VOWEL NASALIZATION IN AMERICAN ENGLISH: ACOUSTIC VARIABILITY DUE TO PHONETIC CONTEXT

Nancy F. Chen[1,2], Janet L. Slifka[1,2], Kenneth N. Stevens[1,2]

[1]Speech & Hearing Bioscience & Technology Program, Harvard-MIT Division of Health Sciences and Technology. [2]Speech Communication Group, Research Laboratory of Electronics, MIT.
nancyc@mit.edu, slifka@mit.edu, stevens@speech.mit.edu

ABSTRACT

This study quantifies acoustic variation of vowel nasalization arising from phonetic context in American English with an emphasis on carryover contexts. While qualitative articulatory trajectories and phonetic descriptions suggest that a vowel is nasalized in carryover contexts, few acoustic studies have examined this issue. Our acoustic analyses investigate the vowel /i/ and show that: (1) a vowel can be nasalized with at least one adjacent nasal consonant, even if the nasal consonant is pre-vocalic; (2) vowels with nasal consonants on both sides (NVN) do not guarantee more vowel nasalization.

Keywords: vowel nasalization, acoustic analysis

1. INTRODUCTION

Over 99% of languages contain nasalized vowels or consonants [1]. Coarticulatory vowel nasalization occurs in virtually all languages [2], and is prevalent in conversational American English, where it may be the only information remaining to indicate the presence of a nasal consonant [3]. Though humans are sensitive to nasalized sounds, automated speech recognizers perform poorly when it comes to nasalized vowels [3]. Research has been done on many aspects of vowel nasalization, including acoustics [4], perception [2], and physiology [5, 6].

Nasalized vowels are unique, because they are the only vowels where air flows through two channels and radiates from the nose and mouth. During vowel nasalization, the open vocal tract is coupled with the nasal cavity, introducing additional pole-zero pairs in the transfer function. Some of the acoustical effects from this coupling are the decrease of A1 (dB), amplitude of the first formant F1, and the emergence of a nasal pole (with amplitude P1 in dB) near 1kHz [4]. The difference in these two values, A1-P1, is proposed to be an acoustic correlate of vowel nasalization [4].

Quantifying vowel nasalization is challenging, because it depends on many factors such as inter-subject anatomical differences, velar coupling area, vowel identity, and phonetic context. For example, Pruthi [7] showed that even within the same subject,

the interaction among the pole-zero pairs is complicated by oral cavity configuration, arising from changes in coupling area due to velar movement or vowel identity.

Phonetic descriptions indicate that vowel nasalization occurs more often and for a longer duration in anticipatory contexts (i.e., in vowel-nasal sequences) than in carryover contexts (i.e., in nasal-vowel sequences) [5]. In an articulatory study on nasalization, Krakow measured velic motion and found a greater degree of velum lowering in stressed and high-speaking-rate conditions [6]. Krakow also showed that the velum is lowered during vowels preceded by nasal consonants. This closing velic movement rate is approximately up to 1.6 times more rapid than the opening movement in vowels followed by nasal consonants, implying that vowels may be more nasalized in pre-nasal positions [6].

Though these articulatory and phonetic studies suggest that nasalization exists in vowels preceded by nasal consonants, some language pedagogy textbooks claim there is no vowel nasalization at all in such phonetic contexts (e.g. [8]). In addition, most acoustical studies on vowel nasalization focus more on anticipatory nasalization than carryover nasalization, so acoustic evidence of carryover nasalization remains limited. The present study quantitatively examines acoustic parameters of nasality in vowels preceded and/or followed by nasal consonants. In particular, we investigate two questions: (1) Is there quantitative acoustical evidence of nasalization in vowels with pre-vocalic nasal contexts? (2) From an acoustical standpoint, does the degree and extent of vowel nasalization increase with the number (0-CVC; 1-NVC, CVN; 2-NVN) of adjacent nasal consonants? This acoustic quantification of nasality from phonetic context can expand our knowledge in phonetic sciences, and potentially be applied to automated speech recognition and speech pathology diagnosis.

2. MATERIALS AND METHODS

Using a read corpus of 150 words, containing vowels in various phonetic contexts, acoustic

parameters were automatically extracted from vowels that had been hand-labeled. Statistical tests (Lilliefors and permutation tests) were used to analyze the acoustic parameters.

2.1.   Speech Corpus

A total of 900 tokens containing the vowel /i/ were collected from six male native speakers of American English. The vowel /i/ was chosen as a first effort because: (1) given a velopharyngeal area, the effect of acoustical coupling is stronger with high vowels than low vowels [5]; (2) high oral vowels next to nasal consonants are nasalized for longer durations than low vowels in the same context [5]; (3) the nasal pole near 1 kHz in /i/ is relatively easy to identify, since it is less affected by F1 and F2, which is typically located some distance from 1kHz for an adult male speaker.

Target words are divided into three groups where the vowel /i/ is in contexts of (1) NVN: surrounded by nasal consonants on both sides (e.g., *mean*); (2) CVN: followed by a nasal consonant (e.g., *team*); and (3) NVC: preceded by a nasal consonant (e.g., *neat*). The vowel of interest is always in a primary stress position. Velar nasal consonants were not included because there are no English words that begin with it. In (2) and (3), the non-nasal adjacent consonant of the vowel is always an obstruent. While 55% of these obstruent consonants are unvoiced, the rest are voiced both in (2) and (3). The control group (CVC) contains words with /i/ surrounded by obstruent consonants. For both these pre- and post-vocalic obstruent consonants, two-thirds are unvoiced and the rest are voiced respectively. The balance among place of articulation of the consonants was not complete due to the limitation of only using real words. The closures and releases of consonants adjacent to the vowel of interest were hand-labeled from a wideband speech spectrogram.

The full set of target words were read in randomized order by each speaker and embedded in the carrier phrase: "Say the word '_____,' please." The speech corpus was recorded in a sound booth with a microphone approximately six inches away from the subject's nose and mouth.

2.2.   Acoustical Parameters

Chen showed that A1-P1 in vowels of NVN differs from those of CVC [4]. While other acoustic correlates of nasality have been proposed (for a summary, see [7]), A1-P1 is adopted in this study. Note that since A1 decreases and P1 increases during vowel nasalization, A1-P1 is expected to be smaller when a vowel is more nasalized.

Though other methods [7] can be used to extract A1-P1, this study focuses on extracting A1-P1 from the spectrum. Spectra of time waveforms were generated by applying 20ms Hamming windows and computing 512-point fast Fourier transforms at 10, 20, 30, 40, and 50ms from the closures or releases of consonants adjacent to the vowel of interest. The durations of the vowels range from 100-250ms. A1-P1 was extracted automatically by adapting criteria in [9]: A1 is measured on the largest harmonic between 300-900Hz, while the nasal resonance amplitude P1 is measured on the largest harmonic between 770-1500Hz.

2.3.   Statistical Tests

T-tests are widely used in testing statistical difference of two populations, yet t-tests assume normally-distributed samples, which is not always true. Hence, this study tests the normal distribution assumption before further acoustic analyses.

Lilliefors tests determine whether the probability distributions of the data are Gaussian [10]. Lilliefors is a relatively weak test, requiring a large number of samples to reject the null hypothesis of normal distribution. Therefore, if the Lilliefors test rejects the null hypothesis, the distribution is very likely not normal, and t-tests are no longer appropriate to use. Instead, non-parametric statistical tests such as permutation tests [11] can be used since they make no assumptions about the probability distributions. Note that even if Lilliefors test does not reject the null hypothesis, it is still possible that the data are not normally-distributed. In addition, permutation tests can be used to test the difference of any statistical parameter, be it mean, median, or variance. In this study, the difference in 'mean' is used.

The Lilliefors test was performed respectively on the subgroups of measurements of A1-P1 at each time-point of each phonetic context for each individual speaker. At least 48% of the probabilistic distributions of the measurements were shown to not be Gaussian, indicating that non-parametric statistical tests are more suitable for acoustic analyses. Permutation tests were thus used to determine statistical significance.

3.   RESULTS AND DISCUSSION

3.1.   Vowels with and without nasal contexts

This subsection investigates whether A1-P1 shows a statistical difference in vowels with and without adjacent nasal consonants.

Hypothesis A: There is a statistical difference in the means of A1-P1 for vowels with adjacent nasal consonants (NVN, NVC, and CVN) vs. vowels with no adjacent nasal consonants (CVC).

The statistical results of testing Hypothesis A are:

(1) A1-P1 in NVN differs ($p<0.001$) from that in CVC for all time-points, for all speakers and in each individual speaker.

(2) A1-P1 in CVN differs ($p<0.001$) from that in CVC for all time-points, for all speakers and for each individual speaker.

(3) A1-P1 in NVC differs ($p<0.001$) from that in CVC for all time-points, for all speakers and for 5 out of 6 subjects when tested individually.

The mean and the lower and upper quartiles of A1-P1 over the initial and final 50ms of the vowel are shown in Figure 1. Chen's work is verified in Result (1) without any assumptions of the probability distributions of the data. Furthermore, Results (2) and (3) show quantitative acoustical evidence that vowels are nasalized when there is only one adjacent nasal consonant, even when in NVC context (at least for the vowel /i/). Subject 1 is the only subject not showing statistical difference in the means of A1-P1 between NVC and CVC. Further analyses of the tokens from Subject 1 indicate that if median were used as the statistical parameter (instead of mean) in permutation tests, then A1-P1 is statistically lower in NVC than in CVC at 30, 40, and 50ms from the consonant release. Since A1-P1 measurements are not always normally distributed, the statistical parameter 'median' provides different information regarding the spread of the data-points. The median of A1-P1 in NVC and in CVC being different suggests that though the average behavior (mean) of the two groups is the same, the distributions of A1-P1 are different; the NVC distribution is skewed toward lower values, which is what is expected when vowels are nasalized. Techniques which provide higher spectral resolution [7] A1-P1 can be used to further determine whether Subject 1's results are due to methodological limitations or individual speaker differences.

3.2. Relationship between vowel nasalization and the number of adjacent nasal consonants

In 3.1, we established that a vowel is nasalized when it has at least one adjacent nasal consonant. In section 3.2, we further hypothesize that the degree and extent of vowel nasalization correlates with the number of adjacent nasal consonants.

Hypothesis B: The degree and extent of vowel nasalization increases with the number of adjacent nasal consonants:
 **B.1 A1-P1: CVC>NVC>NVN**
  B.2 A1-P1: CVC>CVN>NVN

Hypothesis B.1: The results reported here are taken from the initial 50ms of the vowel. The trends shown in Figure 1 generally correspond with Hypothesis B.1. Figure 1 (a) shows that for data combined across all speakers, the means of A1-P1 decrease in the order CVC, NVC, and NVN. According to the statistical results, Subject 2 is the only individual that consistently corresponds with Hypothesis B.1 throughout all the time-points. With the exception of Subject 1, A1-P1 values of CVC are statistically significantly higher than those of NVC or NVN. Two-thirds of the subjects (Subjects 1-3, 6) show that A1-P1 is statistically smaller ($p<0.05$) in NVN than in NVC at all time-points, implying the vowel is more nasalized with two adjacent nasal consonants. However, there is no statistical difference between A1-P1 in NVN and NVC in Subject 5 at any time-point, though the mean of NVN appears lower than NVC in Figure 1. Subject 4 shows no statistical difference in A1-P1 between NVN and NVC for time-points 10, 20, and 30ms, but difference ($p<0.05$) for time-points 40 and 50ms, implying that his velum is relatively raised after 40ms from the nasal consonant release in NVC cases, but not in NVN cases.

Hypothesis B.2: The results reported here are taken from the final 50ms of the vowel. For data combined across all subjects, the means of A1-P1 in CVN and NVN are statistically smaller than that of CVC, but there is no difference between CVN and NVN. This finding implies that on average the vowels in CVN and NVN are nasalized to the same degree. For Subject 2, A1-P1 is statistically smaller ($p<0.05$) in NVN than in CVN for all measured time-points, as hypothesized. In contrast, Subject 1 shows the opposite: A1-P1 is statistically greater in NVN than CVN in all measured regions (i.e., CVN more nasalized than NVN). Subjects 3 and 4 show no statistical difference between NVN and CVN in most time-points, and Subjects 5 and 6 do not show any statistical difference in all time-points. These findings suggest that velum height is not necessarily the lowest in NVN cases, and acoustic evidence of nasalization in NVN and CVN is statistically the same for half of the subjects.

In summary, vowel nasalization occurs in contexts of NVN, CVN, and NVC; A1-P1 in vowels with at least one adjacent nasal consonant is

statistically smaller than in vowels with no adjacent nasal consonants. Two adjacent nasal consonants, however, do not guarantee more vowel nasalization than vowels with only one adjacent nasal consonant, especially in CVN contexts.

Figure 1. The mean and the lower and upper quartiles of A1-P1 over the initial and final 50ms of the vowel. Time 0 denotes the closure/release of the nasal consonant; negative time indices indicate time before the nasal consonant closure, and positive time indices indicate time after the consonant release.
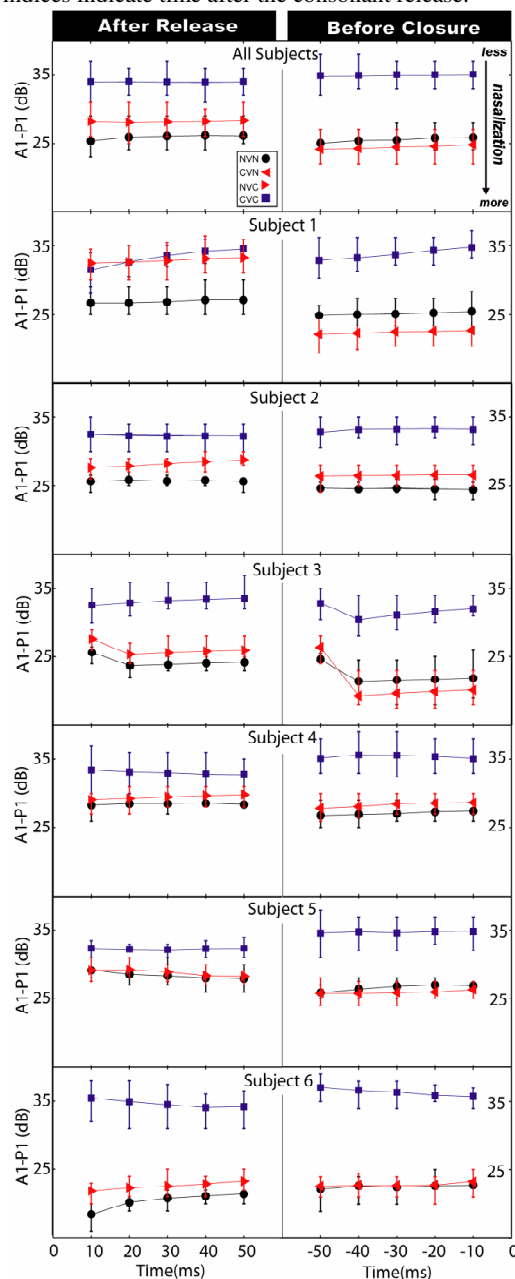
4. CONCLUSIONS

Despite individual differences among vowels in various nasal contexts (NVC, CVN, and NVN) among the subjects, this study shows that A1-P1 is statistically smaller when a vowel has at least one adjacent nasal consonant, be it pre- or post-vocalic.

However, more adjacent nasal consonants do not guarantee more vowel nasalization; the final region of a vowel in CVN context might be more nasalized than in NVN context. Though normalization across subjects was not considered and other acoustic correlates of nasality could have been used, this study still demonstrates the acoustic variability of vowel nasalization due to phonetic context in American English. Our study also gives insight into possible differences in speech motor planning of carryover and anticipatory nasalization. For future studies, acoustic variability can be analyzed in vowels other than /i/, on female speakers, in languages where there are phonemic nasal vowels, and on more speakers to investigate individual speaker differences. .

5. ACKNOWLEDGEMENTS

6. REFERENCES

[1] Maddieson, I. 1984. *Patterns of Sounds*: Cambridge University Press.

[2] Beddor, P. 1993. The perception of nasal vowels, In: Huffman, M.K., Krakow, R.A. (eds), *Phonetics and Phonology: Nasals, Nasalization, and the Velum.*: Academic Press, pp. 171-196.

[3] Hasegawa-Johnson, M.A., Baker, J., Borys, S., Chen, K., Coogan, E., Greenberg, S., Juneja, A., Kirchoff, K., Livescu, K., Mohan, S., Muller, J., Sonmez, K., Wang, T. 2005. *Landmark-based speech recognition: Report of the 2004 Johns Hopkins summer workshop.*, Proc. ICASSP.

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

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

[6] Krakow, R.A. 1993. Nonsegmental influences on velum movement patterns: Syllables, Sentences, Stress and Speaking Rate, In: Huffman, M.K., Krakow, R.A. (eds), *Phonetics and Phonology: Nasals, Nasalization, and the Velum.*: Academic Press, pp. 87-116.

[7] Pruthi, T. 2007. *Analysis, Vocal-Tract Modeling and Automated Detection of Vowel Nasalization*, in *Electrical Engineering*, vol. PhD: University of Maryland.

[8] Moats, L. 2000. *Speech to Print: Language Essential for Teachers*. Baltimore, Maryland: Paul H. Brooks Publishing Co.

[9] Chen, M.Y. (unpublished work). *Nasal Detection Module for a Knowledge-based Speech Recognition System*, in *J. Phonetics*.

[10] Conover, W.J. 1980. *Practical Nonparametric Statistics*: Wiley.

[11] Mukerjee, S., Golland, P., Panchenko, D. 2003. *Permutation Tests for Classification*, in *AI Memo*: Artificial Intelligence Laboratory, Massachusetts Institute of Technology.