AUTOMATIC MEASUREMENT AND COMPARISON OF VOWEL NASALIZATION ACROSS LANGUAGES

Jiahong Yuan & Mark Liberman

University of Pennsylvania, USA
jiahong@ling.upenn.edu; myl@cis.upenn.edu

ABSTRACT

We used an innovative technique to measure the trajectories of vowel nasality in three languages: English, Mandarin, and Portuguese. An SVM classifier was trained on MFCC parameters at five positions in nasal and oral vowels, and the "nasality trajectory" for each language was evaluated in terms of classification accuracy at each of the five positions. The results support the view that Portuguese vowel nasality is phonemic while English and Mandarin vowel nasality is coarticulatory, although in all cases vowel nasalization starts at the vowel onset or even earlier.

Keywords: nasalization, nasality trajectory, corpus phonetics, MFCC, SVM

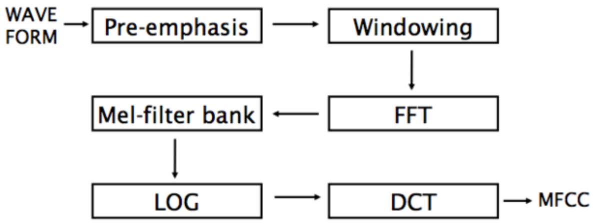
1. INTRODUCTION

Vowel nasalization is the production of a vowel while the velum is lowered and the velopharyngeal port is open, so that the nasal cavities are coupled into the vocal-tract resonance system. Vowels are necessarily nasalized to some extent after or (especially) before a nasal consonant, due to the time required to lower and raise the velum. In many languages of the world, such as French, Portuguese and Yoruba, nasalization is a contrastive feature of vowels, independent of the presence of a nasal consonant.

The production of nasalized vowels introduces nasal resonances and anti-resonances, which interact with the resonances of the oral cavity to produce a complex spectral transfer function. Many acoustic parameters have been found to be related to nasalization, including a reduction in amplitude of the first formant (A1) [17]; the relationship between A1 and the amplitude of the first harmonic (H1) [19]; nasal poles, one below F1 (P0), at around 250-450Hz, and the other above F1 (P1), at around 1000 Hz [22]; the difference between A1 and P0, and the difference between A1 and P1 [6]; nasal pole-zero pairs in the vicinity of the first formant [10, 16]; and a low-frequency center of gravity [2].

As this large number of relevant acoustic parameters suggests, the acoustic consequences of nasalization are complex, and there is no single measurable acoustic correlate of nasality [12]. The time-function of vowel nasalization is conditioned by many factors such as vowel height [3, 14, 26], phonetic context [5, 20], the phonemic or allophonic status of vowel nasality [8, 25], and the speaker and language characteristics [7, 13]. For example, it was found that when high vowels are nasalized, the velum is not lowered to the same extent as when low vowels are nasalized [3], but on the other hand, the nasal flow is greater in high vowel nasalization than low vowel nasalization [26]. Lower vowels require more velopharyngeal coupling to be perceived as nasal [1].

It is desirable to develop a measure of nasality that is robust enough to be used to compare the degree of nasality in different vowels or different languages. One possibility is to apply dimensionality-reduction techniques such as Principal Components Analysis to relevant acoustic parameters, as in [4]. However, these parameters are difficult to measure automatically and accurately, and many of them are vowel dependent. As an alternative, we can start with lower-level and more reliable acoustic parameters such as Mel-frequency cepstral coefficients (MFCC). The MFCC coefficients, widely used in automatic speech and speaker recognition, are based on a linear cosine transform (DCT) of a log power spectrum. The coefficients can be computed automatically through the following procedure:



MFCC parameters have been used in automatic detection of vowel nasality [15, 24, 27], and the results were comparable to or even better than using the higher-level acoustic-phonetic parameters. In [27], for example, the standard MFCC coefficients were extracted at the center of the vowels, and a SVM classifier was built to discriminate between oral and nasalized vowels in a vowel-independent manner. When trained on the TIMIT training set and tested on its test set, the method achieved 88% overall accuracy.

In this study, we investigated the use of MFCC [11] and SVM [9] to study the overall patterns of nasality in large sets of vowel tokens. We assume that vowel regions that are more strongly nasalized, at least in acoustic terms, will be better discriminated from oral-vowel regions by a SVM classifier. In other words, we characterize the "degree of acoustic nasality" of a set of vocalic regions in terms of the proportion of these tokens that are correctly classified as "nasal" vs. "oral" by an SVM classifier.

We used this method to compare vowel nasality in American English, Mandarin Chinese, and Brazilian Portuguese. In the following sections we first introduce the data sets, then we describe our method. The results are presented and discussed in Section 3, followed by conclusions in Section 4.

2. DATA AND METHOD

Three speech corpora were used for this study: the 1997 English Broadcast News Speech (*LDC98S71*), the 1997 Mandarin Broadcast News Speech (*LDC98S73*), and the West Point Brazilian Portuguese Speech (*LDC2008S04*). The broadcast news corpora in English and Mandarin are recordings of broadcasts from television and radio networks. The West Point corpus contains read sentences from multiple speakers. The phone boundaries were automatically determined by using the Penn Phonetics Lab Forced Aligner [23].

Equal numbers of nasal and oral vowels were randomly selected from these corpora to form balanced data sets for the experiment. Portuguese has contrastive nasal vowels. For English and Mandarin, the vowels before a nasal coda were treated as nasalized vowels whereas the others were treated as oral vowels. The data sets are summarized below:

- *English broadcast news speech:*
  Vowels: /ɪ ɛ æ ʌ ɑ ɪ̃ ɛ̃ æ̃ ʌ̃ ɑ̃/
  Training set: 5000 tokens, 500 per vowel
  Test set: 2000 tokens, 200 per vowel
- *Mandarin broadcast news speech:*
  Vowels: /i ɛ ə a ɒ o ĩ ɛ̃ ə̃ ã ɒ̃ õ/
  Training set: 7200 tokens, 600 per vowel
  Test set: 3600 tokens, 300 per vowel
- *Brazilian Portuguese read speech:*
  Vowels: /i e ɐ o u ĩ ẽ ɐ̃ õ ũ/
  Training set: 6000 tokens, 600 per vowel
  Test set: 3000 tokens, 300 per vowel

To investigate the trajectory of vowel nasality, we trained classifiers for each language and for each of five different positions within a vowel – 10%, 30%, 50%, 70%, and 90% of the vowel duration from the onset – by using the MFCC coefficients extracted at those positions from all the nasal and oral vowels in the training data. The 3x5=15 classifiers were speaker- and vowel-independent.

These classifiers were then applied both to the training data and to the test data (set aside and not used for training). The English classifier trained for the 10% position was applied to the coefficients extracted at the 10% position for all English vowels, and similarly for the other 14 classifiers.

The classification accuracies for the vowels in a given language were used to measure their degree of nasality at the five sample points. Vowels that are more strongly nasalized, in acoustic terms, should be more different from oral vowels, and thus all the vowels should be more likely to be classified correctly as "nasal" or "oral". 50% accuracy for a given set of vowels would mean that the classifier cannot do better than chance at discriminating nasal from oral vowels, suggesting that this set of "nasalized" vowels possesses no reliable acoustic indications of nasality (or at least none that this method can find in the MFCC spectra).

The MFCC coefficients were extracted using the HTK Toolkit [18]. The libsvm package of R [21] was used for classification.

3. RESULTS AND DISCUSSION

Figure 1 shows the classification accuracies at different vowel positions. We can see that accuracies are higher for Portuguese than for English and Mandarin except at the last position, which is near the end of the vowel and adjacent to the nasal coda. Moreover, the "nasality trajectory", is different between the phonemic nasalization in

Portuguese and the coarticulatory nasalization in English and Mandarin.

This is consistent with the common-sense view that Portuguese nasal vowels have their own nasality target, while the nasalized vowels in English and Mandarin are nasalized simply by virtue of coarticuation with a nasal coda consonant. There has been a debate about whether nasalization in American English is a result of a phonological rule [25] or not [8]. Our results favor the latter hypothesis: in English and Mandarin, vowels are not phonologically targeted as nasalized; instead, vowel nasalization is purely a transition a closed velopharyngeal port to its peak opening in the nasal consonant following the vowel.

The classification accuracy at the '0.1' position is well above 50% even for English and Mandarin, suggesting that vowel nasalization starts at the vowel onset in these languages. There is also a bit more nasality at the start of the vowel in Mandarin than in English.

Figure 2 shows the accuracies for high and low vowels at the vowel center position. Only the results for the test set are shown.

From Figure 2 we can see that in English and Mandarin high and low vowels have similar degree of nasality at the center of the vowels, whereas in Portuguese, the nasality in high vowels is higher than that in low vowels. The nasality difference between high and low vowels in Portuguese may be a phonological target, because there is no such difference in English and Mandarin where vowel nasalization is not targeted. The difference may, however, also be a pure physiological effect that only appears when the opening of the velopharyngeal port reaches its peak determined by the vowel height.

4. CONCLUSIONS

Many large speech corpora are now accessible, as well as new tools and methods for large-scale analysis. Our study shows that we can make use of technologies from speech recognition and machine learning in phonetically-relevant studies of these large speech corpora.

We investigated the use of MFCC and SVM for automatic measurement and comparison of vowel nasalization in American English, Brazilian Portuguese, and Mandarin Chinese. The results show that both the "nasality trajectory" and the effect of vowel height are different between the phonemic nasalization in Portuguese and the coarticulatory nasalization in English and Mandarin, suggesting that nasalized vowels are not targeted as nasal in English and Mandarin. The results also suggest a small difference between English and Mandarin, especially at the onset of vowels.

Figure 1: Classification accuracies at different vowel positions. The higher the accuracy, the higher degree of nasality the nasalized vowels possess. 50% accuracy represents no nasality.

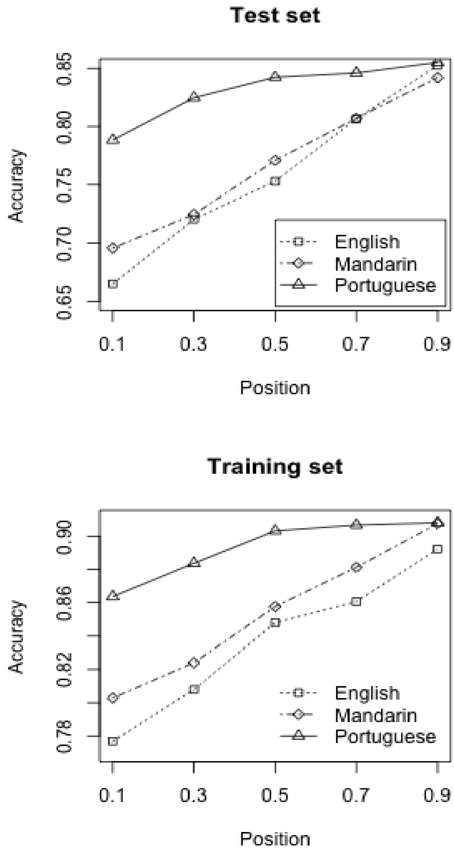
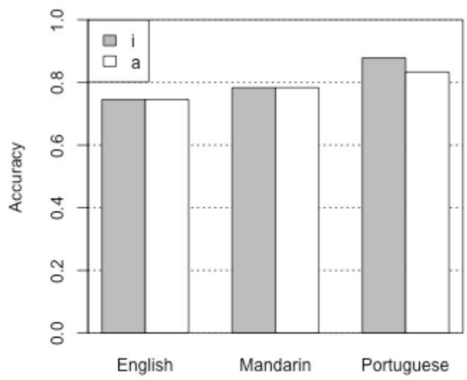


Figure 2: Classification accuracies of high and low vowels. The higher the accuracy, the higher degree of nasality the nasalized vowels possess. 50% accuracy represents no (detected) nasality.



5. ACKNOWLEDGEMENTS

This work is supported in part by NSF grant IIS-0964556.

6. REFERENCES

[1] Abramson, A.S., Nye P.W., Henderson J.B., Marshall C.W. 1981. Vowel height and the perception of consonant nasality. *J Acoust Soc Am.* 70, 329-339.

[2] Beddor, P.S. 1982. *Phonological and Phonetic Effects of Nasalization on Vowel Height*. Ph.D. dissertation, University of Minnesota.

[3] Bell-Berti, F. 1993. Understanding velic motor control: studies of segmental context. In Huffman, M.K., Krakow R.A. (eds.), *Nasals, Nasalization, and the Velum*. New York: Academic Press, 63-85.

[4] Berger, M.A., 2007. *Measurement of Vowel Nasalization by Multi-dimensional Acoustic Analysis*. Master thesis. University of Rochester, NY.

[5] Busà, M.G. 2007. Coarticulatory nasalization and phonological developments. In Solé, M.J., Beddor, P.S., Ohala M. (eds.), *Experimental Approaches to Phonology*. Oxford: OUP, 155-174.

[6] Chen, M.Y. 1997. Acoustic correlates of English and French nasalized vowels. *J. Acoust. Soc. Am.* 102, 2360-2370.

[7] Clumeck, H. 1976. Patterns of soft palate movements in six languages. *J. Phonetics* 4, 337-351.

[8] Cohn, A. 1990. Phonetic and phonological rules of nasalization. *UCLA WPP* 76, 1-224.

[9] Cortes, C., Vapnik, V. 1995. Support-vector networks. *Machine Learning* 20, 273-297.

[10] Dang, J., Honda, K., Suzuki, H., 1994. Morphological and acoustical analysis of the nasal and the paranasal cavities. *J. Acoust. Soc. Am.* 96, 2088-2100.

[11] Davis, S., Mermelstein, P. 1980. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics, Speech and Signal Processing* 28, 357-366.

[12] Glass, J.R. 1984. *Nasal Consonants and Nasalised Vowels: An Acoustical Study and Recognition Experiment*. Master thesis, MIT.

[13] Ha, S., Kuehn, D. 2006. Temporal characteristics of nasalization in children and adult speakers of American English and Korean during production of three vowel contexts. *J. Acoust. Soc. Am.* 120, 1622-1630.

[14] Hajek, J., Maeda, S. 2000. Investigating universals of sound change: The effect of vowel height and duration on the development of distinctive nasalization. In Broe, M., Pierrehumbert J. (eds.), *Papers in Laboratory Phonology V*, Cambridge: Cambridge University Press, 52-69.

[15] Hasegawa-Johnson, M., Baker, J., Borys, S., Chen, K., Coogan, E., Greenberg, S., Juneja, A., Kirchoff, K., Livescu, K., Mohan, S., Muller, J., Sonmez, K., Wang, T. 2004. *Landmark-based Speech Recognition: Report of the 2004 Johns Hopkins Summer Workshop*. Johns Hopkins University 2004 Summer Workshop final report.

[16] Hawkins, S., Stevens, K.N. 1985. Acoustic and perceptual correlates of the nonnasal-nasal distinction for vowels. *J. Acoust. Soc. Am.* 77, 1560-1575.

[17] House, A.S., Stevens, K.N. 1956. Analog studies of the nasalization of vowels. *J. of Speech and Hearing Disorders* 21, 218-232.

[18] HTK: *http://htk.eng.cam.ac.uk/*

[19] Huffman, M.K. 1990. Implementation of nasal: Timing and articulatory landmarks. *UCLA WPP* 75, 112-143.

[20] Krakow, R.A. 1993. Nonsegmental influences on velum movement patterns: syllables, sentences, stress, and speaking rate. In Huffman, M.K., Krakow R.A. (eds.), *Nasals, Nasalization, and the Velum*. New York: Academic Press, 87-113.

[21] libsvm: *http://www.csie.ntu.edu.tw/~cjlin/libsvm/*

[22] Maeda, S. 1982. The role of the sinus cavities in the production of nasal vowels. *Proceedings of ICASSP 1982*, 911-914.

[23] Penn Phonetics Lab Forced Aligner: *http://www.ling.upenn.edu/phonetics/p2fa/*

[24] Pruthi, T. 2007. *Analysis, Vocal-tract Modeling, and Automatic Detection of Vowel Nasalization*. Ph.D. thesis, University of Maryland, College Park.

[25] Solé, M.J. 1995. Spatio-temporal patterns of velopharyngeal action in phonetic and phonological nasalization. *Language and Speech* 38, 1-23.

[26] Young, L.H., Zajac, D.J., Mayo, R., Hooper, C.R. 2001. Effects of vowel height and vocal intensity on anticipatory nasal airflow in individuals with normal speech. *J. Speech Lang. Hear. Res.* 44, 52-60.

[27] Yuan, J., Seidl, A., Cristia, A. 2010. Automatic detection and comparison of vowel nasalization in American English. *J. Acoust. Soc. Am.* 128, 2291.