Acoustics of articulatory constraints: Vowel classification and nasalization

Indranil Dutta, Ayushi Pandey

Department of Computational Linguistics, The EFL University, Hyderabad, India
indranil@efuniversity.ac.in, ayuship.09@gmail.com

Abstract

Nasalization confounds the acoustic space by both crowding acoustic cues and also increasing the bandwidth of F1 which makes vowel identification, and hence classification by machine learning algorithms more complex. We present results from a set of machine learning (ML) algorithms that are trained on 15 acoustic features from the spectral and temporal domains to classify contextually nasalized vowels [1, 2]. The Degree of Articulatory Constraint (DAC) predicts that phonetic segments produced with lesser tongue-dorsum involvement are coarticulatorily more sensitive [3, 4]. We test the predictions of the DAC, using ML algorithms to classify vowels in the context of labials and dentals. Labials /m/ show more coarticulatory sensitivity as compared to dentals /n/. The ML algorithms do align themselves with the predictions of the DAC, showing 12% better acccuracy for identification of vowels in the context of labials than in the context of dentals. Based on the DAC predictions, labials will exert less coarticulatory influence on neighboring vowels, and hence the classification of vowels next to labials should be easier, compared to dentals. The Nasal-Vowel (NV) and Vowel-Nasal (VN) contexts also allow us to test the effects of anticipatory and carryover nasal coarticulation on vowel classification.

Index Terms: coarticulation, machine learning, degree of articulatory constraint

1. Introduction

Nasalization spreads across pre- and post nasal vowels in Bangla as a phonetic feature [5]. While Bangla exhibits contrastive as well as contextual (coarticulatory) nasalization in its vowels, in this paper we examine the effects of coarticulatory nasalization on Bangla vowels which is a consequence of velum lowering during the production of oral vowels. This results in relative nasalization or spread of nasalization in surrounding oral vowels. According to the Degree of Articulatory Constraint (DAC) model [3], varying degrees of coarticulation produce different formant transition effects in neighbouring vowels owing to the consonant place of articulation. Using several acoustic features, both from the spectral and temporal domain, trained on Support Vector Machines (SVM) and Naïve Bayes (NB) classifiers we report the results of identification of 6 vowels; /i/, /e, /a/, /o/, /ɔ/, /u/. We partition the data under two conditions. Under the first condition, we partition the data between dentals and labials to investigate the specific coarticulatory effects of different places of articulation of nasal consonants. Classification of vowels is conducted on these environments separately to investigate how a specific consonant place of articulation affects the identification of a vowel. Labials show better accuracy with SVM compared to dentals. Another partition is made with respect to the directionality of coarticulation. Vowel segments showing carryover coarticulation (NV) are separated from those showing anticipatory coarticulation (VN). SVM results show a reduced accuracy in the anticipatory direction (VN) as compared to the coarticulatory direction (NV). These results correspond to the theory of DAC in showing that effects of nasalization are seen more prominently in the offset of the vowel when there is a following nasal.

The organisation of the paper is as follows: Section 2 relates previous research on nasalization of vowels to our own concerns. In section 3, we discuss in detail the predictions borne out of the DAC [3]. In this section we also discuss the articulatory constraints, both for place of articulation and directionality, in some detail, that help us formulate the predictions for our study that are relevant for vowel classification. Section 4 details the materials and methods employed in this study and also the acoustic features on which the classifiers are trained, in addition to the annotation schemas and the process of feature extraction from a Bangla speech corpus. In section 5, we present the results from automatic classification based on several acoustic features of Bangla nasalized vowels. Here, we also briefly discuss how the prediction accuracies correspond to the predictions of the DAC model, underscoring the need to include articulatory constraints in automatic classification of nasalized vowels. In section 6, we discuss the implications that the DAC model has towards automatic classification of vowels based on the place of articulation of the contextual nasal and the directionality of nasalization, i.e., carryover and anticipatory.

2. Acoustics of Vowel Nasalization

Coarticulation is the process by which articulators involved in producing a conceptually isolated and discrete speech sound get activated either in anticipation of a following gesture or are impacted by articulatory constraints imposed by preceding segments. In the production of nasal consonants, the velum is lowered and some amount of air is allowed to pass through the nasal cavity. Unlike lingual gestures, such as the tongue tip and tongue body, the velic opening gesture is articulatorily found to be more resistant to neighbouring segments, especially in languages like Bangla, where a lexical contrast is found between nasal and oral vowels, quite like French [6] and Brazilian Portuguese [2].

X-ray cineradiography (physiological data evidence from the early 1900s) confirms that nasals have a contextual role to play in vowel production [7]. In the same way, vowels also exert coarticulatory influence on nasals, for instance in closed vowels the velum is higher than in open vowels [8]. Due to the coupling of the nasal and oral cavities in the production of nasalized vowels, there is a reduction of F1 amplitude and an increase in bandwidth around the F1 region [9]. It has also been shown that F1 amplitude gets reduced and nasal peaks are observed around

The authors are listed in alphabetical order.

the first formant. Crosslinguistically, differences in amplitude of the first formant (A1) and amplitudes of the nasal peaks are also observed below and above the first formant (P0 and P1)[6].

Within the context of examining how these various acoustic features associated with nasalization fare in the context of automatic classification, Pruthi et al. [1] propose a set of 9 features (tf9) as acoustic correlates of nasalization. SVM with an RBF and a Linear kernel is implemented on 3 databases (StoryDB, TIMIT and WS96/97). The proposed features (tf9) outperform the other feature sets (mf39 and gs6) in comparison, and give balanced results across nasal and oral vowels, although the accuracies reported are also database-dependent. Assesing the effect of lexical contrast on nasalization, Yuan and Liberman [10] propose that Portuguese (language with contrastive nasalization) has an independent nasalization target, but in English and Mandarin, this target is dependent on the neighbouring vowel. Berger 2007 [11] explores the articulatory dimension of nasalization and asserts that the degree of acoustic nasalization determines the extent of velum lowering.

Rodriguez et al. [12] implement 31 features for automatic detection and classification of the vowels without any linguistic (orthographic, annotating) information using an SVM, Naïve Credal Classifier and a MultiBoost Classifier. In addition to classifier performance, they measure a balanced number of examples for each class (oral and nasal) and the use of frames for nasal vowels. In their results, MultiBoost classifier outperforms Naïve Credal Classifier and the SVM. Using formant tracking techniques in Hindi, Singhal and Das [13] show the differences between nasalized and non-nasalized vowels. Similarly, using HMM models and mel-frequency cepstral coefficients (MFCC), it has been shown that vowel-recognition shows lower accuracy when the vowel is adjacent to the nasals [14]. Independently, acoustic phonetic studies establish how the acoustic space gets crowded, both spectrally and in terms of relative amplitudes of the spectral peaks, be they harmonic and/or formant [6, 9, 15, 2]. Concurrently, and more recently, several studies on automatic classification use an ensemble of machine learning algorithms to test the efficacy of various acoustic features sets in establishing the importance of few set of features over others [1, 11, 14, 13]. While these studies shed a great deal of light on the acoustics of nasalization and also how these acoustic features aid the automatic classification of naslization and nasals, in general, it is not clear how lingual and velopharyngeal coarticulation interact to confound the acoustic space, making identification of vowels under nasalization a complex problem. Especially, in the context of Bangla, where the velic opening gesture is regulated coarticulatorily, both in the production of oral-nasal vowels and contextual nasalization, hence reducing the extent of coarticulation in order to maintain the lexical contrast between oral and nasal vowels. In addition, gestural coordination of lingual and labial gestures along with velic opening and closing gestures provides an acoustic feature matrix where time varying articulatory constraints impose special restrictions.

3. The Degree of Articulatory Constraint

The Degree of Articulatory Constraint (DAC) is a model in which each phonetic segment (vowel and consonant) is given a specified value based on its involvement in closure/constriction formation [3, 4]. This scale has been tested in vowel and consonant dependent contexts to ascertain the relative salience of the coarticulatory effects of phonetic segments on each other. According to the predictions of the DAC, vowels coarticulate maximally with labials, because labials are minimally constrained. The dentals are given an intermediate level of articulatory constraint, and consequently are less sensitive to coarticulation. For example, the F2 for the vowel /u/ is much higher in the context of dental /n/ than it is in the context of the labial /m/. Dental /n/, being more constrained, will exert more coarticulatory influence on the vowel whereas the labial /m/ will not. Paired with varying articulatory constraints, are also language dependent coarticulatory constraints, where segmental inventory and nature of the contrast may also play a role in shaping the extent of coarticulatory effects [16, 17]. While the DAC predicts that labials would tend to exert less coarticulatory influence on the neighboring vowels, a consequence of such low resistance to coarticulation would be less variance in the acoustic vowel targets. This in turn would make classification of vowels within the labial context an easier task, compared to dentals. Additionally, language specific parameters such as the size of the segmental inventory and the density of contrasts would also determine how contextual and coarticulatory nasalization would confound the acoustic targets of a nasalized vowel. A crucial consequence of the latter parameter would be the effects on the directionality of coarticulatory effects; whether, in general greater in the carryover or anticipatory context. This issue is particularly relevant in the context of nasalization because of the interaction of velopharyngeal coarticulation with other articulatory gestures. In the context of V-to-V coarticulation it has been found that carryover effects are greater in extent compared to anticipatory [18], however, quite the opposite may hold in C-to-V and V-to-C coarticulation [3]. Here we propose to show that the articulatory constraints proposed in the DAC model are manifest in the acoustic variablility due to coarticulation which in turn have consequences for automatic classification of nasalized vowels.

4. Materials and Methods

This corpus consisted of 200 read-out sentences from 2 male and 2 female Bangla speakers from the larger Shruti Corpus, a continuous ASR speech corpus [19]. Each sound file was manually annotated using Praat [20]. A single-tier annotation schema was used where both the nasal place of articulation and the vowel identity were marked. The data was marked for 6 vowels /a/, /e/, /i/, /o/, /ɔ/, and /u/ in the context of two nasal consonants /m/ and /n/. Scripts in Praat were then used to extract 13 acoustic features. 2 categorical features showing segment labels and directionality context labels were also included. Each vowel was first split into 3 chunks. In the spectral domain 5 acoustic features were measured, namely, the amplitude of the first harmonic (H1), the amplitude of the second harmonic (H2), frequency of the first formant (F1), the nasal peak before the first formant (P0) and the nasal peak after the first formant (P1). Next, we calculate the differences in amplitude of the first formant, A1 and P0 (A1-P0) and also between A1 and P1 (A1-P1). In addition to these features, we also measure the total duration of the vowel portion. Categorical and predictor variables such as the label of directionality and segments were also present, thus giving a set of 15 features. Given the differences in the lingual gestures of the dental and labials, they are expected to show different coarticulation patterns. Consequently, to observe the different behaviour of the two nasals as given by the predictions of the DAC, we segmented the data between dentals and labials. Each of these segments was later to be passed to the classification algorithms. The chunking of the vowel duration was the main factor in observing patterns of directionality of coarticulation. For the carryover direction (NV) we needed to locate the effect of nasalization at the onset of the vowel. Therefore, we subsetted the data such that only the first two chunks were passed for classification. In a similar way, for the anticipatory direction of coarticulation, the duration needed to be measured at the offset of the vowel. Hence, we passed the acoustic features obtained in the last two chunks for classification.

5. Results

In this section, we present the results from the SVM and Naïve Bayes classifiers on the contextually nasalized vowel identification task. The tables display a matrix such that the predicted identities appear in rows and the the actual vowel identities in the columns. We provide diagonal accuracy percentages and the corresponding Cohen's kappa statistic that helps us interpret the strength of the classification accuracy of the vowels based on the acoustic features.

A Support Vector Machine (SVM) is a supervised learning discriminative classifier formally defined by a separating hyperplane. A hyperplane is obtained such that the line (in case of 2-D classification, hyperplane is a line) has the largest minimum distance from all the data points in a classification system. On the other hand, a Bayes classifier is a simple probabilistic classifier based on Bayes' theorem (from Bayesian statistics) with strong (naïve) independence assumptions. A more descriptive term for the underlying probability model would be "independent feature model". In simple terms, a Naïve Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature. We use the e1071 package in R to perform both the SVM classification and Naïve Bayes classification of contextually nasalized vowels on the speech dataset [21]. A dataframe consisting of the parameters discussed above was passed to the SVM and Naïve Bayes algorithms. We divide the corpus into training and test sets. An 80-20% partition is used for training and test, respectively. The training data contains the labels for the vowel identity but the test dataset does not. In the following subsections, we present the results from these two classifiers for vowel classification with respect to directionality and place of articulation features as discussed above.

5.1. Place of Articulation effects in Vowel Classification

Tables 1 and 2 display the classification results obtained from the data trained on dentals and labials. A diagonal accuracy is computed to test the accuracy of the classifier.

Table 1: *True and false positives for SVM classification - dentals*

	True					
Predictions	a	e	i	o	ɔ	u
a	13	0	0	5	3	1
e	1	18	0	3	0	0
i	0	2	21	1	0	2
o	2	1	0	5	1	0
ɔ	0	0	0	0	1	0
u	0	0	0	0	0	1

With the SVM classifier, we achieve a diagonal accuracy of 84.2% and a Cohen's kappa statistic of 0.78 in the case of labials. The dentals show a reduced, 62.3% diagonal accuracy and

Table 2: *True and false positives for SVM classification - labials*

	True					
Predictions	a	e	i	o	ɔ	u
a	14	0	0	5	3	1
e	0	1	0	3	0	0
i	0	0	10	0	0	2
o	0	1	0	2	0	0
ɔ	0	0	0	0	4	0
u	0	0	0	0	0	1

0.59 kappa statistic. NB classifier, on the other hand, achieves a diagonal accuracy of 73.65% and kappa of 0.65 in the case of labials and a 72.8% diagonal accuracy and a 0.64 kappa in the case of dentals. While the SVM classifer performs vowel classification with greater accuracy for labials, the NB classifier provides a reduced accuracy in the classification of labials compared to the SVM (tables 3 and 4). The reduced accuracy in vowel classification with dentals with an SVM could be attributed to greater coarticulatory effects on the vowel of the lingual gesture in addition to nasalization. Lingual and velopharyngeal coarticulation could be seen as impacting accuracies of vowel classification in the dentals, while the absence of the lingual gesture in labials facilitates the recovery of the vowel, compared to the dentals.

Table 3: *True and false positives for NB classification – dentals*

	True					
Predictions	a	e	i	o	ɔ	u
a	9	0	0	2	3	1
e	2	16	3	3	2	1
i	0	0	20	0	0	0
o	1	1	1	4	1	1
ɔ	2	0	0	3	2	0
u	0	0	2	1	0	0

Table 4: *True and false positives for NB classification – labials*

	True					
Predictions	a	e	i	o	ɔ	u
a	14	0	0	0	0	0
e	0	1	0	0	0	1
i	0	2	4	0	0	0
o	2	0	0	5	1	0
ɔ	0	0	0	1	1	0
u	0	1	0	0	0	2

It needs to be underscored here that general improvement in classification accuracy is observed in the context of the labials compared to the dental. The lingual gesture associated with dentals, along with velopharyngeal coarticulation in contextual nasalization, could potentially lower the classification accuracy of vowels in dentals. Machine learning algorithms are not immune to variablility in acoustic features brought about by articulatory complexity and this complexity needs to be parametrized, in term of operationalizing the DAC model in order to achieve better vowel classification in nasal contexts.

5.2. Directionality effects on Vowel Classification

Contextual nasalization, independent of place of articulation, remains a complex issue, especially when learning mechanisms are deployed to understand ways in which carryover and anticipatory coarticulation interact with articulatory gestures to yield acoustic variance [3, 4]. Here too, an SVM classifier outperfoms a Naïve Bayes classifier to an extent (compare tables 5 and 7, and tables 6 and 8). These data are trained on partitions with NV and VN contexts.

Table 5: *True and false positives for SVM classification - carryover*

	True					
Predictions	a	e	i	o	ɔ	u
a	10	0	0	0	0	0
e	0	6	0	0	0	0
i	0	1	11	1	0	0
o	0	0	0	8	0	0
ɔ	0	0	0	0	2	0
u	0	0	0	0	0	0

Table 6: *True and false positives for SVM classification - anticipatory)*

	True					
Predictions	a	e	i	o	ɔ	u
a	11	0	0	6	4	0
e	0	11	0	2	2	0
i	0	1	22	0	0	3
o	4	1	1	4	1	0
ɔ	2	0	0	1	2	0
u	0	1	0	1	1	0

With respect to directionality, the carryover direction (NV) context shows an accuracy of 94.8% (diagonal accuracy) and 0.93 kappa whereas in the anticipatory direction (VN) the diagonal accuracy of vowel classification is 87.5% and kappa is found to be 0.83 for the SVM. NB classification, however, shows a diagonal accuracy of 69.2% and kappa of 0.6 in the carryover direction whereas in the anticipatory direction (VN) the diagonal accuracy of vowel classification is 85% and kappa is found to be 0.81. While SVM shows that vowels are extractable with greater accuracy in the carryover context, NB shows quite the opposite, yet the SVM classifier outperforms the Naïve Bayes due to better handling of larger feature sets by SVMs, in general.

Table 7: *True and false positives for NB classification – carryover*

	True					
Predictions	a	e	i	o	ɔ	u
a	7	0	0	0	0	0
e	0	4	2	1	0	1
i	1	0	11	0	0	0
o	1	0	0	3	1	1
ɔ	0	1	0	3	1	0
u	0	0	0	0	0	1

Table 8: *True and false positives for NB classification – anticipatory*

	True					
Predictions	a	e	i	o	ɔ	u
a	10	1	0	2	1	0
e	0	9	0	0	1	0
i	0	0	5	0	0	0
o	0	1	0	4	0	0
ɔ	0	0	0	0	4	0
u	0	0	0	0	0	2

6. Conclusions and further Research

Understanding the acoustics of nasalization, independently, has involved identification of several features such as the amplitude of the first harmonic (H1), the amplitude of the second harmonic (H2), frequency of the first formant (F1), the nasal peak before the first formant (P0) and the nasal peak after the first formant (P1), in addition to relative spectral intensity measures such as the differences in amplitude of the first formant, A1 and P0 (A1-P0) and also between A1 and P1 (A1-P1). Studies on nasal coarticulation, especially, those on languages that contrast oral and nasal vowels, have also helped further refine the nature and extent of nasalization on vowels. Concomitantly, studies on articulatory constraints, and especially the DAC model have helped advance our understanding of the nature of both lingual and velopharyngeal coarticulation. Training a set of these spectral features on two separate automatic classification algorithms we show that predictions from the Degree of Articulatory Constraint (DAC) model are largely borne out, in that, vowels in the context of labials show least coarticulatory resistance, and consequently show greater classification accuracy. Directionality of coarticulatory nasalization also shows effects on classification of vowels in nasal contexts, with greater accuracy in the carryover (NV) context compared to the anticipatory context (VN). We propose that automatic classification of nasalized vowels, even in languages that maintain a lexical contrast between nasal and oral vowels, needs to be parameterized within the feature selection process in order to improve accuracy of classification.

7. References

[1] T. Pruthi and C. Y. Espy-wilson, "Acoustic parameters for the automatic detection of vowel nasalization," in *INTERSPEECH 2007*, 2007.

[2] A. A. Bailey, "The Nose knows: Differences in Nasal and Nasalized vowels as produced by native and l3 speakers of Brazilian Portuguese," in *Proceedings of Phonetics and Phonology in Iberia (PaPI)*, 2013.

[3] D. Recasens, M. D. Pallarès, and J. Fontdevila, "A model of lingual coarticulation based on articulatory constraints," *Journal of the Acoustical Society of America*, vol. 102, pp. 544–561, 1997.

[4] D. Recasens and A. Espinosa, "An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan," *Journal of the Acoustical Society of America*, vol. 125, pp. 2288–2298, 2009.

[5] G. Cardona and D. Jain, Eds., *The Indo-Aryan Languages*. London: Routledge, 2003.

[6] M. Y. Chen, "Acoustic correlates of english and french nasalized vowels," *The Journal of the Acoustical Society of America*, vol. 102, no. 4, pp. 2360–2370, 1997.

[7] W. Hardcastle and N. Hewlett, *Coarticulation: Theory, Data and Techniques*. Cambridge University Press, 1999.

[8] F. Bell-Berti, "The velopharyngeal mechanism : an electromyographic study," in *Haskins Laboratory Status Report on Speech Research, Supplement*, 1973, pp. 117–129.

[9] A. S. House and K. N. Stevens, "Analog studies of the nasalization of vowels," *Journal of Speech and Hearing Disorders*, vol. 21, no. 2, pp. 218–232, 1956.

[10] J. Yuan and M. Liberman, "Automatic measurement and comparison of vowel nasalization across languages," in *17th International Conference of Phonetic Sciences*.

[11] M. A. Berger, "Measurement of vowel nasalization by multi-dimensional acoustic analysis," Master's thesis, University of Rochester, Rochester, New York, 2007.

[12] Y. L. Rodriguez and A. Teixeira, "On the detection and classification of frames from european portuguese oral and nasal vowels," *FALA*, 2010.

[13] N. Singhal and P. 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)*, 2013, pp. 305–314.

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

[15] S. Maeda, "Acoustics of vowel nasalization and articulatory shifts in French nasal vowels," in *Phonetics and phonology: Vol. 5: Nasals, nasalization and the velum*, M. K. Huffman and R. A. Krakow, Eds. Academic Press, 1993.

[16] S. Manuel, "The role of contrast in limiting vowel-to-vowel coarticulation indifferent languages," *Journal of the Acoustical Society of America*, vol. 88, pp. 1286–1298, 1990.

[17] ——, "Cross-language studies: Relating language-particular coarticulation patterns to other language-particular facts," in *Coarticulation: Theory, data and techniques*, W. Hardcastle and N. Hewlett, Eds. Cambridge University Press, 1999.

[18] F. Bell-Berti and K. S. Harris, "Some aspects of coarticulation," in *8th International Congress of the Phonetic Sciences*.

[19] B. Das, S. Mandal, and P. Mitra, "Bengali speech corpus for continuous automatic speech recognition system," in *Speech Database and Assessments (Oriental COCOSDA)*, 2011, pp. 51–55.

[20] P. Boersma, "Praat, a system for doing phonetics by computer." *Glot International*, vol. 5, no. 9/10, pp. 341–345, 2001.

[21] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel, and F. Leisch, *e1071: Misc Functions of the Department of Statistics (e1071), TU Wien*, 2014, r package version 1.6-3. [Online]. Available: http://CRAN.R-project.org/package=e1071