An Acoustic Analysis of Labialization of Coronal Nasal Consonants in American English

by

Elisabeth A. Hon

B.S.E., Electrical Engineering

Princeton University, 2003

Submitted to the Department of Electrical Engineering and Computer Science
in Partial Fulfillment of the Requirements for the Degree of Master of Science in
Electrical Engineering and Computer Science

at the

Massachusetts Institute of Technology

September 2005

Signature of Author…………………………………………………………………………
Department of Electrical Engineering and Computer Science
August 31, 2005

Certified by………………………………………………………………………………
Kenneth N. Stevens
Clarence J. LeBel Professor of Electrical Engineering
Thesis Supervisor

Accepted by……………………………………………………………………………..
Prof. A.C. Smith
Chair, Committee on Graduate Students

An Acoustic Analysis of Labialization of Coronal Nasal Consonants in American English

by

Elisabeth A. Hon

Submitted to the Department of Electrical Engineering and Computer Science
on August 31, 2005 in Partial Fulfillment of the Requirements for the
Degree of Master of Science in Electrical Engineering and Computer Science

ABSTRACT

A challenge for speech recognition models is to account for the variation between natural connected speech forms and the canonical forms of the lexicon. This study focuses on one particular sound change common in conversational speech, in which word-final coronal nasal consonants undergo place assimilation toward following word-initial labial consonants. Formant frequency measurements were taken from words ending with coronal nasal consonants in potentially assimilating sentence contexts, and identical words ending in labial nasal consonants, across vowel contexts. The frequency of the second formant at vowel offset and during nasal closure was found to be sufficient to discriminate between underlying forms. There was evidence that even strongly-assimilated coronal segments differ on the basis of these cues from their pure labial counterparts. It is hypothesized that listeners can use these acoustic cues to uncover the intended place of articulation of assimilated segments, without recourse to phonological inference or sentence context.

Thesis Supervisor: Kenneth N. Stevens
Title: Clarence J. LeBel Professor of Electrical Engineering

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank my thesis supervisor, Prof. Ken Stevens, for his guidance and patience in my preparation of this study and this thesis. Many thanks also to Arlene Wint, an invaluable resource in the office, and to all of our office-mates. The helpful spirit and collegiality of the Speech Communication Group at MIT are a direct effect of Prof. Stevens’ kindness and openness, and are a great gift.

I would like to thank Dr. David Gow, Jr. for sharing his work and ideas with me, and my subjects, whose patient contributions have made this study possible. I would also like to thank Yoko Saikachi and Tamas Bohm for keeping me company during long hours of measurements.

Finally, I would like to thank Kevin Hunt for his companionship and patience during all my endeavors, and my family for their constant support.

1. Introduction

1.1 Place Assimilation

An ever-present obstacle to the success of speech recognition models is the vast variation between canonical word forms and natural connected speech. One widespread source of variation has been termed “assimilation”, a process by which neighboring segments in connected speech become more similar to each other. In English, the assimilation of a word-final coronal segment to the place of articulation of a following labial or velar segment is common. For example, when the phrase “green beans” is spoken in casual speech, the /n/ in “green” may be produced with a labial closure similar to that of the /b/ in “beans”, making the phrase to some extent resemble “greem beans”. Such alteration in the surface form of spoken words can pose great difficulties to recognition systems which rely on comparison with canonical forms in a lexicon.

Various theories as to the causes of assimilation have led to differing conclusions about the ability of the perceptual system to successfully recognize the intended forms. Roughly, these theories may be described in two groups: phonology-based theories and articulation-based theories.

Many phonology-based theories of assimilation support the concept of “underspecification” in lexical representation. Under the theory of underspecification, only distinctive feature specifications which are not predictable from context are included in the abstract canonical word forms in the lexicon. Unmarked or default feature values are not specified in the underlying forms, but they are filled in during production if the feature nodes are still empty. Before they are filled in, they are free to assimilate the feature values of neighboring segments. When assimilated forms are presented to the perceptual system, they are successfully recognized as the underlying forms, since they do not mismatch the unspecified features.

Lahiri and Marslen-Wilson (1991) claim to have demonstrated the reality of underspecification by examining vowel nasality in Bengali and English. Since vowel nasality is predictable and non-distinctive in English, they reason that the feature [nasal] is never specified for English vowels. In Bengali, vowel nasality is distinctive, and nasal vowels must be specified as [nasal]. Their experiments confirm that, when hearing nasalized vowels, English and Bengali speakers treat them differently, conforming to the differences in specification in their lexicons. When hearing oral vowels, however, English and Bengali speakers treat them the same. Lahiri and Marslen-Wilson interpret this result as evidence that [oral] is the universal default feature value for vowels, and is therefore unspecified in both languages. Oral vowels are then free to assimilate the nasality of a following nasal segment without mismatching the underlying representation.

Marslen-Wilson et al (1995) apply the theory of underspecification to their treatment of place assimilation in English. Since only coronal consonants regularly assimilate to the place of articulation of a following consonant, [coronal] is widely assumed by phonologists to be the universal default unspecified place feature (Avery & Rice, 1989; Paradis & Prunet, 1989). Since coronals are unspecified for place of articulation, they may freely assimilate to following labial or velar consonants. Labials and velars are specified for place of articulation and may not assimilate the place of another consonant. When the perceptual system is presented with a surface labial or velar consonant, the input is of course compatible with the labial or velar underlying form, but it is also compatible with the coronal underlying form, since it does not mismatch an unspecified feature. Accordingly, Marslen-Wilson et al (1995) found that listeners treat labial and velar surface forms, which could be derived from labial, velar, or assimilated coronal underlying forms, as more ambiguous than pure coronal surface forms, which could not be derived from assimilated labial or velar forms.

The application of underspecification theories to place assimilation requires an assumption about the discrete nature of assimilation changes. An underspecified segment which assimilates the place of articulation of the following segment undergoes a complete phonological feature substitution. A coronal segment which is followed by a labial segment may either remain the default coronal or become a labial segment, with no options in between. This assumption of discrete feature changes in place assimilation has important implications for acoustics and perception, as will be discussed in a later section.

As opposed to phonology-based theories of assimilation, articulation-based theories allow for assimilation to be a gradient process, with a continuum of realizations between unchanged surface forms and complete feature substitutions. From the point of view of articulation, assimilation can be seen as the effect of natural physical interactions between the various articulators in the vocal tract. In Browman and Goldstein's (1989) theory of gestural phonology, speakers' goals of speed and fluency often cause reductions in the size of individual articulatory gestures and increases in their overlap. This can produce assimilation when, for instance, the labial closure for the /b/ in "green beans" overlaps the timing of the /n/ in "green", causing it to be produced somewhat like an /m/. The overlap of successive gestures is not necessarily a deliberate change from the underlying lexical form; it is often merely "direct effects of principles of motor control economy" (Barry, 1985).

Browman and Goldstein (1990) found evidence for their articulatory phonology at work in place assimilation by examining X-ray microbeam displays of casual speech. In place-assimilated coronal consonants, they found that the coronal gesture of the tongue often had not been eliminated, but had merely been hidden by the overlapping gestures of adjacent consonants. Byrd (1992) used a model employing gestural overlap to synthesize speech, and found that the synthetic overlap of articulatory gestures created the perception of place assimilation for listeners. Evidence for place assimilation caused by gestural reduction, as opposed to overlap, was found by Jun (1996).

1.2 Susceptibility of Coronal Nasals

In English, place assimilation in stop and nasal consonants occurs only between word-final coronals and following word-initial labials or velars. Labial and velar segments do not assimilate the place of articulation of a following segment. Nor can place assimilation occur in the reverse direction; word-initial segments do not assimilate the place features of preceding segments. Why such asymmetry in the application of place assimilation in speech?

Underspecification proponents hold that coronal segments are uniquely susceptible to assimilation because they alone are unspecified for place of articulation. Since labial, velar, and other surface forms do not mismatch the underlying unspecified feature, coronal segments are free to assimilate the place of articulation of any neighboring segment without disrupting recognition. Labial and velar segments, on the other hand, are specified for place of articulation, and would no longer match the underlying forms if they were to undergo assimilation. Avery and Rice (1989) present evidence that this asymmetry is true across languages, indicating that [coronal] is the universal unmarked or default place of articulation. Paradis and Prunet (1989) propose that coronal segments lack a Place node altogether, and they present evidence that the Place nodes of adjacent vowels treat coronals as completely transparent.

The special status of coronals with regard to place assimilation may also have an articulatory basis, given that different physical structures are employed in the formation of constrictions for different places of articulation. Coronal constrictions are formed using the tongue tip, while labial constrictions use the lips, and velar constrictions use the tongue dorsum. Barry (1992) suggests that coronals are most often assimilated because the coronal gesture is the most easily interrupted. The tongue tip can be considered essentially massless, and therefore it can be quickly and easily deflected from its intended target. Browman and Goldstein (1990) agree that the tongue tip being faster than the lips or tongue dorsum may be a reason that coronals undergo more assimilation than other segments. They also point out that place assimilation occurs in segment sequences that form unacceptable syllable codas or onsets, including coronal consonants followed directly by labials or velars. They suggest that the gestures involved in such sequences are not timed relative to each other so as to prevent them from overlapping.

Perceptual explanations also exist for the asymmetry present in place assimilation processes. With regard to the regressive direction of place assimilation, it has been suggested that the weaker place cues of the word-final coda are sacrificed through assimilation to the perceptually-salient following word onset. Ohala (1990) points out that CV onsets contain formant transitions plus the release burst as cues to the place of articulation of the consonant, whereas VC codas contain only the formant transitions, making the onsets more perceptually rich (see also Hura et al, 1992). Gow and Gordon

(1995) suggest that the saliency of word onsets is preserved at the expense of word endings because of the importance of onsets in the process of lexical access. According to the "good start" model, the word onset begins the recognition system's search of the lexicon for matching words, and therefore it is most important in getting the system off to a good start.

Perceptual experiments have also provided explanations for weakness of coronals, and especially nasal coronals, against the effects of place assimilation. Byrd (1992) showed that coronal consonants have smaller formant transitions than labial consonants; if a coronal and a labial gesture are produced simultaneously in the vocal tract, the labial will dominate the acoustic result in terms of formant transitions. In addition many studies have found that nasals are more frequently assimilated than are stop consonants (Nolan & Kerswill, 1990; Ohala, 1990; Hura et al, 1992; Hardcastle, 1994). Nolan and Kerswill (1990) suggest that this is due to the nasalization of the preceding vowel, which obscures the formant transitions into the word-final nasal. Since the cues to place of articulation are less perceptually salient, they are less important to word recognition, and therefore they are allowed to blend with those of the following segment. This assimilation may also strengthen the percept of the following segment, thus facilitating the recognition of the following word, with a much smaller detriment to the recognition of the current word.

1.3 Perceptual Studies

Various perceptual studies of listeners' ability to successfully recognize the underlying forms of words affected by place assimilation have been used as evidence supporting the disparate theories of the nature and causes of assimilation. If listeners appear to use acoustic cues to discover the articulatory timing used by the speaker to form the segment in question, then the evidence is in favor of articulation-based theories of gradient assimilation. If, on the other hand, listeners appear to use their knowledge of phonological rules to infer the underlying identity of a changed segment, then the evidence supports phonology-based theories of discrete assimilation.

Several studies have indeed offered support for underspecification and phonological change as explanation for place assimilation processes. Marslen-Wilson et al (1995) used gating techniques to show that listeners would reinterpret labial and velar segments as underlyingly coronal, once they heard a following segment which could license such place assimilation. They interpret the ambiguity with which listeners treat the labial and velar segments as evidence for underspecification, and the fact that listeners reinterpret these segments in the presence of a licensing context as evidence for a process of phonological inference. It appears that the assimilated segments themselves do not contain acoustic cues as to their underlying identity; the listeners' knowledge of phonological rules and the presence of licensing context segments are required for successful access of the underlying forms (see also Coenen et al, 2001).

Gaskell and Marslen-Wilson (1996, 1998) found additional evidence for the use of phonological inference in the recognition of assimilated coronal segments. These experiments found that the coronal forms would be accessed only in the presence of a phonologically licensing context. However all of these experiments used stimuli in which the assimilation of coronal segments would create nonwords in English. For example, place assimilation in the phrase "green beans" produces "greem", which is not a word in English. The presence of a nonword in the surface form may facilitate the access of the underlying word, regardless of phonological context. Gaskell and Marslen-Wilson (2001) conducted experiments in which the assimilated forms were also real words in English; for example, "bean" could be labialized to form "beam", both of which are real words in English. They found that listeners would now access the coronal form only if the sentence context created a lexical bias in its favor. If the sentence context fit either form equally well, then only the surface form would be accessed, regardless of phonological context. Clearly there was no acoustic evidence within the target segment to indicate that it might be an assimilated form; listeners were forced to use neighboring words as phonological and lexical cues to the presence of assimilation.

The lack of acoustic cues to the underlying coronal segments is not surprising, however, since the stimuli in all these experiments were made not using natural tokens of assimilation, but rather using deliberate mispronunciations of the target words, often recorded by the authors of the studies. Since they assumed, as many phonologists have, that place assimilation is a discrete and complete feature substitution, they created their stimuli by pronouncing the words as if the coronal segments were instead labials or velars. It is no wonder that the listeners were unable to find evidence within the target segment for underlying coronality, since the stimulus segment was never in fact underlyingly coronal. Gow (2003b) points out that very different perceptual results obtain when natural assimilation is used as the stimulus.

Gow (2000, 2002), using natural tokens of place assimilation produced by naïve speakers, found that listeners do access the underlying coronal forms, even when the assimilation produces real words and the sentence context is ambiguous between the possible forms. When the following context was removed, he found that listeners access both possible forms, indicating that acoustic cues to both places of articulation are present within the assimilated segment. This lends support to the articulation-based theories of assimilation, which allow for gestures to overlap in a graded fashion, such that place-assimilated segments may be in between coronal and noncoronal articulatory forms. Gow and Zoll (2002) suggest a feature parsing strategy, whereby the perceptual system segregates the acoustic cues present in the assimilated segment, associating the coronal cues with the current segment and the noncoronal cues with the following segment. The noncoronal cues may act to facilitate early recognition of the following segment, since they are present earlier in the incoming speech signal.

Gow (2003a) found evidence for facilitation of the recognition of the following context segment in assimilated forms, supporting the feature parsing theory. Gow and Im (2004) found that this progressive context effect is not language-specific; even listeners hearing assimilation processes in a foreign language exhibit such facilitation. This result contradicts the phonological inference account of Gaskell and Marslen-Wilson, since the nonnative listeners do not possess the requisite knowledge of the phonological rules required for inference. Gow (2001) also found that the underlying forms are accessed in the native language when the phonological context is implausible.

The use of natural tokens of assimilation in perceptual studies rather than deliberate mispronunciations of words gives compelling support to graded articulatory models over the discrete phonological change theories. Even Gaskell (2003) agrees that the choice of stimuli may be a cause of the conflicting results obtained by him and Gow. It is worth noting in addition, however, that the studies of Gaskell and Marslen-Wilson were conducted using British forms of English, while Gow and colleagues worked with American English. British and American dialects differ considerably with respect to various aspects of stop consonant production (release, aspiration, glottalization, etc.), so it is not impossible that they would differ with respect to the scope of place assimilation processes. Although a controlled comparison study would put the question to rest, some articulatory evidence does already run counter to the discrete phonological change theory for British speech. Various articulatory studies, mostly conducted by British researchers, have shown gradient articulatory realizations of place assimilation, as will be described in the next section. These provide further support for gestural accounts and the presence of acoustic cues to the intended underlying forms of assimilated segments.

1.4 Articulatory Studies

Quite a few articulatory studies have demonstrated that place assimilation is a gradient process, with articulatory realizations along a continuum between extreme values of coronality and noncoronality. Hardcastle and Roach (1979) used an electropalatograph to monitor constrictions formed between the tongue and palate, while a camera recorded the movement of the lips. In this way they could observe the timing overlap between successive closures formed by different articulators for stop consonants. Although they reported some degree of overlap in 240 of 272 cases of stop sequences, the coronal closure was only completely absent in 6 of 96 cases. This is direct evidence against discrete feature change, which would require complete absence of the coronal gesture.

Barry (1985) used electropalatography to study the coproduction of coronal and velar consonants. He observed a residual coronal gesture in about half of the cases of assimilation. Moreover, the prominence of the residual gesture varied from case to case, suggesting a graded continuum of assimilation realizations. Nolan and Kerswill (1990) divided this continuum into a three-point scale consisting of complete coronal closure, partial coronal closure, and complete lack of coronal closure. Although no coronal gesture is recorded on the electropalatograph in the third case, a coronal gesture may still exist which does not make contact with the palate. Nolan (1992) demonstrated that listeners can tell the difference between an assimilated coronal with no coronal closure and an underlyingly noncoronal segment with above-chance accuracy.

Hardcastle (1994) used electropalatography plus acoustic airflow and laryngograph measurements to further demonstrate the gradient nature of coronal-to-velar assimilation. He found that different speakers exhibit different degrees and frequency of assimilation. Ellis and Hardcastle (2002) echo this result using electropalatography, and find that speakers differ as to their general assimilation strategies. Some speakers never assimilate, some always assimilate completely, some assimilate in a gradient fashion, and some assimilate in a binary fashion. They suggest that treating place assimilation as a fully gradient process ignores these important between-speaker variations. However, since a good recognition system, human or artificial, should work across all speakers, it seems expedient to treat assimilation as fits its across-speaker behavior, which is manifested as a gradient continuum.

1.5 Acoustic Studies

In combination with perceptual and articulatory studies, our understanding of the recognition system’s response to place-assimilated segments should rely on studies of the acoustic information present in such segments. After all, the acoustic signal is the most significant input received by the listener. Acoustic studies must identify the cues for place of articulation present in the signal, and chart their progression along the continuum between coronal and noncoronal forms.

1.5.1 Acoustic Cues for Place of Articulation

The search for acoustic cues for place of articulation of stop consonants dates back to the 1950’s, when researchers used speech synthesized from hand-painted spectrograms to chart the perceptual effects of controlled changes in various acoustic factors. Schatz (1954) found that the frequency of the release burst was a perceptual cue for place of articulation in voiceless stops, and Delattre et al (1955) found that the transition of the second formant (F2) frequency between consonant and vowel was a cue for voiced stops. Delattre et al also found F2 "locus" frequencies for the best exemplars of each place of articulation. That is, for a given place of articulation and fundamental frequency, the F2 transition could be seen to point to a specific frequency during the consonant closure, independent of vowel context. The F2 loci for labial and coronal consonants were easily discriminated from each other. They found the F1 locus to be the same for all places of articulation, making F1 not a useful cue for this feature.

Harris et al (1957) found that F3 transitions were also cues for place of articulation in voiced stop consonants, in addition to and independent of F2 transitions. However, they also found that the steady-state F3 frequency can vary within a given vowel identity, making the transitions harder to characterize. Evidence for F3 loci was weaker than that for F2. Hoffman (1958) found that release burst frequency was a place cue for voiced stops, as well as unvoiced stops, and that all of the relevant cues act independent of each other. He reasoned that the various cues add as vectors, with cues for the same place creating a strong percept, and cues for disparate places creating ambiguous percepts.

Stevens and Blumstein (1978) found that bursts alone were not sufficient for the identification of place of articulation by listeners. They found that formant transitions alone were sufficient, but better results were obtained by the bursts and transitions combined. They theorized that the perceptual system's primary cue for place of articulation was the gross shape of the short-term spectrum (spectral tilt) at the onset of the stimulus, which would include both the release bursts and formant transitions. The formant transitions would act as secondary cues, which would be invoked if the primary cue were missing or distorted. However, Blumstein et al (1982) later found that listeners do not base their choice of place of articulation on the spectral tilt, but rather on the formant transitions. Walley and Carrell (1983) found this to be true for adults and children, indicating that spectral tilt is never a primary cue for place of articulation.

The combination of these studies, then, points to formant transitions between consonant and vowel, especially those of F2, as the most reliable cues for place of articulation. A high-frequency F2 locus during the consonant is a cue for coronal place, while a low-frequency F2 locus is a cue for labial place; the F2 locus for velar place is more variable depending on vowel context. The formant transitions for nasal consonants have been assumed to be similar to those of stop consonants (Stevens, 2000). In place assimilation studies the formant transitions are especially important cues, since the target segments are in word-final, unreleased positions, and do not show release bursts.

1.5.2 Recent Assimilation Studies

Acoustic studies of the effects of place assimilation in stop consonants and nasals in English have been few in number compared with perceptual and articulatory studies. Zsiga (1994) measured the difference in formant frequencies between the vowel midpoint and the vowel offset for F2 and F3 before a coronal stop consonant. She found that the presence of assimilation does cause changes in ΔF, and she found these changes to vary along a continuum. She did not measure how closely these changes might cause the formant transitions to approximate those of pure noncoronals. It is the acoustic differences between assimilated coronals and pure noncoronals which would prove useful in recognition systems for determining the underlying form of an assimilated segment.

According to Manuel and Stevens (1995), the difference in F2 transitions between coronal and labial consonants has to do with the position of the tongue body. During a labial closure, the tongue body is relatively uninvolved, and therefore remains in a neutral position, since the lips are completely separate articulators. During a coronal closure, however, the tongue tip must make contact with the alveolar ridge, and the tongue body must accordingly be fronted, since it is part of the same physical structure as the tip. (It has been suggested that this fronting process is not required for all languages; however it is generally applicable to English.) It is the fronting of the tongue body which makes F2 higher for coronal consonants than labial consonants. When a coronal consonant is assimilated to a following labial, the coronal closure and the corresponding fronting movement are interrupted by the early labial closure. However, a partial fronting movement may still have been accomplished in preparation for the underlying coronal. This would be manifested in an intermediate F2 transition, between the characteristic transition for a coronal and that for a labial. Indeed, Gow and Hussami (1999) find that the formant transitions for assimilated coronal stops are intermediate between pure coronal and pure noncoronal forms. Their study does not deal with nasal consonants, and also does not account for differences in vowel context.

Dohr (2004) did include various vowel contexts in her study of place assimilation in stop consonants; however most of the results in her particular study did not reach significance. She measured formant frequencies at regular spacings across the entire duration of the vowel preceding assimilated coronals, but only found a significant difference in F2 between assimilated coronals and pure labials at the end of the vowel. This indicates that the steady-state portions of the vowel are not important for distinguishing between underlying forms; however the formant transitions just before the consonant are. At the end of the vowel Dohr’s results are in the predicted direction; F2 is higher for assimilated coronals than it is for pure labial forms.

1.5.3 Current Study

The goal of the study described in this thesis is to expand upon recent acoustic studies of place assimilation in American English to categorize its effects on formant transitions. If significant differences can be found between the transitions present in assimilated coronals and those present in underlying noncoronals, then those differences can be exploited by recognition systems. The presence of adequate acoustic cues to the underlying identity of assimilated segments may also provide support for perceptual theories which are not grounded in underspecification and phonological inference. This study attempts to identify which particular acoustic cues are the most useful in identifying underlying forms, and to quantify these cues across vowel contexts.

The present study focuses exclusively on the assimilation of coronal nasal consonants to following labial segments. Nasals have been chosen because they are known to assimilate more frequently than stop consonants (Ohala, 1989; Nolan & Kerswill, 1990; Hura et al, 1992; Hardcastle, 1994), and also because it is hypothesized that nasals may possess acoustic cues which are not available in stops. Since phonation continues throughout the nasal consonant closure, it is theoretically possible for formants to be tracked throughout the production of the consonant; whereas stop consonants exhibit a silent interval during which the formant trajectories cannot be observed.

In order to narrow the scope of this study, only labializing assimilations are dealt with, as opposed to velarizing assimilations. It has been suggested that the acoustic differences between coronals and labials may be larger and clearer than those between coronals and velars (Delattre et al, 1955; Byrd, 1992; Manuel & Stevens, 1995). In future work this study will hopefully be expanded to include acoustic analyses of velarized segments, as well as to add articulatory and perceptual components to the acoustic categorizations.

2. Method

2.1 Recordings

Subjects for this study were drawn from the Massachusetts Institute of Technology student community. Speech samples were recorded from two male speakers (TH and KH) and two female speakers (MA and LS), all between the ages of 20 and 30 years. All subjects were naïve to the purpose of the experiment and were paid for their participation. All were native speakers of American English, with no self-reported hearing or speech disabilities.

Recordings were made in the Eastham Room, a sound-attenuating chamber in the laboratory of the Speech Communication Group (MIT room 36-512). The subject was seated in a comfortable chair, and a microphone stand was used to position the microphone about six inches from the subject's mouth. Directly in front of the subject, at a comfortable viewing distance, was a computer monitor on which the utterances to be spoken were displayed. Each subject completed Part 1 of the recording session, which took about two hours, on one day, and returned on a later day to complete Part 2, which took about one hour. Three short breaks were taken during each part of the recording session to prevent the subject from becoming tired or uncomfortable.

The MARSHA software tool was used, along with a pre-amplifier and anti-aliasing filter, to digitize recordings at a sampling rate of 12 kHz. A script was written through which MARSHA cycled through the utterances in pre-programmed order, recording each utterance to a separate file on the computer. The printed form of each utterance was displayed on the computer monitor, at which time the subject spoke the utterance into the microphone.

In Part 1 of the recording session, the subject was directed to speak sentences in which were embedded words ending in coronal or labial nasal consonants, followed by words beginning with voiced labial stop consonants or labial nasal consonants. For each of nine different vowels, four test sentences were constructed, for a total of 36 test sentences.

The first test sentence for each vowel contained a one-syllable stress-bearing word with that vowel for the nucleus and the coronal nasal consonant /n/ for the coda, followed by a stress-bearing word with the voiced labial stop consonant /b/ for the onset. (The unvoiced labial stop consonant could also have been used here; the decision to use the voiced version was made arbitrarily for consistency. There is no reason to believe that the unvoiced version would produce different results than the voiced version.) The second test sentence for each vowel contained the same word combination as the first sentence, except the coronal nasal consonant /n/ was replaced by the labial version /m/. The words were chosen so that the version ending in /n/ and the version ending in /m/ were both real words in English. For example, for the vowel /e/, the first test sentence contained the combination “cane back”, and the second sentence contained “came back”.

The third and fourth test sentences for each vowel were a similar pairing of coronal nasal and labial nasal target segments; however the following word here began with the labial nasal consonant /m/ instead of the stop consonant. For instance for the vowel /i/, the third test sentence contained the combination “teen made”, and the fourth sentence contained “team made”. Across all the test sentences, the word with the labial onset was not always exactly identical to its counterpart in the paired sentence. However, they were identical in the onset and nucleus of their first syllables: for example, “beanbag” was paired with “beam back”. The difference in the second words’ codas was allowed in order to facilitate the creation of natural-sounding sentences, and also to help prevent the subjects from detecting the patterns in the sentences they were asked to speak. The difference in the codas of the second words should not have a significant effect on the pronunciation of the nasal codas of the first words.

The purpose of embedding the target words in sentences was to facilitate place assimilation of the word-final nasal coronal consonant to the following labial consonant, since this phenomenon tends to occur in conversational speech. The test sentences were therefore constructed with the goal of creating natural sentences which could be part of an everyday conversation (see Appendix 1 for the list of sentences). Subjects were instructed to speak the sentences “in a casual and somewhat rapid fashion, as if you were having a normal conversation with a friend.” In order to minimize prosody effects, the target words were never the first or last words in the sentence, and they were never separated by clause boundaries.

In order to prevent subjects from guessing the purpose of the experiment, and to draw attention away from the repeated sound combinations of the target words, 36 filler sentences (also shown in Appendix 1) were interspersed among the target sentences within the recording session. The filler sentences were chosen and constructed in order to match the test sentences in approximate length and general prosodic shape, but they did not contain target words. The test sentences were recorded in their entirety, while the filler sentences were not recorded at all. The subjects were prompted to speak the complete set of 72 sentences four times, in a different random order within each of the four sets. Repetitions were necessary to obtain assimilated speech samples, because assimilation does not occur invariably, even in rapid conversational speech (see Ellis & Hardcastle, 2002). A short break was taken after each set of 72 sentences for the subject to stretch and rest.

In Part 2 of the recording session, the subject was directed to speak single words embedded in the carrier phrase, “Say ____ again.” The words included the target words from Part 1 which ended in nasal consonants, as well as versions of these words with the final nasal changed to a voiced stop consonant (see Appendix 2 for the list of 72 words). When the word-final nasal was changed to a stop consonant, the result was sometimes a nonword in English. This was acceptable, since the goal of Part 2 was to capture tokens of the words in isolation, rather than in natural conversation. The recorded tokens of the words ending in nasal consonants were used to compare the assimilated versions from Part 1 with canonical forms from Part 2. The versions of the words ending in stop consonants were used to permit comparison with the words ending in the corresponding nasals, in order to aid in identifying nasal resonances and separating them from measurements of formants.

In Part 2, as in Part 1, the subjects were asked to speak the 72 utterances four times in total, in a different random order in each of the four sets. A short break was taken after each set for the subjects to stretch and rest.

2.2 Measurements

The Xkl software tool in the MIT Speech Communication Group laboratory was used to make measurements of formant frequencies in all recorded speech tokens. The frequency measurements were taken from individual DFT spectra with Hamming window length corresponding to the length of the pitch period. The window length was reset for each utterance to match the pitch period as closely as possible, but was kept constant for all measurement points within a given utterance. Ranges of window lengths used were: 3.5ms – 6.5ms for speaker MA; 4.0ms – 6.5 ms for speaker LS; 4.5ms – 10.5ms for speaker TH; 6.0ms – 9.5ms for speaker KH. The DFT window was positioned with the left edge slightly before the closing motion of the pitch period, as shown in Figure 1.

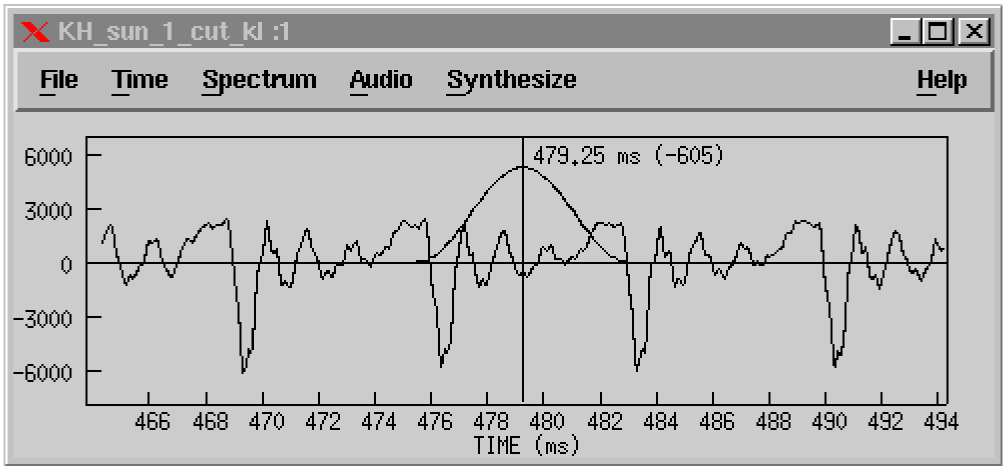

Fig. 1: Position of DFT Hamming window with respect to pitch period for a representative speech input.

Spectra taken over single pitch periods were used in order to obtain formant frequency measurements from very specific moments in time. In this study, a measurement point of interest was the transition from the vowel of the target word into the word-final nasal consonant, during which time the articulators move rapidly, and the formant frequencies may change significantly between consecutive pitch periods. While using longer DFT windows would conflate frequency information from several neighboring pitch periods, restricting the window length to cover a single pitch period allows measurements to be taken with greater time accuracy. Measurements can thus be taken from the point in time closest to the consonant closure, representing the full extent of the formant transitions.

Automated formant tracking programs were not used in this study, as they are prone to errors in identification of individual formant peaks. Also they are ill-equipped to distinguish formant peaks from nasal resonances, which is important in a study of nasal consonants. In this study the formant peaks of interest were identified by hand by comparing the DFT spectra with the visible trajectories of the formants on the spectrograms. The computer program was then used to calculate the exact frequency of the local maximum within the identified peak in the spectrum.

For each /VC#/ target speech sequence, the frequencies of the first four formants (F1, F2, F3, and F4) were measured at the vowel midpoint and vowel offset. The vowel midpoint was defined as the full pitch period closest to the point in time exactly halfway between the onset of vowel phonation and the acoustic discontinuity signaling consonant closure. The consonant closure discontinuity was identified by an abrupt decrease in amplitude visible in the spectrogram, and an abrupt change in shape visible in the waveform of the utterance. The vowel offset measurement point was then defined as the last full pitch period preceding this consonant closure. For the few cases in which this pitch period did not exhibit a clearly distinguishable F3 peak, the closest preceding pitch period with visible F3 was used instead.

The formant measurements taken from the words from Part 2 of the recordings which ended in stop consonants were used as aids in identifying formant peaks, as opposed to nasal resonances, in the words from both parts which ended in nasal consonants. Figure 2 shows an example of the comparison of the two word forms. The top portion of Figure 2 shows a spectrogram of the utterance, “Say ‘bead’ again,” in which the target word ends in a stop consonant. In the bottom portion, the utterance, “Say ‘bean’ again,” ends in a nasal consonant with the same place of articulation. The first nasal resonance, around 1000 Hz, can be seen in the bottom portion of the figure after time index 0.4, but it is absent in the top portion of the figure. Its presence during the vowel portion of “bean” and absence during the vowel portion of “bead” identifies it as a nasal resonance, and discounts it as a potential formant peak.

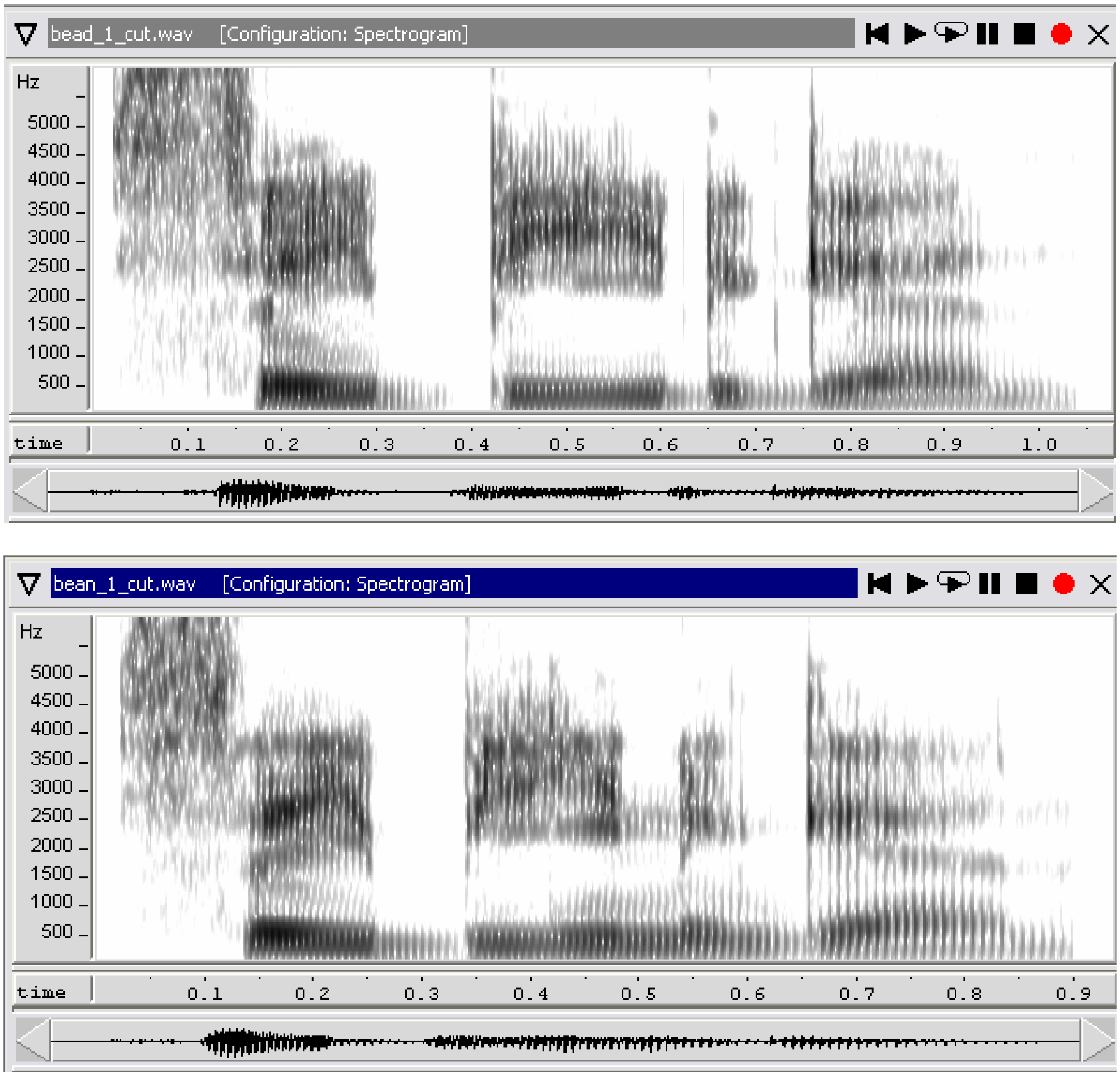

Fig. 2: Comparison of spectrograms of the utterances, “Say ‘bead’ again,” (top) and, “Say ‘bean’ again,” (bottom), produced by speaker KH. Note the nasal resonance which appears around 1000Hz during the word “bean”, but not during the word “bead” (time index 0.4 – 0.6).

For the words from Part 2 which ended in nasal consonants, in addition to the formant measurements at vowel midpoint and vowel offset, a third measurement was taken from the average spectrum across the nasal consonant closure. To find this measurement, the nasal closure was defined as the interval between the acoustic discontinuity signaling consonant closure and the acoustic discontinuity signaling onset of the following vowel. An average DFT spectrum was taken across this interval, using the same window size as the other measurement points, and from this spectrum the formant peaks were identified in the same fashion as for the other measurement points.

The same three formant measurements were taken for the /VN#b/ sequences from Part 1 of the recordings. For the /VN#m/ sequences from Part 1 of the recordings, four measurement points were used: the vowel midpoint, the vowel offset, the average spectrum across the first nasal consonant, and the average spectrum across the following /m/. In many cases, an acoustic discontinuity, usually manifested as abrupt shifts in formant frequencies and/or changes in formant amplitudes, was present to mark the transition point between the first nasal and the following /m/. If this discontinuity was not present, the transition point was taken to be the midpoint between the onset of the first nasal and the offset of the following /m/.

Before the formant measurements were taken, perceptual judgments were made by the experimenter as to the degree of assimilation present in the /Vn#b/ and /Vn#m/ sequences from Part 1 of the recordings. These judgments were made from repeated listening to the sequences in their full sentence context, but without consulting the spectrograms or any other visual aids. Each sequence was scored on a three-point scale: a score of 0 indicated that no assimilation could be heard between the /n/ and the following labial; a score of 1 indicated some perceptual ambiguity between coronal and labial surface form for the underlying /n/; a score of 2 indicated that the underlying /n/ sounded completely assimilated and /m/-like. Nolan (1992) has demonstrated that offline perceptual scoring of this type closely parallels articulatory measures of the prominence of coronal gestures in place assimilation contexts.

3. Analysis and Results

3.1 Cues for Place of Articulation

The analyses for this thesis focused on the target words in sentence context from Part 1 of the recordings. Although measurements were taken from the words from Part 2 of the recordings, these measurements were retained for later analysis in future studies.

For the current study, statistical analyses of the various formant measurements from Part 1 of the recordings were conducted in order to determine which formant cues were the most useful in discriminating between coronal and labial place of articulation. Due to time constraints, not all of the measurements taken were included in the analysis. F1 measurements were excluded, since past research has indicated that the first formant is not a good indicator of place of articulation (Delattre et al, 1955). Although it is possible that the detail of the F1 trajectory near the vowel offset contains information relevant to place of articulation, this detail was not available from the measurements used in this study. For the few measurement points used here (vowel midpoint, vowel offset, nasal closure), the data for F1 were not found to be a good indicator for place of articulation. F4 measurements were also excluded, since they were difficult to obtain due to the low amplitude of the speech signal at high frequencies, and also since they did not appear to exhibit any strong patterns with respect to place of articulation of the consonant.

The remaining measurements to be analyzed were as follows:

- F2 frequency at vowel midpoint.
- F2 frequency at vowel offset.
- F2 frequency from average spectrum across nasal closure.
- F2 frequency from average spectrum across following /m/ (for /VN#m/ sequences).
- F2 frequency at vowel offset minus F2 frequency at vowel midpoint.
- F2 frequency from average spectrum across nasal closure minus F2 frequency at vowel midpoint.

- F2 frequency from average spectrum across following /m/ minus F2 frequency at vowel midpoint (for /VN#m/ sequences).
- F3 frequency at vowel midpoint.
- F3 frequency at vowel offset.
- F3 frequency from average spectrum across nasal closure.
- F3 frequency from average spectrum across following /m/ (for /VN#m/ sequences).
- F3 frequency at vowel offset minus F3 frequency at vowel midpoint.
- F3 frequency from average spectrum across nasal closure minus F3 frequency at vowel midpoint.
- F3 frequency from average spectrum across following /m/ minus F3 frequency at vowel midpoint (for /VN#m/sequences).

Each of these measurements was tested for its usefulness in discriminating between underlying coronal and labial segments in otherwise identical speech sequences. Measurements were taken from target words ending in coronal nasal consonants in potentially-assimilating sentence contexts, and their counterpart words ending in labial nasal consonants in non-assimilating sentence contexts. Measurements with significant differences between the coronal and labial forms should be useful during recognition in discriminating between the two.

Figure 3 shows spectrograms for three /VN#b/ target sequences from sentences spoken by speaker KH, with three measurement points marked by vertical lines and numbers. Measurement point 1 is the midpoint of the vowel; measurement point 2 is just prior to the consonant closure; measurement point 3 is the average spectrum across the nasal consonant closure. The spectrogram at the top of the figure is a partially-assimilated (assimilation score 1) token of the utterance “Ron built”; the spectrogram in the middle of the figure is a strongly-assimilated (assimilation score 2) token of “Ron built”; the spectrogram at the bottom of the figure is a token of the utterance “ROM built”. It can be seen that the trajectory of F2 near the vowel offset is moving upward in the top figure, is moving downward in the bottom figure, and is intermediate in the middle figure. It seems that the F2 trajectories for assimilated coronal forms are in fact

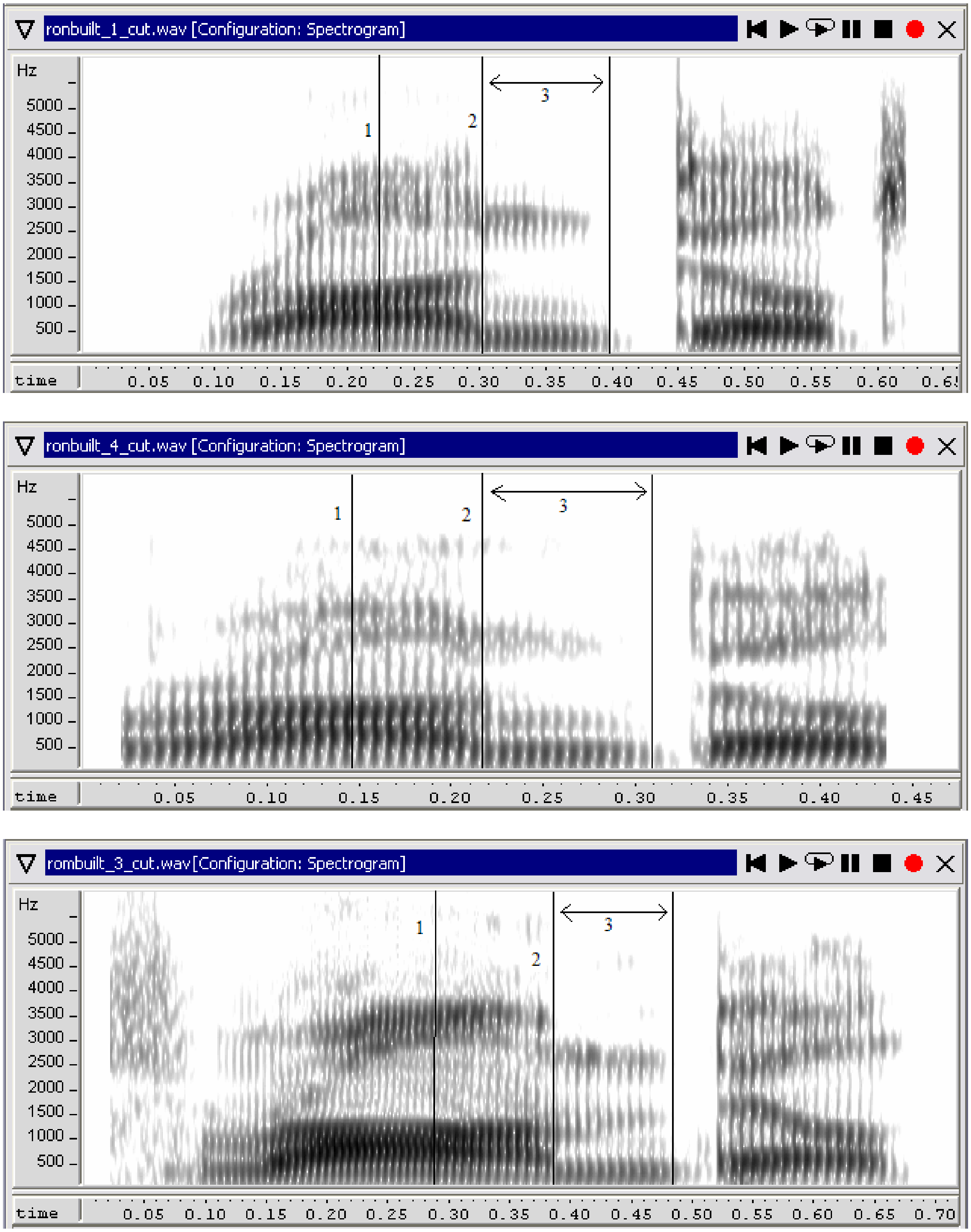

Fig. 3: Spectrograms of a partially-assimilated token of “Ron built” (top), a strongly-assimilated token of “Ron built” (middle), and a token of “ROM built”, produced by speaker KH. Measurement point 1 is the vowel midpoint; measurement point 2 is just prior to the consonant closure, and measurement point 3 is the average spectrum across the nasal consonant closure.

different from those of underlying labial forms; the analyses below were undertaken to confirm these observations.

Separate two-factor analyses of variance were performed for each vowel context in the /VN#b/ and /VN#m/ sequences. One factor was underlying place of articulation of the target nasal segment (coronal or labial), and the other factor was speaker (MA, LS, TH, KH). The averages for the coronal target segments were taken across all assimilation scores, so that they represent averages between strongly and weakly assimilated forms. Each sample was composed of four repetitions of the same utterance; if any data points were missing, they were filled in with the average of the other repetitions.

Results of the ANOVAS for each measurement in each vowel context are shown in Tables 1-4. The figures listed are averages of sixteen measurements (four repetitions for each of four speakers) for minimal pairs with underlyingly coronal and labial target segments. Each column represents a different measurement point, and each row represents a different combination of vowel context and labializing context segment (/b/ or /m/). In cases for which the figures have been left blank, the data were too sparse to perform ANOVAS. This occurred occasionally when the formant peaks were too obscure to measure in these particular recordings.

Measurement pairs shown in darker blue are significantly different at the P<.01 level, with the coronal measurement higher than the labial measurement. Pairs shown in lighter blue are significant at the P<.05 level, with the coronal measurement higher than the labial measurement. Pairs shown in darker red are significant at the P<.01 level, with the coronal measurement lower than the labial measurement. Pairs shown in lighter red are significant at the P<.05 level, with the coronal measurement lower than the labial measurement. Pairs shown in black are not significantly different (P>.05). The effect of speaker was almost always significant, which was to be expected especially given that the speakers were of different genders.

The results of the ANOVAS show that, for the speech signals measured in this particular study, the two measurements which were most useful for distinguishing between coronal and labial underlying forms were the frequency of F2 at vowel offset, and the frequency of F2 from the average spectrum across the nasal closure. For these

		F2 at V midpoint			F2 at V offset			Average F2 across N			Average F2 across /m/	
Context		Cor.	Lab.		Cor.	Lab.		Cor.	Lab.		Cor.	Lab.
/iN#b/		2439	2351		2096	1820		1597	1480			
/iN#m/		2442	2383		2169	1648		1555	1470		1367	1396
/eN#b/		2389	2274		2082	1661		1596	1388			
/eN#m/		2328	2303		1956	1383		1554	961		959	904
/IN#b/		1863	1827		1707	1475		1533	1258			
/IN#m/		1908	1795		1808	1575		1635	1459		1499	1494
/εN#b/		1701	1561		1630	1478		1525	1332			
/εN#m/		1799	1720		1671	1449		1570	1226		1191	1064
/aN#b/		1908	1957		1535	1339						
/aN#m/		1835	1800		1613	1484		1562	1351		1432	1507
/uN#b/		1526	1607		1524	1141		1433	1040			
/uN#m/		1917	1730		1726	1237		1544	1100		1323	1227
/ΛN#b/		1453	1362		1509	1257		1464	1209			
/ΛN#m/		1443	1370		1610	1165		1578	1213		1488	1443
/αN#b/		1239	1153		1371	1084		1397	1096			
/αN#m/		1333	1222		1421	1119		1459	1086		1363	1399
/oN#b/		1251	1123		1201	1071		1184	1040			
/oN#m/		1175	991		1163	912		1103	880		1095	953

Table 1: Average F2 frequencies across vowel contexts for coronal and labial underlying forms. Measurement pairs shown in darker blue are significantly different at the P<.01 level, with coronal frequencies higher than labial frequencies.

two measurements, the differences are always in the direction predicted by Manuel and Stevens (1995), with the coronal frequency higher than the labial frequency; moreover, the differences are highly significant in all vowel contexts for F2 frequency at vowel offset, and in all but two vowel contexts for F2 frequency from average spectrum across nasal closure.

		ΔF2 at V offset			Average ΔF2 across N			Average ΔF2 across /m/	
Context		Cor.	Lab.		Cor.	Lab.		Cor.	Lab.
/iN#b/		-343	-533		-758	-873			
/iN#m/		-274	-735		-859	-972		-1227	-1006
/eN#b/		-308	-612		-798	-882			
/eN#m/		-372	-920		-773	-1342		-1310	-1360
/IN#b/		-157	-352		-332	-569			
/IN#m/		-100	-220					-398	-297
/εN#b/		-70	-87		-172	-239			
/εN#m/		-127	-265		-228	-488		-595	-621
/aN#b/		-372	-618						
/aN#m/		-222	-316		-279	-450		-409	-293
/uN#b/		5	-466		-69	-573			
/uN#m/		-192	-493		-348	-625		-478	-436
/ΛN#b/		56	-105		10	-154			
/ΛN#m/		167	-205		136	-157		47	73
/αN#b/		132	-69		157	-66			
/αN#m/		88	-103		131	-136		34	177
/oN#b/		-50	-53		-56	-83			
/oN#m/		-12	-79		-73	-112		-40	-87

Table 2: Average change in F2 frequency from vowel midpoint to measurement point across vowel contexts for coronal and labial underlying forms. Measurement pairs shown in darker blue are significantly different at the P<.01 level, with coronal changes less negative than labial changes.

The frequencies at vowel midpoint for both F2 and F3 are not always in the predicted direction and are not always significant; it appears that the steady-state portion of the vowel is not as much affected by place of closure of the following consonant as is the offset portion. This finding is similar to those of Dohr (2004), who found no significant F2 differences throughout vowels before coronal and labial consonants, except toward the ends of the vowels. Both F2 and F3 show hardly any significant differences

		F3 at V midpoint			F3 at V offset			Average F3 across N			Average F3 across /m/	
Context		Cor.	Lab.		Cor.	Lab.		Cor.	Lab.		Cor.	Lab.
/iN#b/		2978	2890		2712	2544		2578	2428			
/iN#m/		2984	2877		2700	2534		2562	2457		2457	2451
/eN#b/		2876	2774		2660	2517		2505	2435			
/eN#m/		2791	2745		2602	2465		2515	2471		2501	2480
/IN#b/		2717	2676		2637	2486		2602	2484			
/IN#m/		2769	2708		2749	2566		2609	2480		2416	2445
/εN#b/		2717	2496		2621	2484		2551	2430			
/εN#m/		2742	2618		2685	2533		2644	2575		2496	2607
/aN#b/		2625	2739		2462	2432		2465	2328			
/aN#m/		2682	2747		2598	2522		2556	2468		2421	2476
/uN#b/		2250	2316		2424	2373		2487	2264			
/uN#m/		2594	2606		2570	2477		2504	2481		2397	2446
/ΛN#b/		2846	2842		2777	2722		2736	2484			
/ΛN#m/		2410	2550		2613	2582		2615	2578		2432	2479
/αN#b/		2543	2649		2622	2719		2559	2565			
/αN#m/		2665	2726		2571	2694		2610	2597		2417	2487
/oN#b/		2654	2629		2606	2539		2566	2495			
/oN#m/		2817	2657		2735	2709		2700	2581		2644	2497

Table 3: Average F3 frequencies across vowel contexts for coronal and labial underlying forms. Measurement pairs shown in darker blue are significantly different at the P<.01 level, with coronal frequencies higher than labial frequencies.

in the average spectrum across the /m/ in /Vn#m/ vs. /Vm#m/ sequences, which is not surprising since the word-initial /m/ is the same segment in both sequences. In general the differences for ΔF from vowel midpoint to measurement point are not as significant as the corresponding absolute frequency measurements are.

The result that absolute F3 frequency differences at vowel offset and across the nasal closure are not always significant in the predicted direction is somewhat

		ΔF3 at V offset			Average ΔF3 across N			Average ΔF3 across /m/	
Context		Cor.	Lab.		Cor.	Lab.		Cor.	Lab.
/iN#b/		-267	-346		-400	-461			
/iN#m/		-284	-343		-422	-426		-527	-426
/eN#b/		-215	-258		-371	-340			
/eN#m/		-189	-290		-275	-274		-289	-265
/IN#b/		-81	-191		-116	-192			
/IN#m/		-61	-142		-160	-228		-353	-264
/εN#b/		-97	-21		-177	-56			
/εN#m/		-57	-85		-98	-43		-246	-6
/aN#b/		-163	-308		-104	-301			
/aN#m/		-80	-224		-126	-278		-276	-271
/uN#b/		179	57		243	-148			
/uN#m/		-23	-129		-91	-114		-198	-160
/ΛN#b/		-69	-120		-110	-357			
/ΛN#m/		204	45		205	28		22	-72
/αN#b/		79	71		16	-84			
/αN#m/		-106	-32		-54	-129		-248	-239
/oN#b/		-72	-91		-94	-135			
/oN#m/		-82	52		-117	-76		-173	-160

Table 4: Average change in F3 frequency from vowel midpoint to measurement point across vowel contexts for coronal and labial underlying forms. Measurement pairs shown in darker blue are significantly different at the P<.01 level, with coronal changes less negative than labial changes.

unexpected, given that F3 transitions have been found to be perceptual cues for place of articulation by other researchers (Harris et al, 1957; Gow & Hussami, 1999). However, it has also been found that the steady-state F3 frequency for a given vowel is not constant (Harris et al, 1957), which may make F3 transitions less clear. It appears, then, that the frequency of F2 at vowel offset and nasal closure are the most useful cues for distinguishing underlying coronal and labial forms in potentially assimilating environments. Further analyses in this study will focus exclusively on these two cues.

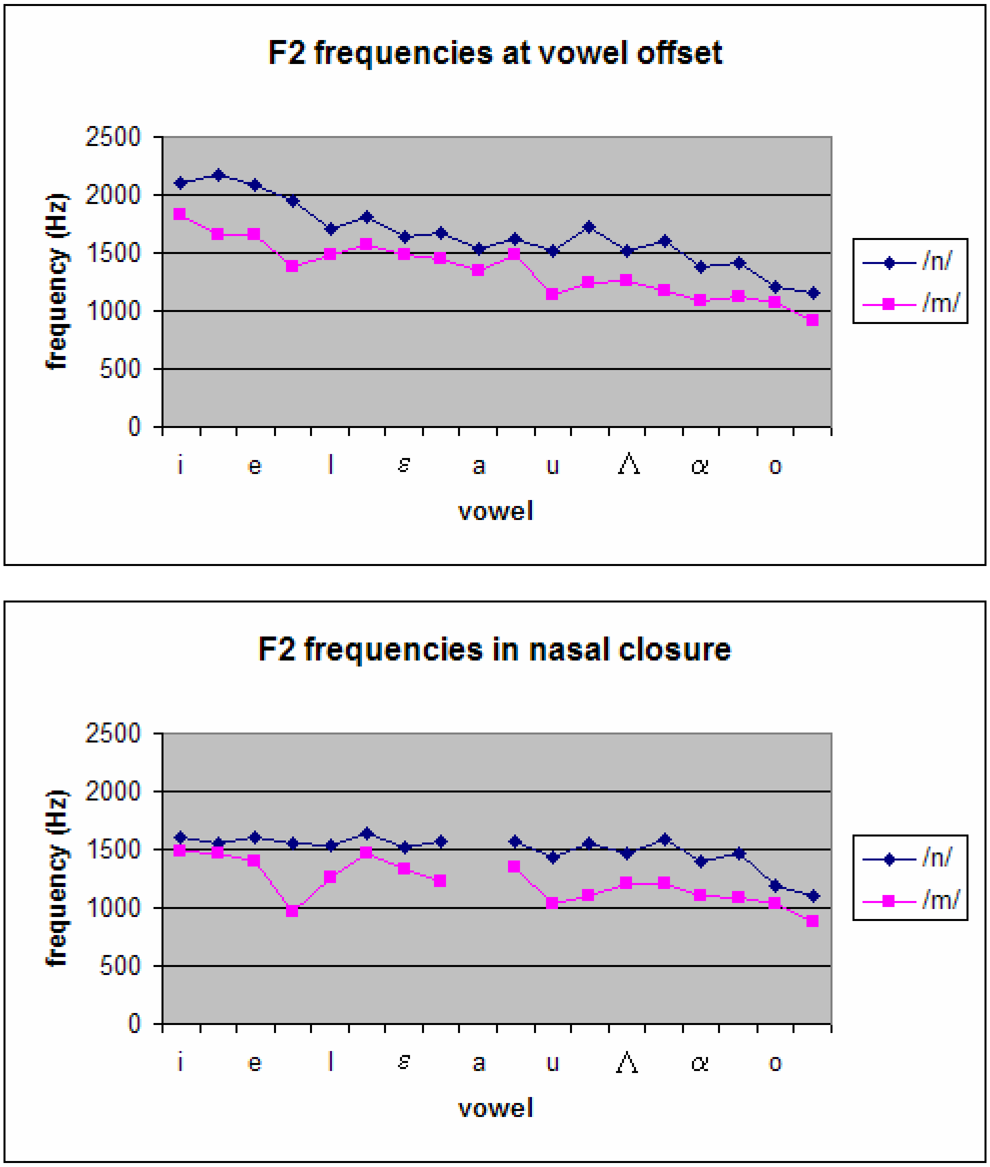

Fig. 4: Plots of average values of F2 frequency at vowel offset (top) and F2 frequency from average spectrum across nasal closure (bottom) for coronal and labial target segments across vowel contexts. The two data points for each vowel context represent /VN#b/ and /VN#m/ sequences.

Figure 4 shows plots of the average values for coronal and labial target segments of the two place cues, F2 frequency at vowel offset and average F2 frequency in nasal closure, across the nine vowel environments used in this study. There is a clear separation between the two line plots in both graphs, indicating that the coronal frequencies are consistently higher than the labial frequencies, as predicted. The back vowels (on the right side of the graphs) in general display relatively larger differences between coronal and labial frequencies for both cues; the tense front vowels (on the left side of the graph) also display relatively larger differences in F2 frequency at vowel offset.

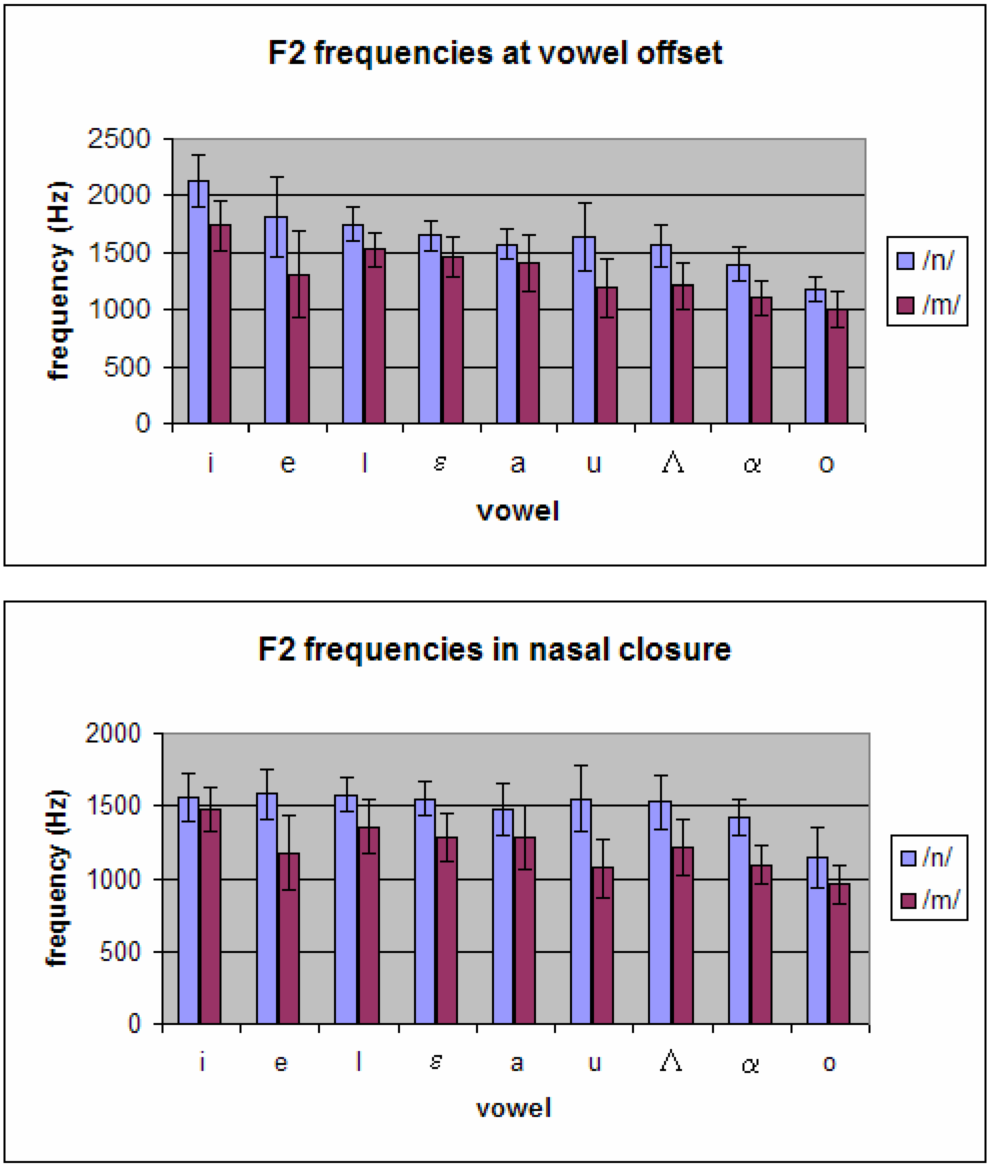

Fig. 5: Bar graphs of average values of F2 frequency at vowel offset (top) and F2 frequency from average spectrum across nasal closure (bottom) for coronal and labial target segments across vowel contexts. In these charts the data have been collapsed across the /VN#b/ and /VN#m/ sequences for ease of viewing. Error bars represent one standard deviation above and below the mean for each data set.

The result that back vowels display larger F2 differences between coronal and labial contexts than front vowels may have an articulatory basis. Back vowels are so named because they are produced with the tongue body in a relatively backed position, resulting in a low value of F2. Front vowels are produced with the tongue body in a fronted position, resulting in a high value of F2. Labial consonants are produced with the lips, and the tongue body is not required to move in the process. Therefore, when a labial consonant is produced after a vowel, the tongue body may essentially remain in position from the preceding vowel, resulting in a relatively lower value of F2 at closure for back vowels, and a relatively higher value of F2 at closure for front vowels. On the other hand, the production of a coronal consonant requires fronting of the tongue body, and the target tongue position for coronal closure may be assumed to be constant across vowel contexts. The value of F2 at closure would thus be nearly the same across vowel contexts, and always higher than for labial consonants. The difference in F2 frequencies between coronal and labial contexts would therefore be larger for back vowel contexts than for front vowel contexts.

There is not an immediate explanation for the relatively larger differences in F2 frequencies at vowel offset between coronal and labial contexts for tense front vowels in this study. It may be that these results reflect random variations in the particular data of this study, rather than more specific trends. Figure 5 shows bar graphs of the average F2 frequencies for coronal vs. labial target segments across vowel contexts, with standard deviations for the measurement sets indicated by error bars. Indeed the error bars for the tense front vowel contexts at vowel offset are relatively large, indicating more variability in these particular data.

In the above analyses the data have been collapsed across assimilation scores for coronals; the following analyses treat each assimilation score separately.

3.2 Analysis by Degree of Assimilation

When the data were segregated by assimilation score, analysis became relatively difficult due to the sparseness of data points. Each speaker produced only four repetitions of each potentially-assimilated coronal target in each vowel context, which would often include only one or two tokens with a given assimilation score. Data were not collapsed across speakers, since each speaker displayed a unique propensity for assimilation. Table 5 shows the frequency (out of 72 total tokens in all vowel contexts) of each assimilation score in the utterances produced by each speaker. A score of 0 indicates no percept of assimilation; a score of 1 indicates some ambiguity between coronal and labial production; a score of 2 indicates complete labial assimilation as judged by the experimenter.

		Assimilation Score	
	0	1	2
Speaker			
MA	15%	43%	42%
LS	0	50%	50%
TH	22%	67%	11%
KH	0	62.5%	37.5%

Table 5: Frequency of place assimilation produced by each speaker. Percentages are based on 72 total utterances.

Speaker TH was the least likely of the four speakers to assimilate the coronal nasal target segments to the following labial segments, although he did produce some tokens with the highest assimilation score. Speaker LS was the most likely to assimilate, producing half of her tokens with the highest assimilation score. Speakers LS and KH never failed to assimilate to some degree, and thus their tokens never received an assimilation score of 0. Speakers MA and TH produced tokens with all three assimilation scores. Clearly assimilation tendencies vary between individual speakers; however each of the four speakers in this study appeared to exhibit a continuum of assimilation forms. The speakers differed from each other in their centers of gravity along the continuum.

Despite the sparseness of data points for individual combinations of speaker, vowel context, and assimilation score, one-tailed T-tests were performed on these data where possible. The F2 frequency at vowel offset and the F2 frequency from the average spectrum across the nasal closure for coronal segments with each assimilation score were compared with the corresponding values for labial segments. Tables 6-9 show the results

MA		F2 at V offset					Average F2 across N			
Context		Lab.	Cor 0	Cor 1	Cor 2		Lab.	Cor 0	Cor 1	Cor 2
/iN#b/		1828		2039	2157		1424		1606	1500
/iN#m/		1618		1898	1953		1325		1406	1375
/eN#b/		1735		2086	1953		1459		1711	1465
/eN#m/		1494	2203	1969	1477		1043	1828	1805	1008
/IN#b/		1447	1992	1629	1570		1318	1828	1453	1313
/IN#m/		1617		1922	1688		1518		1641	1570
/εN#b/		1647		1805	1641		1453		1562	1547
/εN#m/		1570	1945	1711	1547		1400	1805	1653	1266
/aN#b/		1588		1494			1372		1149	
/aN#m/		1635		1711	1629		1547		1676	1547
/uN#b/		1319	1782	1770			1149	1665	1676	
/uN#m/		1360			1565		1141			1336
/ΛN#b/		1500	1594	1696			1442	1570	1688	
/ΛN#m/		1395	1840	1899			1342	1746	1840	
/αN#b/		1272		1477	1344		1225		1617	1289
/αN#m/		1324	1531	1500			1237	1547		
/oN#b/		1231		1242	1289		1149		1078	1187
/oN#m/		1067			1143		992			1015

Table 6: F2 frequencies for speaker MA separated by assimilation score. Coronal measurements shown in darker blue are significantly higher than their labial counterparts at the P<.01 level.

for each speaker. Coronal measurements in darker blue are significantly higher than the corresponding labial measurements (P<.01), etc. Measurements in black are not significantly different from the corresponding labial measurements. Measurements in green did not include enough data points for the performance of a T-test.

Although many data points are missing from these analyses, two points can be made with respect to these results. First, as a general trend for each speaker, a lower assimilation score corresponded with a larger percentage of coronal tokens (out of those

LS		F2 at V offset				Average F2 across N		
Context		Lab.	Cor. 1	Cor. 2		Lab.	Cor. 1	Cor. 2
/iN#b/		1922	2391	2117		1559		
/iN#m/		1688	2367	2285		1547	1910	1430
/eN#b/		1746		1951		1553		1652
/eN#m/		1389	2402	2040		920	1653	1535
/IN#b/		1641	1875	1711		1430	1672	1594
/IN#m/		1682	1953			1606	1688	
/εN#b/		1633	1711	1641		1447	1594	1453
/εN#m/		1635	1764			1359	1676	
/aN#b/		1606	1723	1747		1352	1594	1676
/aN#m/		1641	1723	1746		1506	1688	1582
/uN#b/		1319	1805	1719		1156	1711	1735
/uN#m/		1600	2086			1406	1606	
/ΛN#b/		1389	1647			1383	1547	
/ΛN#m/		1102	1727	1711		1318	1688	1641
/αN#b/		1026	1664	1492		1078	1594	1508
/αN#m/		996		1377		1049		1407
/oN#b/		1061		1049		1061		967
/oN#m/		805	1360	1102		805	1278	867

Table 7: F2 frequencies for speaker LS separated by assimilation score. Coronal measurements shown in darker blue are significantly higher than their labial counterparts at the P<.01 level.

with enough data points to conduct T-tests) whose frequency measurements were significantly higher than their labial counterparts. This helps to confirm the assumption that the experimenter's assimilation scores correspond to the degree of gestural overlap actually produced by the speakers. Second, although an assimilation score of 2 is described as "completely assimilated and /m/-like", in many cases underlyingly coronal segments with assimilation score of 2 show F2 frequency measurements which are significantly higher than their labial counterparts. This suggests that even strongly

TH		F2 at V offset					Average F2 across N			
Context		Lab.	Cor 0	Cor 1	Cor 2		Lab.	Cor 0	Cor 1	Cor 2
/iN#b/		1797		2016			1508		1535	
/iN#m/		1705		2273	2051		1375		1453	1512
/eN#b/		1577	2162				1330	1469		
/eN#m/		1324		1726	1711		949		1500	1547
/IN#b/		1301	1617	1582			1055	1477	1488	
/IN#m/		1389		1664			1277		1562	
/εN#b/		1213	1453	1469			1131	1500	1414	
/εN#m/		1249	1664	1531			1008	1523	1523	
/aN#b/		1037	1453	1406			949	1336	1328	
/aN#m/		1248		1453	1430		1067		1484	1430
/uN#b/		967		1231	1278		930		1453	1477
/uN#m/		1061		1477	1489		961		1407	1547
/ΛN#b/		1067	1336	1305			1008	1289	1242	
/ΛN#m/		1049	1430	1391			1031	1477	1383	
/αN#b/		996	1219	1234			973	1289	1328	
/αN#m/		1020		1325			949		1477	
/oN#b/		1008	1242				967	1400		
/oN#m/		944		1178			914		1161	

Table 8: F2 frequencies for speaker TH separated by assimilation score. Coronal measurements shown in darker blue are significantly higher than their labial counterparts at the P<.01 level.

assimilated coronal forms are differentiable from pure noncoronal forms on the basis of acoustic cues. If these differences are significant to the perceptual systems of listeners, then accurate recognition of underlying forms need not rely on mechanisms of inference from other sections of the speech signal, since the acoustic information within the target segment is already adequate.

KH		F2 at V offset				Average F2 across N		
Context		Lab.	Cor. 1	Cor. 2		Lab.	Cor. 1	Cor. 2
/iN#b/		1735	2215	1957		1430	1723	1664
/iN#m/		1582	2414	2242		1508		1594
/eN#b/		1588	2204	2250		1211	1734	1699
/eN#m/		1324	2109	1930		932	1500	1500
/IN#b/		1512	1719	1594		1231	1547	1359
/IN#m/		1611	1811			1406	1672	
/εN#b/		1418	1635			1295	1582	
/εN#m/		1342	1641	1594		1119	1515	1406
/aN#b/		1125	1516	1430		1078	1383	1289
/aN#m/		1412	1602	1594		1283	1461	1500
/uN#b/		961	1523	1231		926		867
/uN#m/		926	1750	1828		891	1781	
/ΛN#b/		1072	1406			1002	1399	
/ΛN#m/		1113	1442	1453		1160	1477	1418
/αN#b/		1043	1477	1297		1107	1453	1289
/αN#m/		1137	1485	1383		1107	1395	1430
/oN#b/		985	1305	1031		985	1500	914
/oN#m/		832	1078	1109		809	1078	1102

Table 9: F2 frequencies for speaker KH separated by assimilation score. Coronal measurements shown in darker blue are significantly higher than their labial counterparts at the P<.01 level.

4. Discussion

4.1 Problems Encountered

The largest obstacle in extracting significant results from this study was the relative sparseness of data collected. The goal of analyzing gradient, speaker-dependent realizations of place assimilation across vowel contexts produced a large number of factors to be dealt with; each combination of speaker, vowel context, and assimilation score needed to be analyzed separately. Four repetitions of each target by each speaker were not enough to reach statistical significance; however the time constraints of this particular study would not permit more repetitions. Subjects tired of the speaking tasks quickly, so obtaining more repetitions would have required repeated visits on a longer time scale. More speakers would also be required in order to make significant across-population generalizations about the effects of place assimilation, given that considerable variation exists between speakers. The gathering and analysis of much larger amounts of data would require a much larger time frame than that of this current study, however.

Several other challenges were encountered within the logistics of making frequency measurements in this study. Determination of the appropriate DFT window length for each utterance was sometimes difficult, especially if the speaker changed pitch over the course of the target words. In particular, speaker TH displayed rapidly changing pitch contours in his prosodic realizations of the test sentences. It is unclear to what extent the length of the DFT window affects the formant frequency measurements; however even small changes in the spectrum can affect judgment calls about the location of formant peaks, especially in cases where the formant is difficult to distinguish from nearby noise or other resonances.

Identifying the formants within the nasal closure proved to be rather difficult in this study. The first nasal resonance occurs at around 1000Hz, which can be near the vicinity of F2 especially for back vowels. Although the position of F2 could be fairly confidently identified by extrapolating from its trajectory in the vowel offset region, interaction with the nasal resonance may have affected its strength and frequency to some degree. Also, during the nasal closure the spectrum peak between 2500 Hz and 3000 Hz was generally very strong in amplitude, often much stronger than the F2 peak. (See, for example, the bottom portion of Figure 2, in which the peak around 2500 Hz between time index 0.48 and 0.54 is quite strong, while the F2 peak below 2000 Hz is not as clear.) Although this peak was identified as F3, and seemed to vary in a manner consistent with a formant, it is not entirely usual for the F3 peak to be so much larger in amplitude than F2. There is therefore reason to suppose that a second nasal resonance may exist in this frequency region, and may affect the appearance of F3 on the spectrogram. This may partially explain the lack of significant trends in the predicted direction for F3 in this study.

Another factor which may have been overlooked in this study is the coarticulation effect of the context surrounding the target sequences. For example, in each /$C_1V_1N\#C_2V_2C_3$/ sequence, C_1 and V_2 (and perhaps even C_3) may have had coarticulatory effects on the formant transitions in V_1 which were not accounted for. Whether these relatively far-away segments would have produced significant effects is debatable; however it is a possibility that should be investigated. A possible illustration of these effects occurs in this study's data given in Table 1. The F2 measurements at vowel midpoint for the context /uN#b/ appear in red because the average labial measurement is unexpectedly higher than the average coronal measurement, whereas the corresponding figures for /uN#m/ appear in blue because the average coronal measurement is higher than the average labial measurement as expected. The immediate context segments provide no immediate explanation for this difference, but the surrounding context might. When the preceding segment is included, the sequences appear as /ruN#b/ and /tuN#m/. The /r/ segment is characterized by a very low F3, which may impinge upon the frequency range of F2, perhaps helping to cause the anomalous measurements. This helps to explain why the measurement point at the vowel midpoint is not as useful as the measurement point at the vowel offset for determining the place of articulation of the following consonant, since the earlier measurement point has more interfering influence from the preceding consonant.

4.2 General Findings

Despite the obstacles to statistical significance which arise from the particular restrictions of this study, several general suggestive findings can be reported:

1. There is considerable inter-speaker variability with respect to frequency and degree of place assimilation applied (see Table 5). (Future experiments should confirm this point through comparison with natural variation in control sequences without place assimilation.)
2. Segments can be scored as to the degree of place assimilation present, in finer detail than a simple binary choice.
3. For nasal consonants, usable formant frequency information is present during the closure and can be used as cues for place of articulation. This is not possible for stop consonants.
4. The frequency of F2 at vowel offset and the frequency of F2 during nasal closure appear to be the most useful acoustic cues (of those investigated in this study) to underlying place of articulation of potentially-assimilated coronal nasal segments.
5. Underlying /n/ (averaged across assimilation scores) and underlying /m/ are distinguishable on the basis of these two cues across vowel contexts.
6. Evidence suggests that perceptual assimilation scores (based on the experimenter's judgment) correlate with measurements of these two cues.
7. Even strongly-labialized coronal segments (assimilation score 2) can be significantly different from underlying labial segments in these two cues.

Although F2 frequency at vowel offset and across the nasal closure were found to be the most useful cues in distinguishing underlying place of articulation in this study, other cues may exist which were not investigated here. Duration measurements and formant amplitudes, for example, may be other potential acoustic cues for place of articulation in nasal consonants. This study has provided evidence, however, suggesting that the F2 cues may be *sufficient* for distinguishing between underlying coronal and labial forms across vowel contexts. For the data gathered in this study, the average F2 measurements for underlying coronals were significantly higher than their labial counterparts, in all vowel contexts. Even strongly-assimilated coronal forms in general showed higher measurements for these cues than the pure labial forms, although the data were too sparse to make significant pronouncements. These results suggest that acoustic cues are sufficient in many cases for accurate recognition of assimilated segments, without recourse to any inference of phonological processes.

The results of this study do not suggest, however, that the F2 cues used here are foolproof indicators of the underlying place of articulation of the target segment. The presence of overlap between error bars for coronal and labial segments in Figure 5 indicates that, in some cases, the acoustic cue is insufficient to correctly recognize the intended segment. It is hoped that future studies will collect large enough data sets to determine probabilistic distributions of the cues in question, and to set threshold values with appropriate confidence levels for recognition. Future studies should also determine to what extent the cues used in this study are independent of each other. Perceptual studies, perhaps using synthetic speech, may be able to determine the best combination of acoustic cues for successful recognition of place-assimilated speech.

The results of the current study also suggest that place assimilation is a gradient process, with gradient acoustic frequency information which correlates with gradient perceptual and articulatory scales. Although the data are too sparse for statistical significance, the general trend in Tables 6-9 is for unassimilated coronals to show high F2 frequencies, and for the F2 frequencies to decrease as assimilation score increases. The assimilation scores in this study were the subjective judgments of only one listener, but it is hoped that future work will confirm these trends with more objective findings.

5. Conclusions and Future Work

This study has found evidence in favor of articulation-based theories of gradient place assimilation, in which cues within the assimilated segment are sufficient for successful recognition of its underlying form. Acoustic cues were found in this study to be sufficient for discriminating labialized coronal nasal consonants from pure labial consonants in identical surrounding contexts. In particular, the frequency of F2 at vowel offset and the frequency of F2 from the average spectrum across the nasal closure were found to be significantly higher in underlyingly coronal segments than in underlyingly labial segments across vowel contexts. Even strongly assimilated forms were found to differ in these measurements from the pure labial forms, although there was not enough data in this case for statistical significance. In future work hopefully much more data will be collected in order to confirm these trends.

In addition to collecting many more utterance repetitions from many more speakers, future studies may also use more measurement points for each utterance. In particular, several pitch periods before and after consonant closure could be measured in order to chart the shape of the formant trajectories between vowel and consonant. These trajectory shapes may provide more detailed information about the movements and timing characteristics involved in overlapping articulatory gestures. Future work should also compare acoustic analyses of nasals to their corresponding stop consonants, to determine if their place of articulation and formant transitions are truly comparable.

Although acoustic cues in this study have been found to differ between assimilated coronal and noncoronal forms, it remains to be determined whether listeners use these cues to uncover the underlying form, or indeed whether they may rely on these cues exclusively. Future perceptual experiments using synthetic speech could quantify the effects of varying these cues with other factors controlled. Combining these studies with articulatory experiments could also determine the correlation between gestural overlap, acoustic effects, and perceptual results. It is also hoped that the study will be expanded to include velarized as well as labialized coronal segments, and stop consonants in addition to nasals. In this way progress will be made toward the goal of a comprehensive analysis of place assimilation processes and their acoustic effects across contexts in American English. Eventually this work may be expanded to include other languages, in an effort to determine the extent to which the processes are universal to human speech in general.

Appendix 1: Sentences Used in Recordings, Part 1

Test sentences

/i/
Throw me the beanbag again.
Kirk told Spock to beam back to the ship.
The cops arrived, and the teen made a run for it.
Our team made a goal at the last minute.

/e/
He gave the cane back to the old man.
The next day the police came back to the crime scene.
Jim’s coach wanted him to gain more weight.
In order to win, he had to play the game more seriously.

/I/
He emptied the gin bottle before dinner.
Mary and Jim bought a car together today.
Dorothy told the Tin Man she wanted to go home.
Max took the lookout, and Tim manned the defenses.

/ɛ/
She went out to the store, then back to the office.
Billy’s mom was late to take them back home.
He didn’t know that Jenn minored in math.
The cave collapsed, but the gem miners escaped.

/a/
Paul told their mom that Dan broke the window.
The dam broke under pressure and flooded the valley.
Turn off the heat when the pan makes a sizzling sound.
I’ve heard that Pam makes excellent cookies.

/u/
The natives believe that the rune brings good luck.
Renting out the room brings in money for the family.
The words are the same, but the tune may change a bit.
Visitors to the tomb may light candles for prayers.

/Λ/
The winter is cold, and the sun barely comes up.
There are a lot of numbers, but the sum barely reaches 100.
Their coach made them run many laps after school.

That man drinks rum many times in a week.

/α/
Just yesterday Ron built a treehouse for his kids.
I wonder if there’s enough ROM built into this machine.
The judge wanted the con man to serve time in prison.
The captain got the message from the com man just in time.

/o/
Put the ice cream cone back in the box.
She was trying to comb back her hair.
I do check my e-mail, but the phone might be faster.
Pour at an angle, or the foam might spill.

Filler sentences

I ask a question and then they debate.
He saw the woman in the water.
A few busy women began with no money.
The cup is small and Debbie may keep it.
His small cousin was very lazy.
My cousin wanted to support the system.
Tom would never perch on the potato.
It was sudden for this time of year.
She said goodbye to just nineteen women.
From hip to toe the baby was small.
Another big book was in the water.
The box did not contain any money.
Jake had gone without his rabbit.
Below the cups were fifteen big pails.
A number of pigs looked into the canoe.
An ache began beside his toe.
Many other families said goodbye.
Suzie wants to seal up the cookies.
Mary likes to suppose she can leave.
Below the city is a smelly sewer.
A sudden shake made the rope come loose.
The day was long and the weather was bad.
They go by cab, but it costs a lot.
A small number of leaky pails were on the rug.
The box contained just a few cookies.
A bad blow from the system did him in.
He gave her a long look and then left.
Taking his time, Tom went to the zoo.
The city gave the zoo a pig.
Because of the debate the class was postponed.
We took a hike and got lost on the way.
The money is coming today by post.
The busses show a loss for the day.
My cousin can never remember my name.
Drinking too much gave him a headache.
Someone ate up all the food.

Appendix 2: Words Used in Recordings, Part 2

/i/

bead	tee’d
bean	teen
beeb	tee’b
beam	team

/e/

kay’d	gay’d
cane	gain
kay’b	Gabe
came	game

/I/

jidd	tidd
gin	tin
jib	tibb
Jim	Tim

/ε/

thed	jedd
then	Jenn
theb	jebb
them	gem

/a/

dad	pad
Dan	pan
dab	pabb
dam	Pam

/u/

rude	two’d
rune	tune
rube	tube
room	tomb

/Λ/

sud	rudd
sun	run
sub	rub
sum	rum

/α/

rod	cod
Ron	con
rob	cobb
ROM	com

/o/

code	foe’d
cone	phone
cobe	phobe
comb	foam

References

Avery, P., and Rice, K. (1989) Segment structure and coronal underspecification. *Phonology* **6**, 179-200.

Barry, M.C. (1985) A palatographic study of connected speech processes. *Cambridge papers in phonetics and experimental linguistics*. Vol. 4.

Barry, M.C. (1992) Palatalisation, assimilation and gestural weakening in connected speech. *Speech Communication* **11**, 393-400.

Blumstein, S.E., Isaacs, E., and Mertus, J. (1982) The role of the gross-spectral shape as a perceptual cue to place of articulation in initial stop consonants. *Journal of the Acoustical Society of America* **72 (1)**, 43-50.

Browman, C.P., and Goldstein, L. (1989) Articulatory gestures as phonological units. *Phonology* **6**, 201-251.

Browman, C.P., and Goldstein, L. (1990) Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston and M.E. Beckman (Eds.), *Papers in Laboratory phonology I: Between the grammar and physics of speech.* Cambridge: Cambridge University Press. pp. 341-376.

Byrd, D. (1992) Perception of Assimilation in Consonant Clusters: A Gestural Model. *Phonetica* **49**, 1-24.

Coenen, E., Zwitserlood, P., and Bölte, J. (2001) Variation and assimilation in German: Consequences for lexical access and representation. *Language and Cognitive Processes* **16 (5/6)**, 535-564.

Delattre, P.C., Liberman, A.M., and Cooper, F.S. (1955) Acoustic Loci and Transitional Cues for Consonants. *Journal of the Acoustical Society of America* **27 (4)**, 769-773.

Dohr, S. (2004) Acoustic cues to word identity in vowels preceding assimilated consonants. M. Phil. thesis. University of Cambridge.

Ellis, L., and Hardcastle, W.J. (2002) Categorical and gradient properties of assimilation in alveolar to velar sequences: evidence from EPG and EMA data. *Journal of Phonetics* **30**, 373-396.

Gaskell, M.G. (2001) Phonological variation and its consequences for the word recognition system. *Language and Cognitive Processes* **16 (5/6)**, 723-729.

Gaskell, M.G. (2003) Modelling regressive and progressive effects of assimilation in speech perception. *Journal of Phonetics* **31**, 447-463.

Gaskell, M.G., and Marslen-Wilson, W.D. (1996) Phonological Variation and Inference in Lexical Access. *Journal of Experimental Psychology: Human Perception and Performance* **22**, 144-158.

Gaskell, M.G., and Marslen-Wilson, W.D. (1998) Mechanisms of Phonological Inference in Speech Perception. *Journal of Experimental Psychology: Human Perception and Performance* **24**, 380-396.

Gaskell, M.G., and Marslen-Wilson, W.D. (2001) Lexical Ambiguity Resolution and Spoken Word Recognition: Bridging the Gap. *Journal of Memory and Language* **44**, 325-349.

Gow Jr., D.W. (2000) Assimilation, ambiguity, and the feature parsing problem. In *ICSLP-2000*. Vol. 2. pp. 535-538.

Gow Jr., D.W. (2001) Assimilation and Anticipation in Continuous Spoken Word Recognition. *Journal of Memory and Language* **45**, 133-159.

Gow Jr., D.W. (2002) Does English Coronal Place Assimilation Create Lexical Ambiguity? *Journal of Experimental Psychology: Human Perception and Performance* **28**, 163-179.

Gow Jr., D.W. (2003a) Feature parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics* **65**, 575-590.

Gow Jr., D.W. (2003b) How representations help define computational problems: commentary on Grossberg, Gaskell, and Greenberg. *Journal of Phonetics* **31**, 487-493.

Gow Jr., D.W., and Gordon, P.C. (1995) Lexical and Prelexical Influences on Word Segmentation: Evidence From Priming. *Journal of Experimental Psychology: Human Perception and Performance* **21 (2)**, 344-359.

Gow Jr., D.W., and Hussami, P. (1999) Acoustic modification in English place assimilation. *Journal of the Acoustical Society of America* **106**, 2243.

Gow Jr., D.W., and Im, A.M. (2004) A cross-linguistic examination of assimilation context effects. *Journal of Memory and Language* **51**, 279-296.

Gow Jr., D.W., and Zoll, C. (2002) The Role of Feature Parsing in Speech Processing and Phonology. *MIT Working Papers in Linguistics* **42**, 55-68.

Hardcastle, W.J. (1994) Assimilations of alveolar stops and nasals in connected speech. In J.W. Lewis (Ed.), *Studies in General and English Phonetics in Honour of Professor J.D. O'Connor*. London: Routledge. pp. 49-67.

Hardcastle, W.J., and Roach, P.J. (1979) An instrumental investigation of coarticulation in stop consonant sequences. In P. Hollien and H. Hollien (Eds.), *Current issues in the phonetic sciences*. Amsterdam: John Benjamins. pp. 531-540.

Harris, K.S., Hoffman, H.S., Liberman, A.M., Delattre, P.C., and Cooper, F.S. (1957) Effect of Third-Formant Transitions on the Perception of the Voiced Stop Consonants. *Journal of the Acoustical Society of America* **30 (2)**, 122-126.

Hoffman, H.S. (1958) Study of Some Cues in the Perception of the Voiced Stop Consonants. *Journal of the Acoustical Society of America* **30 (11)**, 1035-1041.

Hura, S.L., Lindblom, B., and Diehl, R.L. (1992) On the role of perception in shaping phonological assimilation rules. *Language and Speech* **35 (1,2)**, 59-72.

Jun, J. (1996) Place assimilation is not the result of gestural overlap: evidence from Korean and English. *Phonology* **13**, 377-407.

Lahiri, A., and Marslen-Wilson, W.D. (1991) The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition* **38**, 245-294.

Manuel, S.Y., and Stevens, K.N. (1995) Formant transitions: teasing apart consonant and vowel contributions. In K. Elenius and P. Branderud (Eds.), *Proceedings of the XIIIth International Congress of Phonetic Sciences*. Stockholm: Royal Institute of Technology and Stockholm University. Vol. 4. pp. 436-439.

Marslen-Wilson, W.D., Nix, A., and Gaskell, M.G. (1995) Phonological Variation in Lexical Access: Abstractness, Inference and English Place Assimilation. *Language and Cognitive Processes* **10 (3/4)**, 285-308.

Nolan, F. (1992) The descriptive role of segments: evidence from assimilation. In G.J. Docherty and D.R. Ladd (Eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*. Cambridge: Cambridge University Press. pp. 261-280.

Nolan, F., and Kerswill, P.E. (1990) The description of connected speech processes. In S. Ramsaran (Ed.), *Studies in the Pronunciation of English: A Commemorative Volume in Honour of A.C. Gimson.* London: Routledge. pp. 295-316.

Ohala, J.J. (1990) The phonetics and phonology of aspects of assimilation. In J. Kingston and M.E. Beckman (Eds.), *Papers in Laboratory phonology I: Between the grammar and physics of speech.* Cambridge: Cambridge University Press. pp. 258-275.

Paradis, C., and Prunet, J.-F. (1989) On coronal transparency. *Phonology* **6**, 317-348.

Peterson, G.E., and Barney, H.L. (1952) Control Methods Used in a Study of the Vowels. *Journal of the Acoustical Society of America* **24 (2)**, 175-184.

Schatz, C.D. (1954) The role of context in the perception of stops. *Language* **30**, 47-56.

Stevens, K.N. (2000) Diverse Acoustic Cues at Consonantal Landmarks. *Phonetica* **57**, 139-151.

Stevens, K.N., and Blumstein, S.E. (1978) Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America* **64 (5)**, 1358-1368.

Sussman, H.M., McCaffrey, H.A., and Matthews, S.A. (1991) An investigation of locus equations as a source of relational invariance for stop place categorization. *Journal of the Acoustical Society of America* **90 (3)**, 1309-1325.

Walley, A.C., and Carrell, T.D. (1983) Onset spectra and formant transitions in the adult's and child's perception of place of articulation in stop consonants. *Journal of the Acoustical Society of America* **73 (3)**, 1011-1022.

Zsiga, E.C. (1994) Acoustic evidence for gestural overlap in consonant sequences. *Journal of Phonetics* **22**, 121-140.