classification displays greater accuracy under carryover direction compared to anticipatory direction
- Naïve Bayes shows trends that contrast with the predictions of the DAC

Conclusions

What we find

- Vowel classification with respect to place of articulation shows predictable results
- Anticipatory direction shows lower accuracy in classification compared to carryover
- Predictions of DAC are largely borne out

Implications

- Lingual and velopharyngeal coarticulation interact to complicate the acoustic space, making identification of vowels under nasalization a complex problem.
- Machine learning algorithms, in this case, SVM and Naïve Bayes, are sensitive to acoustic variation, that are a consequence of articulatory constraints.
- Automatic classification of nasalized vowels, even in languages that maintain a contrast between oral-nasal vowels, needs to be parameterized within the feature selection process in order to improve accuracy of classification.
- Classification of nasalized vowels in languages with only coarticulatory nasalization would also need closer attention in terms of articulatory constraint parameterization.

References

[1] Tarun Pruthi and Carol Y. Espy-Wilson. Acoustic parameters for the automatic detection of vowel nasalization. In *INTERSPEECH*, pages 1925–1928. ISCA, 2007.

[2] Nishant Singhal and Pratip Das. Study of acoustic properties of nasal and non-nasal vowels in temporal domain. In *Proceedings of the International Conference on Signal, Image Processing and Pattern Recognition (SIPP-2013)*, pages 305–314, 2013.

[3] Bilal A. Raja. Recognition of Nasalized and Non-Nasalized Vowels. Master's thesis, University of Maryland, College Park, 2006.

[4] Marilyn Y. Chen. Acoustic correlates of english and french nasalized vowels. *The Journal of the Acoustical Society of America*, 102(4):2360–2370, 1997.

[5] Arthur S. House and Kenneth N. Stevens. Analog studies of the nasalization of vowels. *Journal of Speech and Hearing Disorders*, 21(2):218–232, 1956.

[6] Jiahong Yuan and Mark Liberman. Automatic measurement and comparison of vowel nasalization across languages. In *17th International Conference of Phonetic Sciences*, 2011.

[7] Yosvany Llerena Rodriguez and Antonio Teixeira. On the detection and classification of frames from European Portuguese oral and nasal vowels. *FALA*, 2010.

[8] Biswajit Das, Sandipan Mandal, and Pabitra Mitra. Bengali speech corpus for continuous automatic speech recognition system. In *Speech Database and Assessments (Oriental COCOSDA)*, pages 51–55, 2011.

[9] Paul Boersma. Praat, a system for doing phonetics by computer. *Glot International*, 5(9/10):341–345, 2001.

[10] Lo-Soun Su, K.-P. Li, and K. S. Fu. Identification of speakers by use of nasal coarticulation. *The Journal of the Acoustical Society of America*, 56(6), 1974.

[11] David Meyer, Evgenia Dimitriadou, Kurt Hornik, Andreas Weingessel, and Friedrich Leisch. *e1071: Misc Functions of the Department of Statistics (e1071), TU Wien*, 2014. R package version 1.6-3.