Acoustic and Perceptual Similarity in Coarticulatorily Nasalized Vowels

Rebecca Scarborough¹, Georgia Zellou¹²

¹Department of Linguistics, University of Colorado, Boulder, CO, USA
²Department of Linguistics, University of Pennsylvania, Philadelphia, PA, USA
rebecca.scarborough@colorado.edu, georgia.weissman@colorado.edu

Abstract
This study investigates the acoustic and perceptual consequences of nasal coarticulation in American English. Nasalized (coarticulated) vowels were found to be closer in the F1-F2 acoustic vowel space than corresponding oral (non-coarticulated) vowels, indicating that contrast is reduced in the nasal vowel space, relative to the oral vowel space. With respect to perception, listeners are, perhaps unsurprisingly, more accurate in identifying oral vowels than nasalized vowels. Interestingly, however, while listeners take longer to identify nasalized vowels than oral vowels when they hear those vowels in isolation, this difference in processing time disappears when the nasalized and oral vowels are heard in lexical contexts. We take these findings to indicate that listeners compensate for nasal coarticulation, albeit sometimes incompletely, attributing the acoustic effects to their consonantal source, and that this compensation takes place instantaneously.

Index Terms: nasality, coarticulation, perception, similarity

1. Introduction

Although vowel nasality is not phonologically contrastive in English—there is no meaningful difference that can be expressed by using a nasalized vowel versus a fully oral vowel—there is pervasive vowel nasality nonetheless. This nasality is due to nasal coarticulation and occurs on vowels adjacent to nasal consonants. If the acoustic effects of this nasality are not critical for perceiving contrasts, we are left to ask: what are the perceptual consequences of this nasality? There are essentially two conflicting perspectives about the perception of coarticulation in the literature. In the first, coarticulation is viewed as a result of target-undershoot that occurs as long as such deviation from canonical is not prohibited [1]. From this stance, coarticulation should be minimized when the perceptual demands of the listener are weighted more heavily and can be increased when the articulatory demands of the speaker are weighted more heavily. On the other hand, coarticulation is not just random deviation from canonical—it is predictable, systematic variation that occurs due to context. Thus the acoustic effects of coarticulation, as long as they are interpretable—for instance in the appropriate coarticulatory context, can be perceptually informative. In other words, coarticulation structures speech to assist listeners [2].

This study examines the effect of nasal coarticulation on vowel quality and the resulting consequences for perception. There are several fundamental questions about the role of coarticulation in speech communication that this study intends to address. First, what is the effect of nasal coarticulation on the acoustic and perceptual distance between vowels? Cross-linguistically, languages with contrastively nasal vowels tend to have fewer quality distinctions among the nasal vowels than among oral vowels [3], suggesting that vowel nasalization may reduce the acoustic and/or perceptual distance between vowels. We predict similar acoustic and perceptual effects for non-contrastive nasality as well. Namely, we predict that vowels in nasal consonant contexts will occupy a contracted vowel space relative to vowels in oral contexts in American English and that this reorganized vowel space will lead nasalized vowels to be perceived as more similar to one another than oral vowels.

But is nasal coarticulation then detrimental to perception? In real listening situations, listeners hear and must perceive whole words. Although coarticulation results in deviation from canonical forms for individual segments, it is also contextually predictable—and therefore predictive of its context. In other words, because nasal coarticulation always occurs in the context of a nasal consonant, the presence of coarticulatory vowel nasality indicates the presence of an adjacent nasal consonant. If coarticulation can indeed provide perceptually useful information in this way, it may not be perceptually detrimental, despite increased vowel similarity.

In fact, the perceptual effects of coarticulation may have everything to do with context. It has been shown that nasalized segments in non-nasal contexts (or in isolation) are perceived by English-speaking listeners as “nasal”; in an appropriate nasal consonant context, however, the segments are judged as less nasal or even oral [5], [6], as the nasality is attributed to the source (the nasal consonant). If this compensation for coarticulation applies not only to the nasality itself but also to the associated F1 and F2 effects, nasal coarticulation may only be perceptually detrimental when there is no context.

Thus, this study explores additionally whether vowels (and nasal coarticulated vowels, in particular) are perceived differently in isolation versus lexical context. We predict that nasal vowels identification will be more difficult (slower and/or less accurate) in isolation than in a natural lexical context.

2. Acoustic Description

We begin with an acoustic description of the speech that will be used in our main experiment.

2.1. Methods

The target vowels for this study (/ɑ/, /ɛ/, /æ/, /ʌ/) are four contrastive non-high vowels in American English—chosen because they tend to be most nasal. There is an inherent link between the velum and the jaw, such that as the jaw lowers, it tends to pull the velum lower too. Thus, lower vowels, with a lower jaw position, tend to be more nasal.

The subject for this study was one male native English speaker. The recordings were made in a sound booth at a 44kHz sampling frequency. Eight real words were selected for each of
the four target low vowels, (/a/, /ɛ/, /æ/, /ʌ/), consisting of four oral-nasal minimal pairs where the final consonant differed only in nasality (e.g. “bud”, “bun”). (The nasal and oral stops for each minimal pair match in place of articulation.) Each word was repeated 5 times, for a total of 160 tokens. Vowel quality measurements of the first and second formants were taken at the midpoint of each token. Since we are concerned with acoustic contrast and perceptual distinctiveness the data were bark-scaled in order to analyze the data with psychoacoustically relevant values [7].

2.2. Results

Means for each vowel, separated by nasal or oral context, are shown in Fig. 1.

Acoustic distances between adjacent vowels of the same nasality and the same backness (front or back) were calculated, along with individual formant differences for these pairs, as displayed in Fig. 2.

These data indicate that oral vowels are further apart acoustically than nasalized vowels, and this greater distance is reflected on both the F1 and F2 dimensions for the back vowel pair /a/ - /ʌ/ and on the F1 dimension for the front vowel pair /ɛ/ - /æ/. Furthermore, we can see that the influence of nasality context on vowel quality was greatest for the /a/ - /ʌ/ pair, especially with respect to F1, where vowel space is reduced by .74 bark between /a/ and /ʌ/ relative to /a/ and /ʌ/. Vowel space is reduced by .33 bark from /ɛ/ - /æ/ to /ɛ/ - /æ/.

The observation that nasalized vowels are acoustically closer than corresponding oral vowels was confirmed by a series of two-way analyses of variance performed on the bark-scaled data with Euclidean distance, F1, and F2 as dependent variables and vowel pair (back pair /a/-/ʌ/ or front pair /ɛ/-/æ/) and nasality (oral, nasalized) as independent variables. With respect to Euclidean distance, the analysis revealed a significant interaction between vowel pair and nasality [F(2, 1596)=79.7, p<.001]. For both vowel pairs, the oral vowels were further apart than the nasalized vowels, but the distance for the back pair was greater than the distance for the front pair. The F1 ANOVA revealed a similar significant interaction between vowel pair and nasality [F(2, 160)= 7.6, p<.01]. The F1 difference was greater for oral than for nasalized vowels for both the front and the back pair, but again, the difference was greater for the /a/-/ʌ/ pair. Finally, the F2 analysis showed an interaction between vowel pair and nasality as well [F(2, 160)= 10.1, p<.01]. In the case of F2, however, the effect was different for the front pair than it was for the back pair. For F2 for the front vowels, /ɛ/ - /æ/, the difference for the nasalized vowels was actually greater than the oral vowel difference.

The comparison of coarticulated and non-coarticulated vowels, then, revealed that nasality does affect vowel quality. The coarticulated vowels were consistently raised (with lower F1) relative to corresponding non-coarticulated vowels, as in [8], explainable by the spectral skewing that occurs with the introduction of the “nasal formant” in the lower frequencies of the spectrum. And these F1 differences netted smaller differences between nasalized vowels than between oral vowels. F2 was also affected by nasal coarticulation, though this effect was different for the front vowels than for the back vowels: front vowels were fronter (with a higher F2) for the nasalized variant than for the oral variant; back vowels were backer (with a lower F2) for the nasalized variant. These differences netted a smaller nasalized vowel difference for the back pair, but a smaller oral vowel difference for the front pair. Importantly, though, together these spectral differences cause oral vowels to be more acoustically dispersed than nasalized vowels. Or put differently, nasalized vowels occupy a more reduced vowel space than oral vowels. We might expect, then, that nasalized vowels would be less perceptually distinct than their oral counterparts. We will address this prediction in the following perception experiment.

3. Perception Study

The aim of this experiment is to investigate how the acoustic consequences of nasal coarticulation, where nasality reduces distance between non-high vowels, influences perception. As in the acoustic description, the target vowels of this study are the four English vowels (/a/, /ɛ/, /æ/, /ʌ/), which are potentially most susceptible to the effects of nasal coarticulation. As stated above, since nasal coarticulation reduces the acoustic distance between vowels, it may be predicted that there will also be a reduction of perceptual contrast. Such an effect would be evidenced in greater confusability between nasalized vowels, as well as slower processing, in tasks such as identification, relative to responses for corresponding oral vowels.

However, in real speech, vowels are not perceived in isolation; rather, of course, they are embedded in lexical contexts. Because it has been shown that coarticulatory vowel nasality may be attributed to its consonantal source when the source is present, rendering the percept of the vowel essentially “oral” [9], [5], we will investigate potential differences in perception between vowels in lexical contexts and those in isolation. If the compensation for coarticulation parses not only
the nasality itself to the source nasal but also its consequences for vowel quality in the F1 and F2 dimensions, it is predicted that differences in confusability and processing for oral and nasalized vowels in isolation will disappear when the vowels are perceived as parts of words.

3.1. Methods

Oral and nasalized vowel stimuli were presented in two contexts: a lexical context and an isolation context (created by excising the vowel from the lexical context). The stimuli were created from data collected in the production experiment. Recall that these words consisted of oral-nasal minimal pairs (e.g., “bud”, “bun”). Three repetitions of each target word were selected and mixed with Gaussian noise in order to increase the difficulty of the task and decrease the likelihood of ceiling effects on identification.

The stimuli were presented in a forced-choice identification task controlled by PsyScope X. The experiment consisted of four blocks: [/ɛ/, /æ/] in lexical context, [/ɛ/ /æ/] in isolation, and [/ʌ/ /ɑ/] in isolation. A short break and instructions were given at the beginning of each block. Participants used an ioLabs button box to select which vowel they heard: <α> “as in ‘bad’” representing /ɛ/, <ɛ> “as in ‘bed’” representing /æ/, <ο> “as in ‘bud’” representing /ʌ/, and <ο> “as in ‘bud’” representing /ɑ/ (the two options available for response in a given block were displayed continuously during the block). Participants were instructed to respond as quickly and as accurately as possible. A total of 192 stimuli were presented, randomized within each block. Responses and reaction times (calculated from offset of vowel for isolation context stimuli and from offset of word in lexical context stimuli) were collected for analysis. Ten undergraduates (8 female, 2 male) were recruited and paid 5$ for participating in the study.

3.2. Results

3.2.1. Accuracy Analysis

Overall, listeners were more accurate when identifying oral vowels (95%) than nasalized vowels (91%) and more accurate when identifying vowels in a word context (96%) than vowels in isolation (91%). Listeners were least accurate with nasalized vowels in isolation (89%). Accuracy results, broken down by vowel pair are illustrated in Fig. 3.

These results indicate that lexical context does indeed help improve perception of nasalized vowels, though this lexical benefit is seen for oral vowels as well.

3.2.2. Reaction Time Analysis

Response times (RTs) from correct responses were log transformed and fitted to a linear mixed-effects model using the lmer function in R (p-values were obtained by the pvals.fnc function in the languageR package) [10]. Nasality (oral, nasalized), context (word, isolation), and vowel pair (/ɛ-/ /æ/ or /ʌ-/ /ɑ/) were set as fixed effects, while listener and trial number were set as random effects. The summary statistics from the RT model are given in Fig. 4.

![Figure 3: Identification accuracy, by context and vowel pair.](image)

![Figure 5: Mean RTs, by vowel nasality and context.](image)
4. Discussion

Production data revealed that nasalized vowels (i.e., vowels in nasal consonant contexts) in American English are acoustically closer to one another in the vowel space than corresponding oral vowels. (This effect was carried by both F1 and F2 for the difference between /æ/ and /ɑ/ pair and by F1 for the difference between /ɛ/ and /ɜ/.) This finding suggests perceptual consequences: in particular, nasalized vowels should be harder to perceive. This would also seem to support various claims (e.g., [3], [4]) that nasalized vowels are less perceptually distinct. However, we predicted that lexical context might mediate this perceptual effect, since nasalized vowels in nasal contexts have been shown to be perceived as essentially oral, with the nasality attributed to the nasal context [5], [6], [9].

Results from a forced-choice identification task show that listeners were indeed better at identifying oral vowels than nasalized vowels, both in isolation and in word contexts, consistent with the differences in acoustic distance between oral vowels vs. between nasalized vowels. This picture is complicated, however, by the pattern of response times in the identification task. Critically, while listeners were faster at identifying oral vowels than nasalized vowels when the vowels were heard in isolation, this difference was neutralized when the oral and nasalized vowels were heard in lexical context. In other words, despite the contracted nasalized vowel space, identification was just as fast for nasalized vowels as for oral vowels.

How is it that nasalized vowels require no more time for identification than oral vowels in word contexts (suggesting equal processing difficulty) but yield more errors (suggesting greater processing difficulty)? The first piece of the result, that nasalized and oral vowels have the same identification time, is consistent with coarticulatory compensation: when a nasalized vowel occurs in the context of a nasal consonant, listeners attribute the effects of the nasality to the consonant, and the perception of the vowel is not hindered. In fact, as time is not affected, we see that this compensation happens instantaneously—the nasality seems to be perceived in real time as part of the nasal consonant without additional processing. On the other hand, nasality on a vowel in isolation results in variation that is not attributable to any element of the signal, so its influence affects perception because it has to be factored out or otherwise accommodated as part of the identification of the vowel. This additional processing takes additional time.

The second piece of the result, that accuracy is better for oral vowels than for nasalized vowels, suggests however that compensation is not perfect. If the effects of the coarticulation on the vowel are not fully attributed to the consonant, some vestiges of the effect may remain and be misinterpreted as inherent to the quality of the vowel. Recall that the accuracy results mirror the production results: nasalized vowel pairs were acoustically closer than oral vowel pairs, and they are more likely to be confused in perception. In other words, in cases where compensation is incomplete, the reduced acoustic distance between nasalized vowels is not compensated for, leading to greater likelihood of misidentification.

5. Conclusions

In sum, the results from this study give a dynamic picture of the perception of coarticulatory nasalization in English. Nasalized vowels are acoustically more similar, but this increased similarity is largely, but not fully, compensated for when these effects can be attributed to their coarticulatory source. Reaction time data indicate that listeners compensate for the acoustic effects of coarticulation effortlessly—they identify a nasal vowel in context as quickly as they identify an oral vowel in context. But accuracy data indicate that sometimes this compensation is incomplete, leaving nasal vowels more likely to be misidentified.

6. Acknowledgements

We gratefully acknowledge the contributions of the members of the 2010 Linguistics 7030 class at the University of Colorado, who participated in carrying out the initial piloting of this study.

7. References