Nasal codas in Standard Chinese – a study in the framework of the distinctive feature theory

by

Xiaomin Mou

B.S., E.E., Massachusetts Institute of Technology, 2000

M. Eng., E.E., Massachusetts Institute of Technology, 2001

Submitted to the
Harvard-MIT Division of Health Sciences and Technology
In partial fulfillment of the requirements for the degree of

Doctor in Philosophy
In Speech and Hearing Bioscience and Technology

at the
Massachusetts Institute of Technology

June 2006

Author ____
Harvard-MIT Division of Health Sciences and Technology
May 26, 2006

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

Accepted by ____
Martha L. Gray, Ph.D.
Edwin Hood Taplin Professor of Medical and Electrical Engineering
Co-Director, Harvard-MIT Division of Health Sciences and Technology

This page is intentionally left blank.

Acknowledgement

I thank my advisor, Ken Stevens, for fostering my interest in the field of speech communication, for encouraging me to cross the disciplines of engineering, linguistics, and physiology, and for being a brilliant role model, both in and outside the lab. I thank Stefanie Shattuck-Hufnagel for raising the difficult questions with infectious enthusiasm and helping me to develop a way of looking at the big picture and evolve as a researcher. I thank Suzanne Flynn for hiring me as her UROP student in bilingual acquisition and making learning linguistics fun. I thank David Gow for challenging me to examine the data using other theoretical frameworks. I thank Michael Kenstowicz for inspiring my interest in nasal codas, for reading various versions of the thesis draft, and for offering insight into many of the ideas developed in this work. Outside my thesis committee, I also thank Nelson Kiang for being a great mentor in my pursuit to connect my graduate training to real-world issues.

I thank the members of the Speech Communication Group whose bustling activities are testimony to the complexity of our field. I thank Zhiqiang Li and Seth Cable for the speech recordings and Seth Hall for help with computers. I also thank Janet Slifka who served on my oral exam committee and Joe Perkell who read drafts on topics that eventually lead to this thesis. I thank Arlene, our lab mother, for being the best lab mother. I thank my labmates, especially Xuemin Chi, Sherry Zhao, Lan Chen, Steven Lulich, Tony Okobi, Tamas Bohm, and Virgilio Villacorta, who offered their valuable insights during heated discussions and practice talks, and ultimately pulled me through anxious moments.

I am grateful to my dear friends, Teresa Santos, Lita lee, Asako Masaki, Mary Yang, Gussie Abrahamse, Sara Jablon, Cherry and Jane Liu, Shruti and Chaitra Chandrasekhar, Tara Sainath, Amanda Graves, Jenny Lee, Sara Etemadi, Eva Jud, Elke Hodson, Catherine Lippuner, Melissa Hornstein, Ozge Gozum, Tetsuya Matsuguchi, Stanley Hu, and Bryant Mclaughlin, who cheered me on all the way and never let me give up. And to Josh Cates, I wish you were here, you would have been proud.

Finally, I thank my mother and father for setting an example of excellence and instilling in me a love of learning and questioning, for comforting me during the toughest moments, and for believing in me and helping me discover the courage to be all that I want to be.

This research has been supported by an NIH training grant and the Chyn Duog Shiah Memorial fellowship.

This page is intentionally left blank

Nasal codas in Standard Chinese
– a study in the framework of the distinctive feature theory

By

Xiaomin Mou

Submitted to the Harvard-MIT Division of Health Sciences and Technology
on May 26, 2006 in Partial Fulfillment of the
Requirements of the Degree of Doctor of Philosophy

Abstract

Nasal codas in English and Standard Chinese (SC) are compared to distinguish between the acoustic correlates of language-universal distinctive features and language-specific enhancing attributes. The distinctive feature theory and the theory of enhancement provide a framework for quantifying the acoustic and articulatory patterns observed in the two languages. An acoustic model of nasalization is first presented, in which the area of the velopharyngeal port and the place of oral constriction are varied, in order to observe the behavior of the acoustic correlates for the feature [nasal] and to establish a quantal relation between the continuous displacement of the primary articulator and the acoustic consequence of this displacement. The first two experiments identify differences in the distribution of acoustic correlates of nasalization contained in the vowel transition and the murmur regions in vowel-nasal environments in English and SC. Results for the low vowel /a/ show a mapping based on vowel rather than coda similarity. Acoustic analysis shows that the SC vowel /a/ shifts in the frequency of the second formant (F2) depending on the nasal coda, while the English vowel does not. The SC mid vowel /ə/ shifts in F2 while the SC high vowel /i/ does not. Furthermore, analysis of syllable-initial nasals in Chinese and English shows that the SC nasals behave like the English nasals. The third experiment is a perceptual study in which subjects are asked to make judgments of the place of articulation based on limited portions of stimuli that can be either nasal or non-nasal and contain one of the three vowels. The nasal place of articulation was identified best when the nasal was preceded by the mid vowel /ə/, was identified less when followed by the low vowel /a/, and was identified the worst when the nasal was preceded by the high vowel /i/. Together, the results of these experiments suggest that language-specific constraints play an important role in determining the enhancing attributes that occur alongside language-universal features. The interactions of the distinctive features and the enhancing gestures may lead to differences in the acoustic manifestation of the same feature in different languages.

Thesis Supervisor: Kenneth N. Stevens
Title: Clarence J. LeBel Professor of Electrical Engineering and Professor of Health Sciences and Technology

This page is intentionally left blank

Table of contents

List of figures

List of tables

Chapter 1: Introduction

1.1. Motivation

The human speech production system consists of articulators and structures that undergo continuous movements within particular ranges. The properties of the sounds that are generated, however, fall into distinctive categories. Each sound category can be described by a set of binary articulatory parameters, which we refer to as distinctive features. A change in a distinctive feature in the representation of a word leads to a new sound and potentially to a new word. This binary change is a reflection of the quantal nature of speech, and this is the basis of the quantal theory (Stevens, 1972, 1989, 2002).

The quantal theory, which will be discussed in more detail in the next section, draws its strength from studies of coupled acoustic resonators, of vocal-tract wall effects and of acoustic energy losses, principles of sound source generation in the vocal tract, and observations of the relation between acoustic parameters observed in speech and auditory responses to the sound described by these parameters.

The distinctive features are thought to be universal for all languages, but different subsets of them are used to distinguish sounds in different languages. Furthermore, although a distinctive feature has its defining articulatory and acoustic correlates, additional acoustic properties may arise through articulatory actions that are not specified directly by the feature. These actions, which we call enhancing gestures, may be introduced to augment the perceptual contrast defined by the distinctive feature when the feature occurs in certain phonetic or prosodic contexts that render it less salient (Keyser & Stevens, 2001). Different languages differ in their phonotactic constraints and in their inventory of phonetic contrasts, and we would expect the enhancing gestures to be language dependent. The interactions of the language-independent distinctive features and the language-dependent enhancing gestures could lead to differences in the acoustic manifestation of the same feature in different languages.

This thesis seeks to quantify and interpret the distribution of the acoustic correlates contained in the vowel-transition and nasal-murmur regions in syllable-final nasal consonants in Standard Chinese and English, as one step toward understanding the acoustic and articulatory actions for the distinctive features that are language universal and the enhancing gestures that are language specific. We consider a particular linguistic category, namely, the nasal coda, in two languages governed by different phonotactic constraints, and examine the relationship between that linguistic category and its articulatory-acoustic manifestations.

1.2. The quantal nature of speech

The quantal nature of speech is thought to be a basis for the distinctive feature theory. The articulatory-acoustic relations are said to be quantal because as an articulatory parameter is varied through a range of values, the associated acoustic parameters change from one state to another. This non-linear relation can be represented schematically in Fig. 1.1 (Stevens, 1989), which shows a hypothetical relation between some acoustic parameter in the sound radiated from the vocal tract and some articulatory parameter that takes on some arbitrary values.

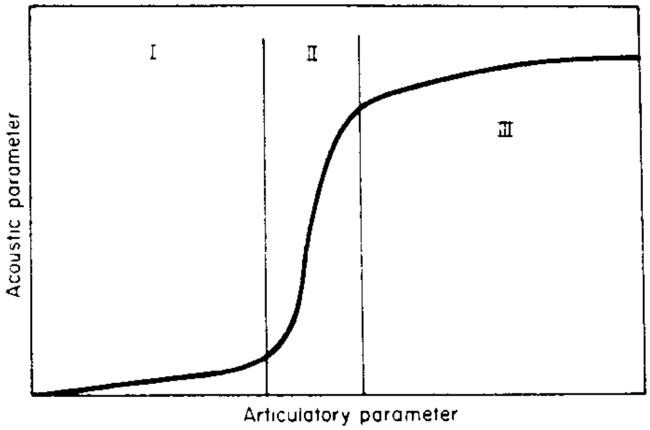

Figure 1.1 Schematization of the change in an acoustic parameter as an articulatory parameter is manipulated, (from Stevens, 1989).

The articulatory and acoustic attributes that occur within the plateau-like regions of I and III are effectively the correlates of some distinctive feature. Within these two regions, the articulatory structures are close to the target state specified by the relevant distinctive feature, moderate changes in articulatory position can occur without significantly altering the relevant attribute in the acoustics. In region II, on the other hand, small changes in the articulatory position can lead to large changes in the acoustic parameter. In this unstable region, the articulatory structures will tend toward either region I or III, and there is a threshold whereby the unstable state shifts to the target state specified by the distinctive feature.

An example of a threshold effect in a quantal auditory-acoustic relation can be illustrated by vowel nasalization. When a vowel becomes nasalized, acoustic coupling of the oral cavity to the nasal cavity through the opening of the velopharyngeal port introduces, among other things, an extra spectral peak or perturbation in the region normally occupied by the first formant. The major effects on the spectrum are the addition of a spectral peak or shift in the frequency of F1 and a decrease of the amplitude of the F1 peak in the spectrum. Stevens, Fant and Hawkins (1987) added a suitably located pole-zero pair to the vocal-tract transfer function of a synthetic vowel. They manipulated the spacing between the pole and zero and asked listeners from different language backgrounds to make nasal-non-nasal judgments. Listeners made similar identification responses regardless of whether a nasal-non-nasal distinction for vowels existed in their language. The threshold for the nasal-non-nasal distinction appeared when the maximum perturbation introduced into the basic non-nasal F1 prominence in the vowel spectrum was in the range 6-9 dB.

The quantal articulatory-acoustic relation in Fig. 1.1 suggests that if the articulators moved to the more extreme configurations in region I or III, the acoustic parameters will show a greater contrast. This implies that different speakers could in principle implement particular features and feature combinations differently, sometimes depending on the degree of intended perceptual contrast. Studies of speaker differences in articulator coordination have been carried out by Johnson, Ladefoged, Lindau (1993) under the assumption that a relatively small set of articulatory features are universally used in the sound systems of the world’s languages. From x-ray microbeam pellet trajectories during the production of vowels of American English by five speakers, they found differences in the patterns of articulation. They point out, for example, that the "r" sound in English can be produced either by using retroflexion or tongue bunching, as long as the sound can be recognized. Individual differences in speech production may be interpreted as indirect evidence that the acoustic goal determines the organization of speech articulation, rather than spatiotemporal targets or gestures. Speakers thus have at their disposal additional shaping of the vocal tract to meet the same acoustic goal.

Languages differ in their contrasts and acoustic goals and we might expect that the value of a particular articulatory or acoustic parameter used to implement a feature in one language will be somewhat different from that in another, even though both implementations may lead to a similar perception. Often, some additional attribute or cues might be introduced, in addition to the defining attributes, so that the acoustic manifestation of a distinctive feature may be quite different.

We make the assumption that the distinctive features in a mental lexicon are instructions to articulators and that the articulatory gestures are intended to achieve acoustic and perceptual goals. Keyser and Stevens (2001) have proposed that the enhancement process operates as a quality control device between the articulatory information and the acoustic information stored in memory. A speaker would presumably accesses additional shaping of the vocal tract if the intended acoustic output is not sufficiently distinct from a nearest neighbor. They present the example of /ʃ/ and /s/. The distinctive feature [-anterior] for the consonant /ʃ/ requires the tongue blade to be placed so that the constriction is made behind the alveolar ridge. The primary acoustic correlate of a [-anterior] obstruent is that the lowest spectral prominence of the frication noise be in the region of F3. But the placement of the tongue blade may only lead weakly to this acoustic property. Lip rounding is a gesture that most speakers of English use to strengthen this acoustic property. Here we have a situation where implementing the articulatory gestures specified by a feature results only in a weak acoustic correlate and that the additional gesture of lip-rounding is introduced to strengthen the acoustic contrast between /ʃ/ and /s/.

1.3. Lexical access model of speech perception

One model of speech perception that provides an account of the process whereby the analog acoustic signal is interpreted as a sequence of discrete phonological units is the land-mark and cue-based model of distinctive feature recognition for lexical access (Stevens, 2002). In this model, features are thought to be binary and universal and are defined in terms of articulatory gestures and the resulting distinctive acoustics. Words are presumably stored in memory as sequences of discrete segments, each of which is comprised of a set of features called the feature bundle. Speech production involves the preparation of feature bundles and their implementation by articulators; speech perception calls for the retrieval of the underlying features from acoustic cues in the speech signal.

This model is the basis for an approach to speech recognition proposed by the Speech Communication Group at MIT. This thesis research fits into a component of such a knowledge-based speech recognition system, which uses landmarks to guide the search for distinctive features. Landmarks identify localized regions in a speech utterance, where the acoustic manifestations of the distinctive features are most salient. They are considered the starting points in time in the speech signal to look for the acoustic correlates of the distinctive features that correspond to each segment.

According to the lexical access model, the processing of the speech signal proceeds in five steps. The first step is to identify acoustic landmarks by detecting peaks, valleys, and discontinuities in particular frequency ranges of the signal. The type of landmark corresponds to a subset of distinctive features called articulator-free features, such as [vowel] or [consonant]. The second step is to derive acoustic parameters from the signal in the vicinity of the landmarks that correlate to the actions of the articulators responsible for the type of landmark and to extract acoustic cues by sampling selected attributes of these parameters in these regions. The third step is to combine the acoustic cues and to give estimates of "articulator-bound" features associated with each landmark, such as [lips] or [nasal]. The fourth step is to parse the features into bundles. The articulator-bound features and the articulator-free features together make up the feature bundles that are the output of the model. The final step is to determine what words in the lexicon are consistent with these features, and then to determine which of these words to select.

Associated with each feature is a set of gestures for various articulators, including, but not limited to, the defining features. In implementing a sequence of segments, the gestures for a feature in one segment may overlap with some gestures for a feature in some adjacent segment. Lindblom (1979) has suggested that gestural overlap has the function of shortening the utterance and thereby reducing the overall speaking effort, as a low-cost motor behavior. In most cases, evidence for the distinctive features that characterize each segment still remains observable in the signal. The timing of articulatory movements seems to be organized to achieve perceptual goals, such as the proper landmarks and the events around these landmarks. Yet in some cases, overlap leads to the absence of some of the acoustic correlates for the distinctive features in the speech signal. Entire segments may even be deleted, or appear as allophonic variants. Nevertheless, the sounds are still perceptible to the listener. Presumably, some other acoustic cues have become present to enhance the weakened or even absent distinctive features. A robust lexical access model should have not only a representation of the distinctive features associated with the underlying lexical item but also knowledge of the surface representation and enhancing gestures that can be language and context dependent.

The goal of this thesis is to evaluate the role of language-specific enhancing gestures that are language specific in speech production and perception, by comparing the production and perception of the nasal coda in English and Standard Chinese, two languages that differ greatly in the number of contrasts and in the inventory of vowels. A goal of this study is to incorporate findings about the role of enhancing gestures into an algorithm for classifying the nasal consonant place of articulation, which captures not only the universal features of nasalization but also some important language-specific phonotactic constraints.

1.4. Nasals in English and Standard Chinese

The perception of nasality is thought to arise from the opening of the velopharyngeal port, whereby a pathway for air from the pharynx to the nose is created. This opening is believed to be responsible for the binary distinction between nasal and oral segments.

The binary values of the distinctive features describing the three nasal consonants in English and Standard Chinese are shown in Table 1 below (Stevens, 1998; Duanmu, 2000).

Table 1.1 Feature values for nasal consonants in English and Standard Chinese

Feature	m	n	ŋ
Vocalic	-	-	-
Consonantal	+	+	+
Continuant	-	-	-
Sonorant	+	+	+
Labial	+		
Coronal		+	
Dorsal			+
Nasal	+	+	+

Table 1.1 indicates that nasals in both English and Standard Chinese are specified by the same features. They are all consonantal and sonorant. /m/ is produced with a closure at the lips, while /n/ is produced with a constriction made with the tongue blade and /ŋ/ is produced with a constriction made with the tongue body.

Standard Chinese (SC) is the result of half a century of reform after the founding of the People's Republic of China in 1949 to establish a standard spoken language. It is based on the pronunciation of the Beijing dialect. Other terms for Standard Chinese are Beijing Mandarin, Standard Mandarin, Mandarin Chinese, or simply Mandarin (Duanmu, 2000). In Standard Chinese, /m/ does not occur in the syllable-final position.

In English, the velopharyngeal orifice is usually thought of as closed for all sounds except /m/, /n/, and /ŋ/. For these sounds, the lowering of the velum opens a passage from the pharynx to the nasal cavity, which allows air to escape through the nasal cavity. Nasal consonants are produced with a complete closure of the oral tract with a clear pathway maintained from the glottis to the nose, so that some air can flow through the airway between the glottis and the output of the vocal tract without pressure buildup above the glottis. The bilabial /m/ is produced with a closure at the lips. The alveolar /n/ is produced with a closure formed at a point about 1.5 to 2.5 cm posterior to the lip opening. The velar /ŋ/ is produced with a high tongue body position with a constriction farther from the lips, about 5 to 6 cm (Stevens, 1998). The opening of the orifice for nasal consonants also tends to nasalize vowels that precede and follow these consonants (Krakow & Huffman, 1993). In English, /ŋ/ never appears in the beginning of a word.

Although the opening of the velopharyngeal port may trigger the perception of nasality, the degree to which a speaker wishes the listener to perceive that contrast can be controlled by manipulating the physiological parameters that are responsible for the acoustics. For example, the size of the velopharyngeal port and the timing of the opening can in principle be controlled by the speaker.

In Standard Chinese, the only consonants allowed in syllable-final position are the nasals /n/ and /ŋ/. The nasal codas in Standard Chinese have no other consonant codas to compete with for perceptual contrast. This unique position may influence the way individuals express the nasal-non nasal contrast and we might expect that the acoustic manifestations of the distinctive features for Chinese nasal codas are somewhat different from those of the English nasal coda.

In addition, there are only five underlying vowels in Chinese (Duanmu, 2002) while there are approximately twelve (monothongs) in English (Ladefoged, 2005; Chomsky & Halle, 1968). Table 1.2 (Duanmu, 2000) shows the feature values for the five underlying SC vowels. Out of these five underlying SC vowels, only the high vowel /i/, mid vowel /ə/, and low vowel /a/ can precede both the alveolar and velar nasal. These three vowels will be the focus of study in this thesis. Note that the mid vowel /ə/ is a fully specified vowel and should not be confused with the English schwa. This thesis will adopt this convention used by Duanmu's (2002) treatment of SC vowels.

Table 1.2 Feature values for Standard Chinese vowels

Feature	i	u	y	ə	a
High	+	+	+	-	(-)
Low	(-)	(-)	(-)	-	+
Back	-	+	-		
Round	-	+	+		
Labial		√	√		
Coronal	√		√		
Dorsal	√	√	√	√	√

( ) indicates predictable values of articulator-bound features; changeable or irrelevant values are left empty; '√' indicates presence of an articulator

The features that make up the phonemes of a language are thought to possess a hierarchical structure (Halle, 1995) based on the observation that only a small fraction of the logically possible pairs of features occur in actual phonological rules. The feature tree splits the universal list of features into mutually exclusive subsets of features and groups the subsets into higher-order sets.

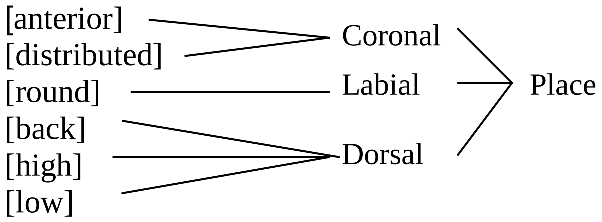

Figure 1.2 Part of the feature geometry tree illustrating the hierarchical structure of features associated with the three articulators that are associated with the place node.

The features [low], [high], and [back] are located under the Dorsal (tongue body) node, while the features [anterior] and [distributed] are located under the Coronal node. Halle points out that in many languages nasal consonants in coda position modify the 'Place" from the onset consonant of the following syllable. He gives the example that in

Sudanese Arabic the Coronal nasal [n] becomes [m] before the Labial [b], and [ŋ] before the Dorsal [k x]. In Standard Chinese, the nasal [n] has the feature [coronal] and the mid and low vowels are unspecified for the feature [back]. Duanmu considers this unspecification to allow the articulator (tongue body) to take on changeable values. In the feature tree shown in Figure 1.2, the articulators are represented by the features assigned to them and the Place node dominates the three articulators. The Place node is believed to have an anatomical motivation in that it combines three articulators that are adjacent to one another. In the case of the SC coronal [n] and the mid and low vowels, the SC the node for the tongue body and the node for coronal could presumably be brought together in their places of articulation. The SC dorsal [ŋ] and the mid and low vowels would also be brought together in their places of articulation, although the extent of modification is presumably less than that required for the coronal [n] since [ŋ] is already produced by the same articulator as the mid and low vowels.

This modification of the place of articulation based on feature geometry could be the articulatory basis for one of the phonotactic constraints observed in Standard Chinese, which restrict the number of logically possible combinations of features. This constraint, called Rhyme Harmony, states that the nucleus and the coda should not have opposite values in [back] or [round], (Duanmu, 2002). In VN (vowel-nasal) rhymes, the alveolar [n] is specified for the feature [coronal] and the velar [ŋ] is specified for the feature [dorsal], while the SC vowel /a/ is unspecified for the feature [back]. The feature geometry would predict some kind of modification to take place when the vowel /a/ precedes the nasal codas. Specifically, we would expect /a/ to be anterior when it is followed by the coronal [n] and [back] when it is followed by the velar [ŋ].

This kind of modification of the vowel by the place of articulation of the neighboring consonants is also seen in other languages. Studies of Marshallese, a Austronesian language whose vowel inventory is highly restricted in its phonetic distribution (Choi, 1995, 1992) have shown that when the articulation of C_1 and C_2 differ from each other in a C_1VC_2 environment, the resulting vowel changes quality, such that the resulting phonetics arise from the consonantal influences. Marshallese has a vowel inventory of only three vowels consisting of a high, a mid, and a low vowel, and they are all unspecified for the feature [back]. Choi found that in symmetrically palatalized environments, only front vowels occur, in symmetrically velarized environments, only back unrounded vowels occur; and in symmetrically rounded environments, only rounded vowels occur.

In SC, only /a/, /i/ and /ə/ can occur before both the velar and alveolar nasal and in addition, /a/ and /ə/ are unspecified for the feature [back], as Table 1.2 shows. It would be very probable for those SC vowels, which are unspecified for the feature [back], to behave like the Marshallese vowels, in that we might expect the SC vowel to become fronted before the alveolar [n] and backed before velar [ŋ]. The difference in the place of articulation of the nasal coda could potentially modify the frontness or backness of the Standard Chinese vowel that is unspecified in the feature [back].

The rhyme harmony constraint does not apply in English, which has a larger vowel inventory. In English, for the front vowel /æ/, back vowel /ɑ/, alveolar /n/, and velar /ŋ/, we have four possible Vowel-Nasal (VN) combinations: [æn], [ɑn], [æŋ], and [ɑŋ] respectively. Manuel (1990) has proposed that although languages occupy a universally available articulatory-acoustic space, this space is divided up differently, so that languages differ in their inventories of distinctive sounds and systems of contrast. A language with a closer vowel space such as English would have a smaller range of production and a language with a more diffuse vowel space, such as Standard Chinese, could tolerate larger production ranges.

In the lexical access model, the influence of the nasal coda in SC on the preceding vowel would be considered an enhancing gesture. Together with some universal attributes of nasalization, this enhancing gesture helps to distinguish the place of articulation in nasal codas in Chinese. On the other hand, the place of the articulation of the nasal coda would have a much less effect on a preceding high vowel [i], which is already specified for the feature [-back] in Standard Chinese. The high vowel would presumably not be able to tolerate a large production range, and we would expect the vowel-nasal combinations, where the vowel is [i], and the nasal [n] or [ŋ], to be distinguished from one another based on acoustic cues not so much in the vowel region but rather in the nasal murmur region.

1.5. Literature Review

1.5.1 Acoustic and articulatory description of nasalization

According to the acoustic theory of nasalization for vowels (Fant, 1960; Fujimura and Lindqvist, 1971), the main difference between nasal and oral vowels is that additional formants are introduced as a result of the coupling of the nasal and oral tract. Sweep-tone measurements of the vocal tract transfer functions of nasals and nasalized vowels (Fujimuar & Lindqvist, 1971), combined with the more traditional model of the nasal tract (House and Stevens, 1956, Fujimura, 1960), have provided the basis for the acoustic study of nasalization.

The lowest pole for the transfer function for nasal consonants is estimated to be in the range of 250 to 300 Hz, which is a Helmoholtz resonance between the acoustic compliance of the vocal tract volume and the acoustic mass of the nasal passages, which are convoluted, with extensive surface areas (Chen, 1997). The second pole of the transfer function during the nasal murmur is in the range of 750-1000 Hz.

The production of oral-nasal segments requires a speaker to manipulate the size of the velopharyngeal port to either prevent or enhance the acoustic and aerodynamic coupling of the nasal tract to the oropharyngeal tract (House & Stevens, 1956). Depending on the set of contrasts in a particular language, there can be differences in the implementation of particular features or feature combinations.

One way these differences may be manifested is in the cross-sectional area of the opening in the velopharyngeal port. In order to maintain sonorancy, the cross-sectional area of the opening in the velopharyngeal port should be no less than 0.2 to 0.3 cm^2 and the total time from the beginning to the end of the lowering-raising movement of the soft palate is about 200 to 250 ms, based on the timing in a typical utterance in English (Stevens, 1998). Comparisons of normal and impaired nasal resonance (Warren, Dalston, et al., 1993) have shown that the velopharyngeal opening for nasal consonants must be larger than 0.2 cm^2 for speech to be perceived as normal in adults and that an opening less than 0.2 cm^2 physically restrains airflow through the nose and the speech outcome is often perceived by a listener to be hyponasal. The size of the velopharyngeal opening was derived from measurements of the pressure difference between the nose and the mouth and the volume rate of the airflow through the nose during speech and shown to be 0.5 to 1.0 cm^2 for nasal consonants and greater than 1.0 cm^2 during breathing through a pressure-flow technique (Warren, 1982).

Differences in the cross-section area of the velopharyngeal port and in the timing of lowering and raising the soft palate, can potentially lead to different acoustic manifestations of nasality. In addition, speakers differ in the properties of their vocal tracts, and they may manipulate other factors, such as spreading the glottis to allow more coupling between the subglottal system and the rest of the nasal system, to possibly further reduce the amplitude of the second formant at the nasal landmark, which is considered one of the acoustic correlates of nasalization.

1.5.2. Acoustic correlates of nasalization

Studies of articulatory-acoustic relations have given a robust picture of the articulatory events surrounding a nasal landmark and the acoustic correlates of nasalization (Stevens, 1998, Chen, 1997). During the interval of a few tens of milliseconds prior to the oral closure and following the release of the nasal consonant, there is coupling of the nasal cavity to the oral cavity and the output is a combination of the volume velocity from the nose and the mouth. The rapid change in the frequency of the low-frequency prominence at the closure and at the release indicates the nasal landmark in the speech waveform. The nasal landmark would be the starting point for the speech recognizer to look to the left and right of the speech signal for acoustic attributes and cues that are correlated with the articulator-bound features, such as [coronal], [dorsal] and [back] and [high].

The single pole corresponding to the first formant of the non-nasal transfer function is replaced by the pole-zero-pole triplet, the lower pole in the range of 300-900 Hz and the higher one in the range of 600 to 1200 Hz, and the zero usually above the nasal pole. The zero will reduce the amplitude of the first and second formants. During the period where the vowel is non-nasal preceding the closure or following the release, one of the poles merges with the zero, leaving a single pole that is the first formant of the non-nasal vowel. The abrupt shift in the first zero Z1 has a large effect on the spectrum of the sound, and is expected to account for a significant change in the spectrum amplitude in the vicinity of the boundary, particularly in the region of the second formant (Stevens, 1998). This results in an abrupt increase in spectrum amplitude in the vicinity of the frequency of the second formant. This is the frequency region in which there was a zero during the nasal murmur, which is gone at the release of the nasal.

Several acoustic correlates of nasalization can be tracked throughout an utterance in order to make comparisons of these trajectories between the Standard Chinese and English codas. The reduction in the amplitude of the first formant spectral peak (A1) has been observed to be the primary cue of nasalization. Perceptual experiments using synthetic stimuli showed that lowering F1 amplitude by 6-8 dB is necessary to achieve a significant level of nasality perception (Stevens, 1998). Nasal coupling during vowel production was found to introduce a pole and a zero in the region of the first formant, and a spectral peak around 1 kHz (House and Stevens, 1956). Another nasal peak between 250 and 450 Hz has been observed and this has been attributed to the pole-zero pair introduced by the paranasal sinuses (Chen, 1997).

Chen (1997) made acoustic analysis of vowel nasalization in the frequency domain and noted the presence of these extra peaks. Chen designated the amplitude of the second peak, found between the first two formants, as P1, and the amplitude of the lower peak as P0. She confirmed that the first-formant amplitude, A1, was reduced relative to its amplitude for an oral vowel. She quantified the extent of vowel nasalization by determining the values for the acoustic correlates A1-P1, and A1-P0 in dB. The lower the value of A1-P1 or A1-P0 is, the higher the extent of vowel nasalization. Because the second nasal peak is near 1 kHz, it can be obscured by the F2 of a back vowel, such as /A/, which occurs at about 1100 Hz. Similarly, the first nasal peak can be obscured by the F1 of a front vowel, such as /i/, at about 300 Hz. Therefore, the acoustic correlate A1-P1 becomes more robust when we examine vowel nasalization of front vowels and A1-P0 becomes more robust when we examine back vowels.

Another acoustic correlate of nasalization is the change in the amplitude of F2 at the discontinuity between the vowel and the nasal murmur. At the transition from the vowel to the nasal murmur in the vicinity of this landmark, the amplitude of F2 should decrease significantly and F2 should also shift noticeably.

The spectrograms on the top panel of Figure 2.2 show the male English speaker's production of 'dɑn' on the left and 'dɑ' on the right. There is an abrupt decrease in the amplitude of F2 at the landmark of the nasal coda, at about 450 ms. The peaks in the spectrum on the bottom left panel indicate the presence nasality with a near zero A1-P0 and the peaks in the spectrum of the bottom right panel indicate non-nasality with a positive A1-P0 of about 10 dB.

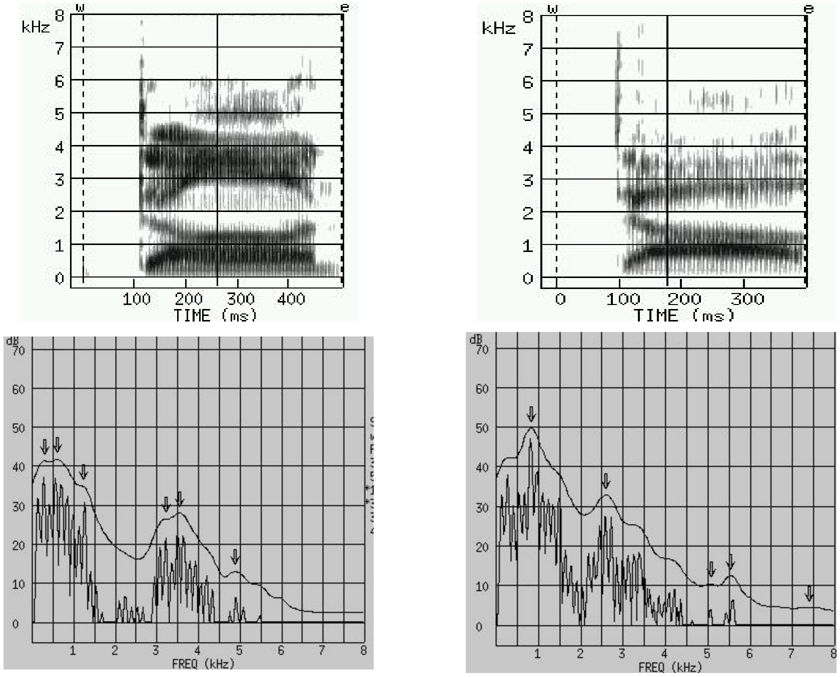

Figure 1.3 Spectrograms of a male English speaker’s production of ‘dɑn’ on the left and ‘dɑ’ on the right (top), and spectra of ‘dɑn’ and ‘dɑ’ (bottom) sampled during the vowel.

While this thesis is concerned with the acoustic correlates of nasalization, it focuses mainly on those acoustic correlates that are responsible for distinguishing the place of articulation of the nasal coda. The acoustic correlates of nasalization are used to build a general and simple acoustic model which captures the poles and zeros of the nasalized vowel and the nasal consonant. This acoustic model is language-independent and based on our knowledge about nasal production in terms of the place of oral constriction and the cross-sectional area of the velopharyngeal port. This model serves as a basis upon which additional acoustic correlates may be introduced or existing acoustic correlates may be modified in the regions of the vowel or nasal murmur to enhance the contrast in the place of articulation for the nasal coda.

One of these additional acoustic correlates that will be the focus of this thesis is F2, the second formant frequency, which correlates with the feature [back]. The second formant frequency is a resonance of the cavity between glottis and the constriction in the oral tract. It indicates the front-back position of the tongue body. A higher F2 indicates a more fronted tongue body and a lower F2 indicates a more backed tongue body. The relative front-back tongue body position of the nasal and the nasalized vowel may be estimated from the F2 measured at the landmark between the vowel and the nasal murmur and tracked several 10s of milliseconds before the landmark and after it.

1.5.3 Classification experiments

A number of classification experiments have examined the relevant contributions of the acoustic cues contained in the vowel region and in the nasal murmur region to the correct identification of the place of articulation for nasals (Recasens, 1982; Kurowski & Blumstein, 1987; Repp & Svastikula, 1998, Harrington, 1994). These studies have focused mainly on English and other languages that have the nasal and oral vowel distinction, such as French and Portuguese. Recasens examined the perceptual contributions of formant transitions and murmurs to place identification in Catalan [m n ŋ] and found that formant transition provided more cues than murmur although murmur made significant contribution to [n ŋ] distinction. Kurowski and Blumstein experimented with a combination of murmur and transition for perception of nasal place of articulation [m n]. They found that the combination that consisted of six glottal pulses on either side of the transition cued place of articulation best. Repp and Svastikula studied syllable-final vs. syllable-initial nasals [m n] and found that spectral change provides less information about place of articulation in VN than NV due to lack of salient spectral change between vowel and nasal coda. Harrington was able to detect the nasal landmark based on abrupt change in energy and in formant frequencies on spectrogram displays. These studies show that the acoustic cues necessary for identifying nasal place of articulation are found in the nasal murmur as well as the formant transitions.

1.5.4 Studies of nasality in Standard Chinese

Nasal and oral airflow studies have shown that the Standard Chinese nasal codas are produced somewhat differently from English nasal codas. Wang (1993) examined the spectra of the intervocalic nasal onset and coda in the Chinese words “fa nan” and “fan an” and found them to be distinct from each other. She measured nasal and oral airflow with an airflow detector. Her results showed that airflow measured at the nose differed significantly between the nasal onset and coda, which were in identical vowel environments. For the nasal onset, nasal airflow was nearly 100%. For the nasal coda, however, nasal airflow was only 68% of the total. That is to say, nearly one third of the airflow passed through the mouth. Wang suggested that the Chinese nasal coda is actually a nasal approximant and that it is more like a semivowel produced without a complete closure in the vocal tract.

Lin and Yan (1991) found that the tongue position at the end-point of the vocalic portion in a vowel-nasal (VN) syllable, where the nasal is either the velar or the alveolar, plays a role in the recognition of the nasal coda. The end-point they refer to is essentially the nasal landmark this thesis refers to. They found significant differences between the F2 measured at the end-point for the non-high vowels that are followed by the velar vs. the alveolar nasal codas. They also found a smaller difference between F2 measured at the end-point for the high vowels that are followed by the velar or alveolar nasal coda (See Appendix A). However, Lin and Yan focused only on the end-points. Although F2 differences reflect tongue body position, measurements at the end-point only confirms an overlapping of the gestures that are required to make the vowels and the gestures that are required to make the nasal coda. It is the F2 movement throughout the vowel that we are interested in, which may reflect a movement of the tongue body to enhance the recognition of the nasal coda.

1.6. Thesis outline

The goal of this work is to examine through acoustic measurements and perceptual studies the differences in the distribution of acoustic cues in the murmur region and the vowel transition region in syllable-final nasal codas in English and Standard Chinese. We will first explore what is language universal for the feature [nasal], and then what is language dependent in terms of enhancing gestures, by making comparisons of production and perception of the nasal sound in English and Standard Chinese. We will review the existing acoustic model of nasalization and vary some parameters to explore the relation between articulatory and acoustics of nasalization. Next, we will examine possible differences in the distribution of acoustic correlates of nasalization that are contained in the vowel and murmur regions English and Standard Chinese. Then we search for differences in the distribution of acoustic correlates of nasalization in all possible vowel-nasal environments as well as in the nasal-vowel environments. We then examine how listeners use differences in the distribution of the acoustic correlates to make judgments of the place of nasal articulation.

The first study is a parametric model of nasalization, which explores the relationship between the physical opening of the velopharyngeal port and the quantal nature of the nasal feature, as an attempt to relate the continuous and graded opening of the velopharyngeal port area to the abrupt changes in the acoustics that signal a nasal-nonnasal distinction.

The second experiment compares differences in the distribution of acoustic cues for the nasal place of articulation between English and Chinese nasal codas. Native speakers of SC are asked to listen to utterances in the form of AXB, where A and B are SC syllables, and X is the matched English syllable. The syllables are of the CVN format, where C is a consonant, N is either the alveolar /n/ or the velar /ŋ/, and V can be one of several vowels, including /a/ and /i/. The subjects are asked to choose from the A and B the SC syllable that sounds closest to the English syllable X. We expect that the choices the subjects make would reveal the distribution of acoustic cues for the SC nasal coda, relative to the distribution in English. The results of the perceptual experiment will be discussed with the aid of acoustic analysis.

The third experiment is an acoustic analysis that focuses on the weighting of the acoustic correlates of nasalization in vowel-nasal coda environments, where the vowels are the high vowel /i/, mid vowel /ə/, and low vowel /a/. Nasals in the syllable-initial position will also be examined. Because the phonotactic constraint that may have been responsible for the weighting differences in the coda position between English and Chinese does not apply for the syllable-initial position, we would not expect any differences in the weighting of acoustic cues between the two languages for nasals in this position.

The fourth experiment is a perceptual experiment that asks subjects to listen to different segments in the vowel transition or the nasal murmur regions of vowel-nasal combinations and to make judgments on the place of articulation. This experiment seeks to correlate nasal production with nasal perception and how SC subjects make use of the acoustic attributes and enhancing gestures for nasalization to identify the place of articulation of nasal codas.

The final chapter gives a summary and discussion of the findings.

Chapter 2. Acoustic model of nasalization

2.1 Acoustic model of nasalization

Nasal consonants are produced by forming a complete closure at some point along the oral region of the vocal tract and during the time there is a supraglottal closure, lowering the velum to allow air to flow through the opened velopharyngeal port to the nose, so that there is no pressure increase behind the constriction (Stevens, 1998).

Fujimura (1962) modeled the articulatory system for nasal production as three subsystems that are acoustically coupled to each other at their ends at the velum: the pharynx extending from the glottis to the velum, the oral cavity with a complete closure at the anterior end, and the nasal tract. In Fujimura's model, the three subsystems are approximated by three acoustic tubes that transmit plane waves and have varying cross-section areas.

A schematization of a nasalized vowel is shown in Figure 2.1 with two outputs, the volume velocity at the mouth, $U_{m,}$ from the main tube and the volume velocity at the nostrils, $U_{n,}$ from the side branch. In the case of a nasal consonant, $U_m = 0$. U_s is the source volume velocity at the glottis. The spectrum of the volume velocity at the nose is the product of the volume velocity at the glottis and the transfer function from the glottis to the nose.

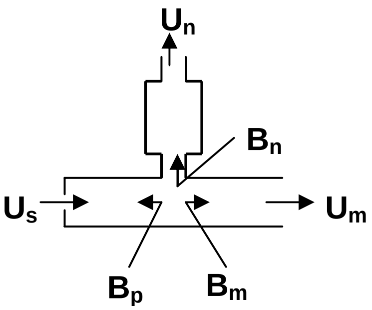

Figure 2.1 Schematization of a simple model for production of nasalized vowels. The degree of nasalization can be varied by changing the area of the velpharyngeal port.

The susceptances looking from the coupling point of the three tubes into the pharynx, the mouth, and the nose are represented by B_p, B_m, and B_n respectively. The natural frequencies of the system, or the poles, are given by the frequencies where the sum of the susceptances looking in all possible directions at any arbitrary point in the system is zero: $B_n + B_m + B_p = 0$. In other words, the impedance looking into all points from the coupling point is ∞. The zeros of the output at the nose occur at frequencies for which $B_m = \infty$. At these frequencies the mouth cavity short-circuits transmission to the nose. That is, the impedance looking into the mouth, $Z_m = 0$. The zeros of the output at the mouth occur at frequencies for which $B_n = \infty$. At these frequencies the nasal cavity short-circuits transmission to the mouth. In other words, the impedance looking into the nose, $Z_n = 0$. The zeros of the nasalized vowel system are the zeros of $\frac{U_m + U_n}{U_s}$. We are interested in the acoustic correlates of nasalization in the vowel region, and in particular the behavior of the extra pole indicative of a nasal, so although the zero of the output is of some interest, we will concentrate on the poles of the nasalized vowel system.

2.2 Measurements and data from early models of nasalization

Early models of the acoustics of nasalization (House & Stevens, 1956, Fujimura, 1962), combined with sweep-tone measurements of vocal-tract characteristics (Fujimura & Lindqvist, 1970, Lindqvist & Sundberg, 1976) carried out with stationary articulations of nasal consonants and nasalized vowels have characterized the transfer function of nasalization. Figure 2.2 shows the susceptance curves of B_n and $-(B_p+B_m)$ for a nasalized vowel with a uniform cross-sectional area of 3 cm^2. The frequencies at which the two susceptances cross indicate the poles of the system. For a non-nasal vowel with uniform cross-sectional area of 3 cm^2, the natural frequencies can be found by setting the impedance looking into the mouth to infinity. The first three formants of this quarter-wavelength tube, assuming an oral tract of 17 cm are approximately 520 Hz, 1560 Hz, and 2600 Hz.

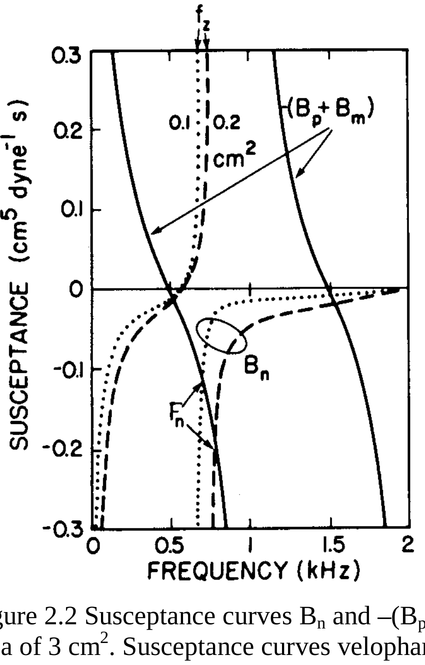

Figure 2.2 Susceptance curves B_n and $-(B_p+B_m)$ for a nasalized vowel with a uniform cross-sectional area of 3 cm^2. Susceptance curves velopharyngeal openings of 0.1 cm^2 and 0.2 cm^2 are shown (from Stevens, 1998).

The opening of the velopharyngeal port introduces an extra pole and zero that can be seen below 2 kHz. The poles are also slightly shifted, as shown in Figure 2.3. The dotted line in the figure indicates the transfer function for the non-nasal vowel. The solid line indicates the transfer function for the nasalized vowel, with a pole-zero pair just before 1 kHz.

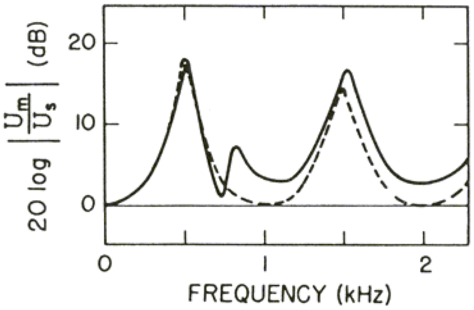

Figure 2.3 Calculated transfer function U_m/U_s for a nonnasal vowel produced with uniform cross-sectional area of the vocal tract (solid line) and a nasalized vowel produced with a velopharyngeal port opening of 0.2 cm^2 (from Stevens, 1998).

The transfer function of a nasal consonant is shown in figure 2.4. There is a prominent pole at about 250 Hz. The zero at around 1200 Hz is the frequency at which the impedance looking into the mouth from the coupling point is zero. It is indicative of a labial nasal, with a constriction about 8 cm from the coupling point.

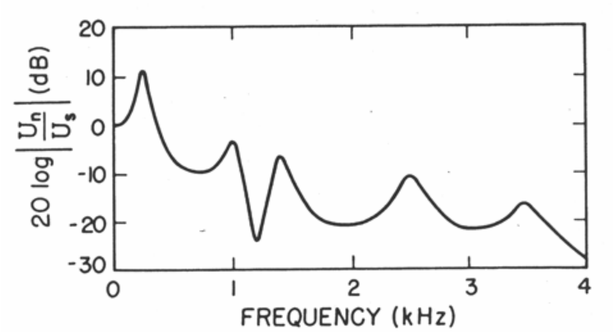

Figure 2.4 Transfer function of U_n/U_s for a nasal consonant. The zero at 1200 Hz is the frequency at which the impedance looking into the right at the coupling point is zero (from Stevens, 1998).

For an alveolar nasal consonant, the constriction is made about 4 cm from the point of coupling; for a velar nasal consonant, the constriction is made close to the point of coupling. As we shall see in Section 2.4, the characteristic prominences in the acoustics of the nasal murmur reflect whether the constriction is made at the mouth, near the hard palate, or the soft palate. This implies that if the primary gestures that are associated with the distinctive features of each nasal are properly made, then the identity of the nasal consonant can be determined from the acoustics of the nasal murmur. The primary gestures refer to the opening of the velopharyngeal port and the complete closure somewhere along the oral tract. The place of constriction in the oral tract distinguishes the three nasals from one another.

However, these primary gestures may be made incompletely or not at all, so that the acoustics of the nasal murmur may not contain enough of the characteristic prominences that identify the nasal consonant. We have pointed out in chapter one that the nasal codas in SC are the only consonants allowed in the coda position and have no other consonants to compete with for perceptual contrast. Wang's (1993) airflow measurements also suggested that the SC nasal codas are not made with complete oral constrictions and that they are nasal approximants. Under these circumstances particular to SC, the acoustics of the nasal murmur may not provide enough information to identity of the nasal coda and some kind of enhancing gestures may be expected to further distinguish the nasal coda. These enhancing gestures may very well occur during the production of the vowel preceding the nasal coda.

Before we explore the interaction of the primary gestures and the enhancing attributes involved in the production of the nasal coda, we will first establish models of nasal consonant production and of vowel nasalization.

2.3 Vowel Nasalization

For the rest of the discussion of vowel nasalization, we are interested in the behavior of the extra pole in the nasalized vowel, as this is an important acoustic correlate of vowel nasalization. In the following section, we present a simplified model of vowel nasalization, where the nasal tract is represented by a very short and thin velopharyngeal port connected to a longer and wider nasal cavity that has a narrowing at the nostrils. We will estimate the poles of the system based on this simple model and observe the behavior of the poles as we vary the area of the velopharyngeal port.

2.3.1. Pole calculations

The poles of the system are given by the frequencies where the sum of the susceptances looking in all possible directions at any arbitrary point in the system is zero: $B_n + B_m + B_p = 0$. We derive expressions for the susceptances from the impedances, which are more familiar concepts.

We model the pharyngeal and mouth cavities as a tube of uniform cross sectional area, as shown in Figure 2.5.

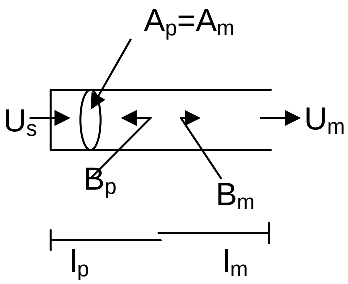

Figure 2.5 Schematization of the pharyngeal and mouth cavity with a uniform cross-sectional area.

The impedance, Z, the admittance, Y, the conductance, G, and the susceptance, B, are related as follows.

$$\mathrm{Z} = \mathrm{R} + \mathrm{jX}$$
$$\mathrm{Y} = \mathrm{G} + \mathrm{jB}$$
$$\mathrm{Z} = \frac{1}{\mathrm{Y}} = \frac{G - jB}{G^2 + B^2} = \frac{G}{G^2 + B^2} + \frac{-jB}{G^2 + B^2} = \frac{-j}{B}$$

The impedance looking back into the pharynx can be approximated by the impedance of a quarter-wavelength tube and the impedance looking into the mouth at the junction point can be approximated by the impedance of a half-wavelength tube. The susceptances can then be derived from the impedances as follows:

(1) $\mathrm{B_p} = \frac{\mathrm{A_p}}{\rho \mathrm{c}} \tan(\mathrm{kl_p})$,

(2) $\mathrm{B_m} = -\frac{\mathrm{A_m}}{\rho \mathrm{c}} \cot(\mathrm{kl_m})$, where $k = \frac{2\pi f}{c}$.

The impedance looking into the nasal cavity from the junction point can be approximately as the sum of the impedances of the two tubes, assuming that the length of the velopharyngeal port, represented as l_{vpp} in Figure 2.6, is much less than the length of the nasal tract, represented as l_{nasal}.

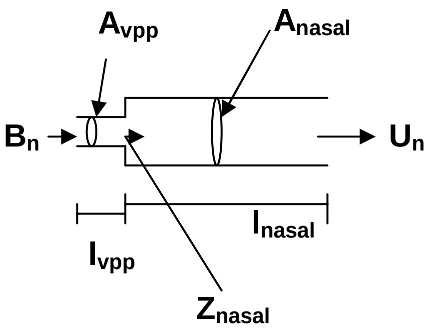

Figure 2.6 Schematization of the nasal tract, consisting of a short tube representing the velopharyngeal port and a longer tube representing the nasal cavity.

The impedance looking from coupling point of the two tubes into the nasal tract can be approximated by the impedance of a quarter-wavelength tube,

$$Z_{nasal} = \frac{j\rho c}{A_{nasal}} \tan(kl_{nasal}) ,$$

where A_{nasal} and l_{nasal} are the dimensions of the nasal tract.

The impedance looking from the opening of the velopharyngeal port into the port itself can be approximately by a half-wavelength tube.

$$Z_{vpp} = j \frac{\rho c}{A_{vpp}} \tan kl_{vpp}$$

If we make the assumption that the dimensions of the short tube representing the velopharyngeal port are much smaller than the dimensions of the larger tube representing the nasal cavity,

$$\frac{2\pi f l_{vpp}}{c} << 2\pi ,$$

Then we can simplify the expression for the impedance looking into the velopharyngeal port as an acoustic mass, where l_{vpp} and A_{vpp} are respectively the length and cross-sectional area of the velopharyngeal opening.

$$Z_{vpp} = j\frac{\rho c}{A_{vpp}}\tan kl_{vpp} \approx j2\pi f\frac{\rho l_{vpp}}{A_{vpp}}$$

The total impedance looking through the velopharyngeal port and through the nasal cavity, which we call Z_{vpp_nasal} here, can be approximated as a sum of Z_{vpp} and Z_{nasal}

(3) $$Z_{vpp\_nasal} = \frac{j\rho c}{A_{nasal}}\tan(kl_{nasal}) + j2\pi f\frac{\rho l_{vpp}}{A_{vpp}}$$

The poles of this system occur at frequencies where $B_n + B_m + B_p = 0$, or where

(4) $$\frac{-j}{Z_{vpp\_nasal}} + \frac{A_p}{\rho c}\tan kl_p + -\frac{A_m}{\rho c}\cot kl_m = 0$$

We solve this equation graphically, using MATLAB. The susceptance curves are shown for the nasalized vowel in Figure 2.7 for two values of velopharyngeal openings, 0.1 cm^2 and 0.2 cm^2. The curves are generated by making the approximation that for this simple vowel vocal tract configuration, the length of the mouth cavity and the pharynx are both 8 cm and that their cross-sectional areas are 3 cm^2. Furthermore, the cross sectional area of the nasal tract is approximated as 2.4 cm^2, based on average physical dimensions of the vocal tract (Chen, 1997). The length of the nasal tract is 10 cm and the length of the velopharyngeal port is 2 cm. The intersections of the curves B_n and $-(B_m+B_p)$ are the poles of the system.

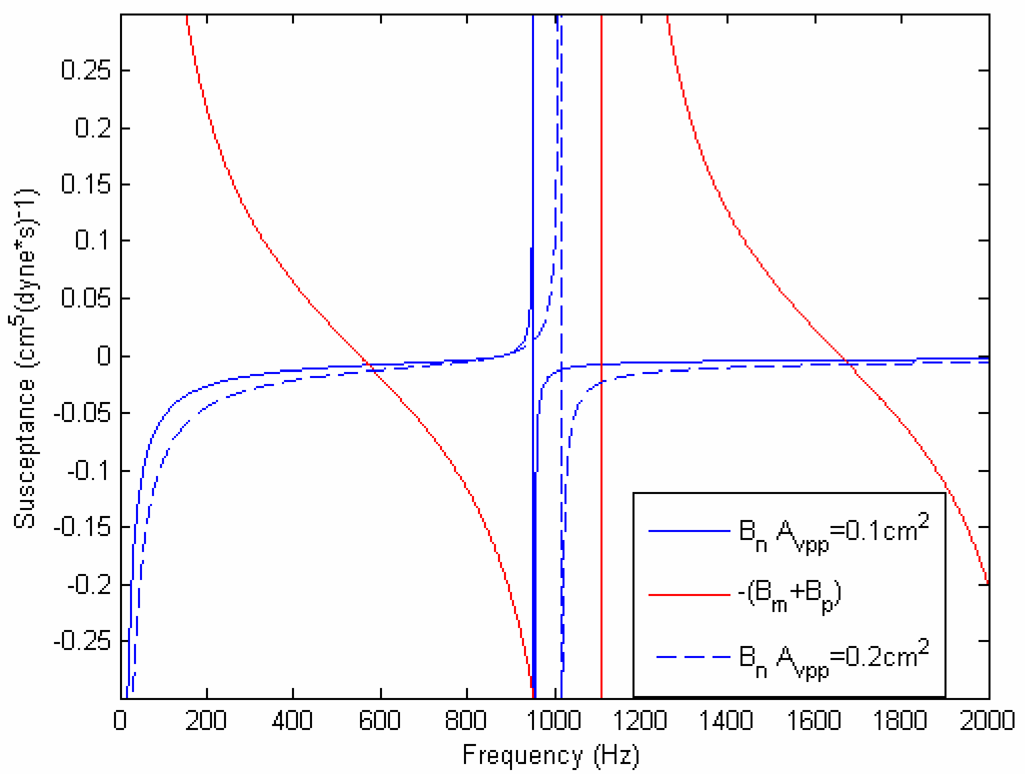

Figure 2.7 Susceptance curves for a simplified model of vowel nasalization. The curves are shown for two cross-sectional areas of the velopharyngeal port.

Nasalization introduces an extra pole-zero pair to the spectrum of the vowel. Figure 2.7 indicates that the extra pole occurs at about 950 Hz when the velopharyngeal opening is 0.1 cm^2 and that it increases to about 1000 Hz when the velopoharyngeal opening is increased to 0.2 cm^2. Note that the straight line in red in the figure is simply the asymptote of the susceptance curve for $-(B_m + B_p)$, which MATLAB displays, and should not be considered as a curve for the intersection with B_n.

In equation (3), the impedance looking into the nasal tract through the coupling point is described by the sum of two terms. The first term represents the impedance looking into the nasal cavity from the other end of the velopharyngeal port and the second term represents the impedance looking into the velopharyngeal port through the coupling point. Only the second term, which is an acoustic mass of the short tube representing the velopharyngeal port, varies with the cross-sectional area of the velopharyngeal port. The first term stays constant.

Equation (4) indicates an inverse relationship between the cross-sectional area of the velopharyngeal port and the overall impedance looking into the nasal tract. As the velopharyngeal port is opened further, the impedance decreases. The susceptance is inversely related to the impedance, so the susceptance will increase. The susceptance curve representing B_n would therefore shift to the right.

However, there is a limit to this increase in B_n. Even if we increase the cross-sectional area of the velopharyngeal port hypothetically to match that of the nasal cavity, at 2.4 cm^2, which is physiologically impossible, the nasal pole only increases to about 1350 Hz, given the dimensions we have used so far to approximate the nasalized vowel, as shown in Figure 2.8.

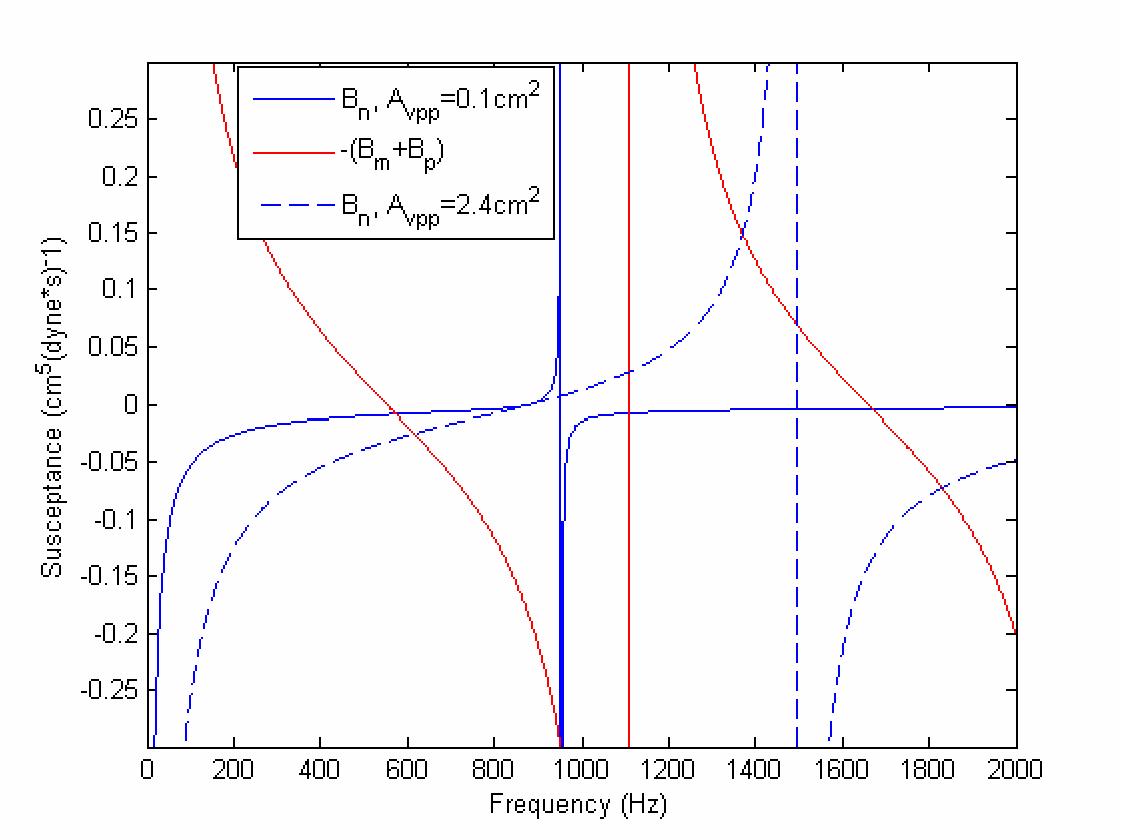

Figure 2.8 Susceptance curves for a simplified model of vowel nasalization. The curves are shown for two cross-sectional areas of the velopharyngeal port, one that is physiologically possible, and the other that is hypothetical.

This is a highly simplified model of the vowel nasalization. The nostrils, which are not part of this model, play an important role in shaping the output at the nose. The nostrils have a smaller cross-sectional area than the rest of the nasal cavity leading up to them, and based on perturbation theory (Stevens, 1998), this constriction will in effect cause a downward shift in all formants. We would therefore expect to observe a nasal pole that is no higher than 1350 Hz in the spectra of nasalized vowel utterances.

2.4 Nasal consonants

2.4.1. Poles

The poles of a nasal consonant can be found by summing up the susceptances looking into the three cavities so that $B_m + B_p + B_n = 0$. As shown in Figure 2.9, in an idealized production of a nasal consonant, the output at the mouth U_m, is zero.

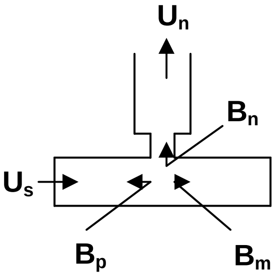

Figure 2.9 Schematization of a simple model for production of nasal consonant. The mouth output is zero.

The expressions for the susceptances looking into the pharynx and the nasal cavity that we use to approximate the poles of a nasal consonant are the same as those used to solve the poles for a nasalized vowel. The only change is the expression for the susceptance looking into the closed mouth cavity. We have

$$(5)\quad B_n = \frac{-j}{Z_{vpp\_nasal}}, B_p = \frac{A_p}{\rho c}\tan kl_p, B_m = \frac{A_m}{\rho c}\tan kl_m$$

Figure 2.10 shows the susceptance curves generated by MATLAB with the above expressions for a labial nasal with $l_m = 8$ cm. The cross-sectional area of the velopharyngeal port used is 0.2 cm^2. The poles occur at the frequencies where the susceptance curve of B_n crosses the susceptance curve of $-(B_n+B_p)$. The vertical line in red at about 1100 Hz in the graph is simply the asymptote of $-(B_n+B_p)$, which

MATLAB displays, and should not be considered when finding the cross-over points with B_m.

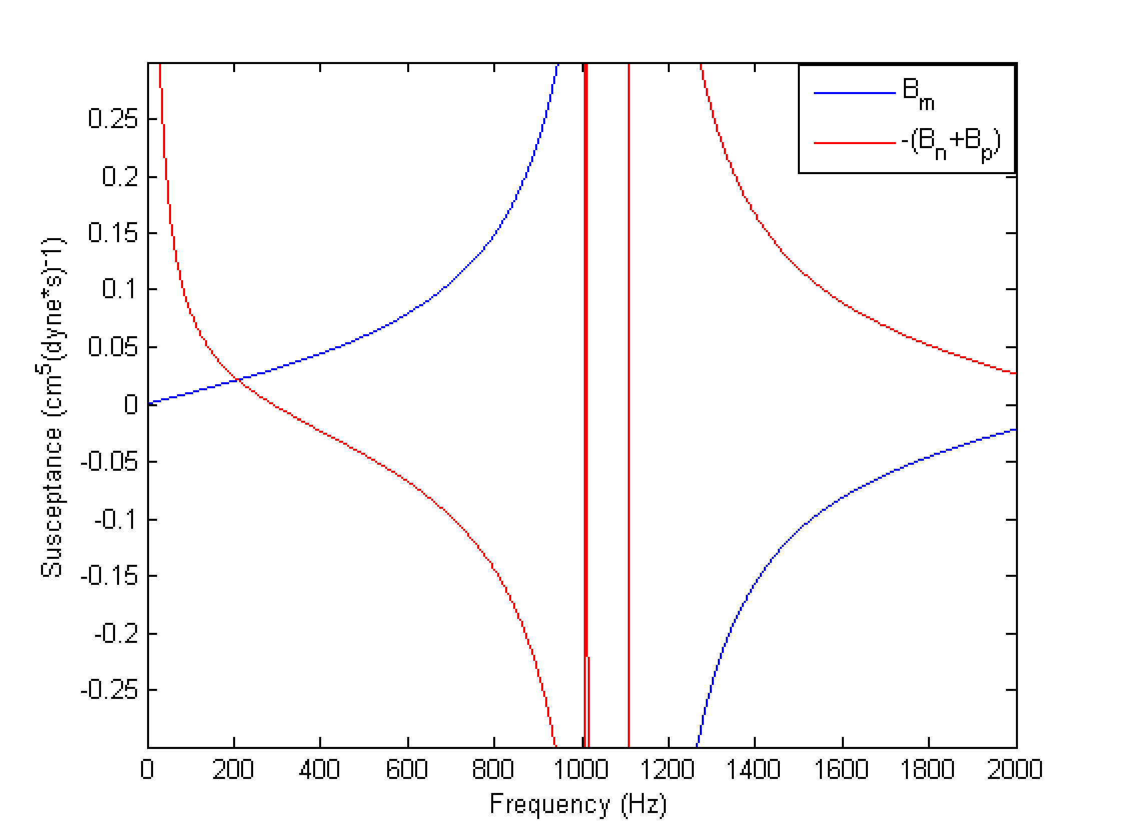

Figure 2.10 Susceptance curves for a simplified model of a labial nasal consonant, with a velopharyngeal opening of 0.1 cm^2.

From equation (5), the susceptance curve $-(B_n+B_p)$ is the sum of three terms: a cotangent function that represents the susceptance looking into the nasal tract, a term proportional to the cross-sectional area of the velopharyngeal port that represents the acoustic mass of this short tube, and a tangent function that represents the susceptance looking into the pharynx. This curve is shown in red in Figure 2.10. The curve shown in blue represents the susceptance looking into the mouth cavity. The first two poles occur at approximately 200 Hz and 1000 Hz. The third pole occurs around 2200 Hz (not pictured). The first zero occurs when B_m goes to infinity, at about 1100 Hz. These estimated values from an idealized and simplified model of the labial nasal agree with the measurements described in section 2.2 and given in Stevens' discussion of sonorant consonants (1998).

Equation (5) indicates that the susceptance curve $-(B_n+B_p)$ is the sum of three terms, only one of which is proportional to the cross-sectional area of the velopharyngeal port. As the velopharyngeal port opens further, we would expect the curve to shift to the right and see an increase in the second nasal pole. There is again a limit to how large that nasal pole in the murmur can be, just as there was a limit to the nasal pole in the nasalized vowel. The effect of the increase in the cross-sectional area of the velopharyngeal port on the second nasal pole in the nasal murmur, however, is not expected to be as great as it is in the nasalized vowel, since the area of the velopharyngeal port effects only one of three terms in the calculation of the nasal pole in the murmur, where as it effects one of two terms in the calculation of the nasal pole in the nasalized vowel.

The susceptance curves generated by MATLAB for an alveolar nasal consonant with a constriction made about 5 cm from the coupling point, with all other parameters kept equal, are shown in Figure 2.11. The first two poles occur at approximately 200 Hz and 1000 Hz. The third pole occurs at approximately 1400 Hz. The first zero occurs when B_m goes to infinity, at about 1750 Hz.

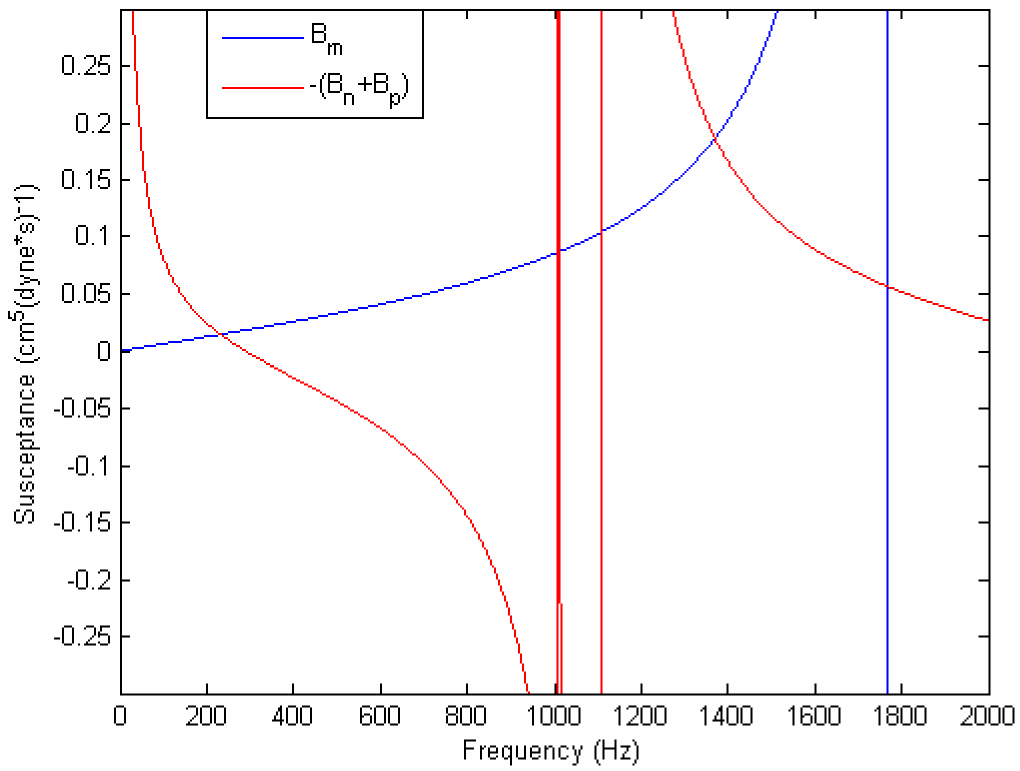

Figure 2.11 Susceptance curves for a simplified model of an alveolar nasal consonant.

The susceptance curves generated by MATLAB for velar nasal consonant with a constriction made at about the coupling point are shown in Figure 2.12. The first two poles occur at approximately 200 Hz and 1000 Hz. The third pole occurs at approximately 3000 Hz. There are no zeros in this configuration.

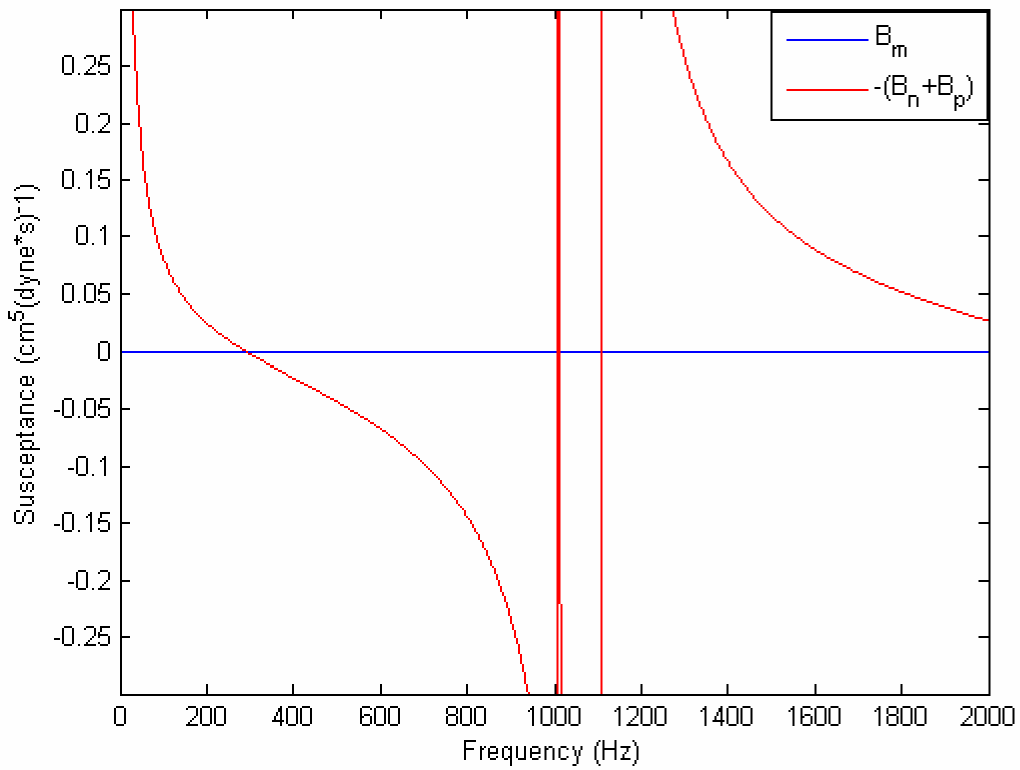

Figure 2.12 Susceptance curves for a simplified model of a velar nasal consonant.

2.4.2 Zeros

Another way of finding the zeros, beside the graphical method we have just described in the previous section, is to locate the frequencies where the impedance looking into the mouth is $Z_m = 0$, as given by equation (6)

$$(6)\ Z_m = \frac{\rho c}{A_m} \cot k l_m = 0,$$

Where l_m is the distance from the oral constriction to the coupling point and A_m is the cross sectional area of the mouth cavity. l_m is approximately 8cm for a labial nasal consonant, 4cm for an alveolar nasal consonant, and 0cm for a velar nasal consonant, since the constriction is very close to the coupling point.

Equation (6) can be solved by setting $\cos(\mathrm{kl}_{\mathrm{m}}) = 0$ to satisfy the boundary condition, where $k = \frac{2\pi f}{c}$, and $l = \frac{c}{f}$, $\mathrm{kl}_{\mathrm{m}} = \frac{\pi}{2}(2n-1)$, and we have equation (7).

(7) $f_z = \frac{c}{4\mathrm{l}_{\mathrm{m}}}(2n-1)$

The first zero for a labial nasal consonant, assuming a constriction made approximately 8cm from the point of coupling, is 1100 Hz. The first zero for an alveolar nasal consonant, assuming a constriction made approximately 4cm from the point of coupling, is 1770 Hz. For a velar nasal consonant, there are no zeros, and it is an all pole system.

2.5 Quantal nature of nasality

We have discussed in this chapter the methods for estimating the location of the poles and zeros in a nasalized vowel and nasal consonant. The presence or absence of the extra poles and zeros, which is thought to correlate with the opening and closing of the velopharyngeal port, has been considered the acoustic correlate of the feature [+/- nasal]. When we speak of the quantal nature of nasality, we are concerned with an articulatory event that leads to the perception of nasality. We are essentially trying to establish the three regions that capture the quantal nature of speech in Figure 1.1 of section 1.2, where moderate changes in some arbitrary articulatory parameter does not significantly alter the relevant acoustic attribute in regions I and III and where small changes in the articulatory parameter leads to large changes in the acoustic parameter in region II. In terms of the feature [nasal], we are interested in whether there exists a range of values within which the velopharyngeal opening takes that leads to the most prominent perception of nasality and whether beyond this range, changes in the velopharyngeal opening would not lead to much change in the perception of nasality.

To do this, we examine what happens to the primary acoustic correlate of the feature [nasal] when the primary articulator, the velopharyngeal port, is increased in cross-sectional area continuously from a closed state. Earlier modeling works of nasalization (House & Stevens, 1956, Fujimura, 1962) have shown that as the area of the velpharyngeal port increases, the extra pole increases as well. By varying the cross-sectoinal area of the velopharyngeal port, we expect to observe abrupt changes in the acoustics that indicate the nasal-nonnasal distinction. We hypothesize that as the area of the velopharyngeal port is increased from zero (closed state), there exists a range within which the primary acoustic correlate of nasality will be the most prominent.

The relationship between the cross-sectional area of the velopharyngeal port and the first two formants of the nasalized vowel, as well as the nasal pole around 1 kHz, are shown in Figure 2.13. The formant values and the location of the nasal pole are tracked using an algorithm that picks the intersections of B_n and $-(B_m+B_p)$ from each figure, such as Figure 2.12, which was generated by MATLAB for each cross-sectional are of the velopharyngeal opening.

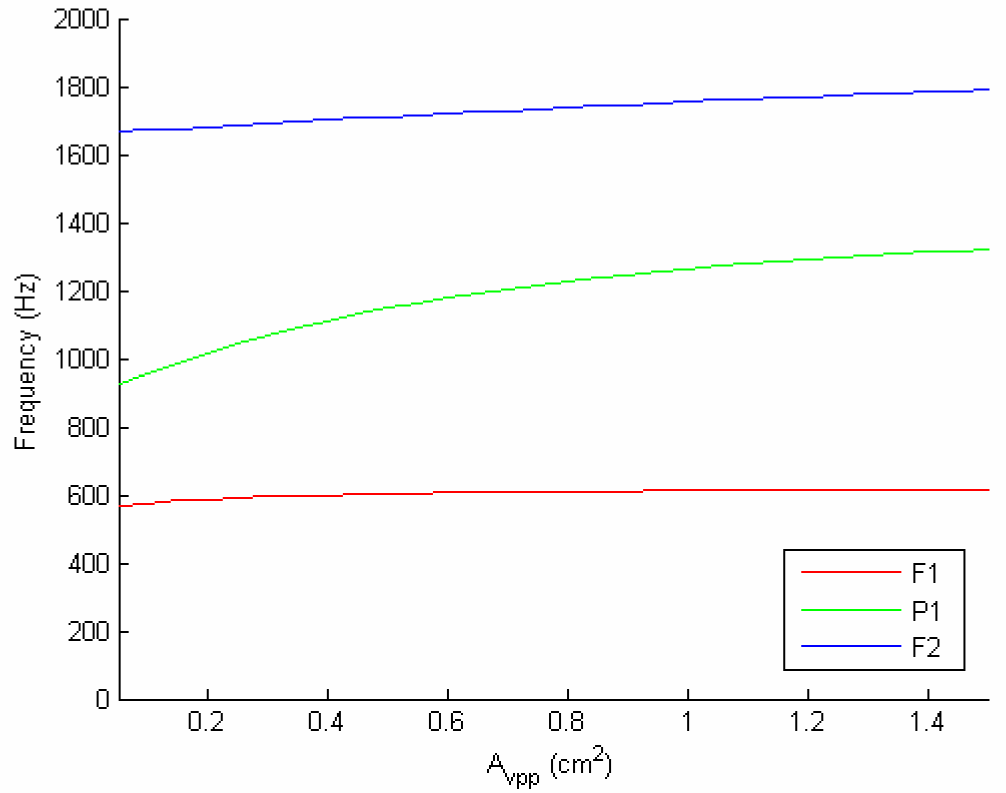

Figure 2.13 Relationship between the A_{vpp} and the first formant, F1, the second nasal pole, P1, and the second formant, F2.

The x-axis in Figure 2.13 has been extended to 1.4 cm^2, so that we can more easily observe a general trend in the behavior of the nasal pole as the velopharyngeal port cross-sectional area is increased, although aerodynamics studies (Warren, 1997, 1993) have shown that the average opening of the velopharyngeal port during the production of nasals is usually between 0.1 cm^2 and 1cm^2. The first and second formants do not increase much from the values for a non-nasal neutral vowel, which we have shown in section 2.2 to be approximately 520 Hz and 1560 Hz. The nasal pole, however, increases from about 900 Hz to about 1250 Hz for this configuration. The slope of the curve levels off when the velopharyngeal opening size increases beyond 1cm^2. The slope of the curve is largest from the moment the velopharyngeal port opens to when it is opened to about 0.4 cm^2.

We have discussed that as the cross-sectional area of the velopharyngeal opening increases, the poles also shift to higher frequencies. The zeros of the system with the dimensions specified above for a relatively open vowel, however, depend mainly on the length from the coupling point to the oral closure, and are not affected very much by the change in the size of the velpharyngeal opening. Thus, the movement of the poles away from the zeros gives rise to the prominence of the nasal pole that is observable in the acoustics. The movement of the poles in Figure 2.13 simply gives us a range within which to look for the presence of a nasal pole in the acoustic signal. It does not give the amplitude of this pole, nor does it say anything about the difference between the amplitude of this pole and the corresponding zero. It is the difference between the amplitude of the nasal pole-zero pair that gives the nasal pole its prominence in the acoustics. This amplitude difference correlates nonlinearly with the degree of the velopharyngeal port opening and gives rise to the quantal relation between the articulator movement and the acoustics that we have been looking for.

The amplitude difference between the nasal pole-zero pair is found as follows. The nasal poles are picked from Figure 2.13. We assume that the zero does not move and that it takes on the lowest value of the poles. This marks the starting point of the pole-zero separation. The bandwidths for the pole and zero are set to a default 100 Hz (Klatt, 1980). For each value of the nasal poles, a spectrum was generated (Fant, 1960), and the spectrum of the zero at a fixed frequency is added to it, in order to simulate the nasal resonance, as shown in Figure 2.14. The minimum value of the summed spectrum was subtracted from the maximum value to obtain the difference amplitude of the nasal pole-zero pair.

Figure 2.14 shows the pole-zero spectra generated with three different values of nasal poles. The zeros are all located at the lowest value of the pole and the poles are values picked from figure 2.13.

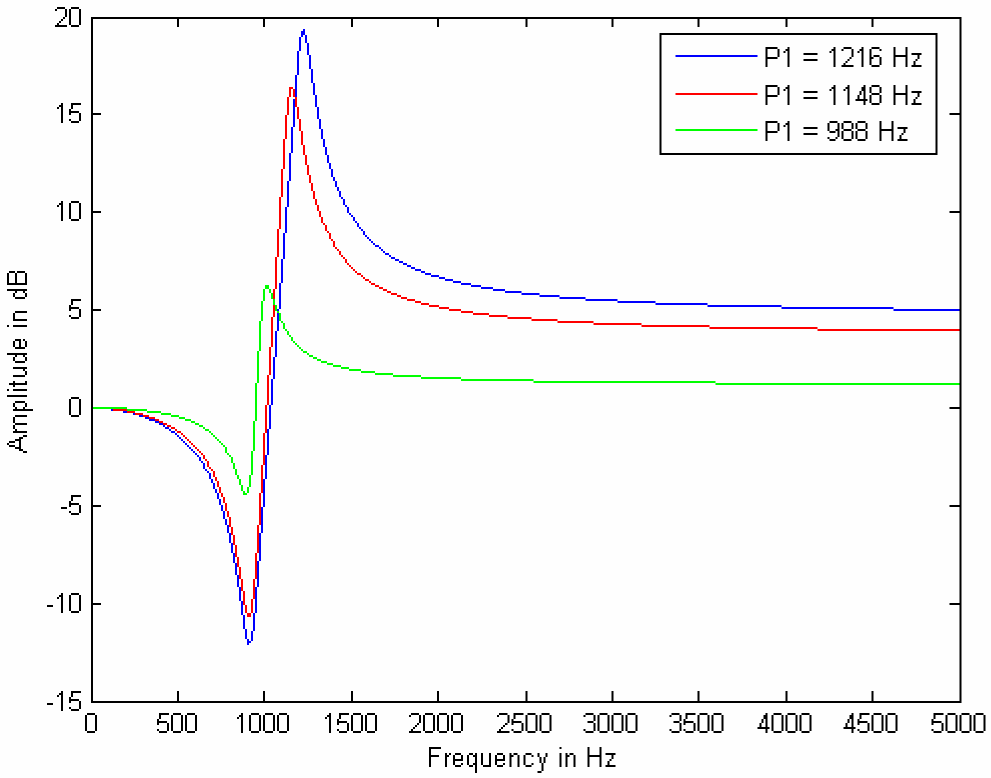

Figure 2.14 Spectra of the pole-zero pairs for different pole locations corresponding to different velopharyngeal opening sizes.

The pole located at 1216 Hz is about 20 dB in amplitude and its corresponding zero is about 12.5 dB in amplitude. The difference in amplitude between the pole and zero is

32.5 dB. The pole located at 988 Hz is about 6 dB in amplitude and the zeros is about 4 dB. The difference in amplitude between this pole and zero pair is about 10 dB.

The curve describing the differences in amplitude of the nasal pole and zero as the velopharyngeal port is opened and increased to 1.5 cm^2 is shown in Figure 2.15. The amplitude of P1-Z1, for a velopharyngeal opening of about 0.125cm^2 is about 5 dB. The amplitude of P1-Z1, for a velopharyngeal opening of about 0.25 cm^2 is about 20 dB. The amplitude of P1-Z1, for a velopharyngeal opening of about 0.5 cm^2, is about 27.5 dB. The amplitude of P1-Z1, for a velopharyngeal opening of about 1 cm^2 is about 32 dB.

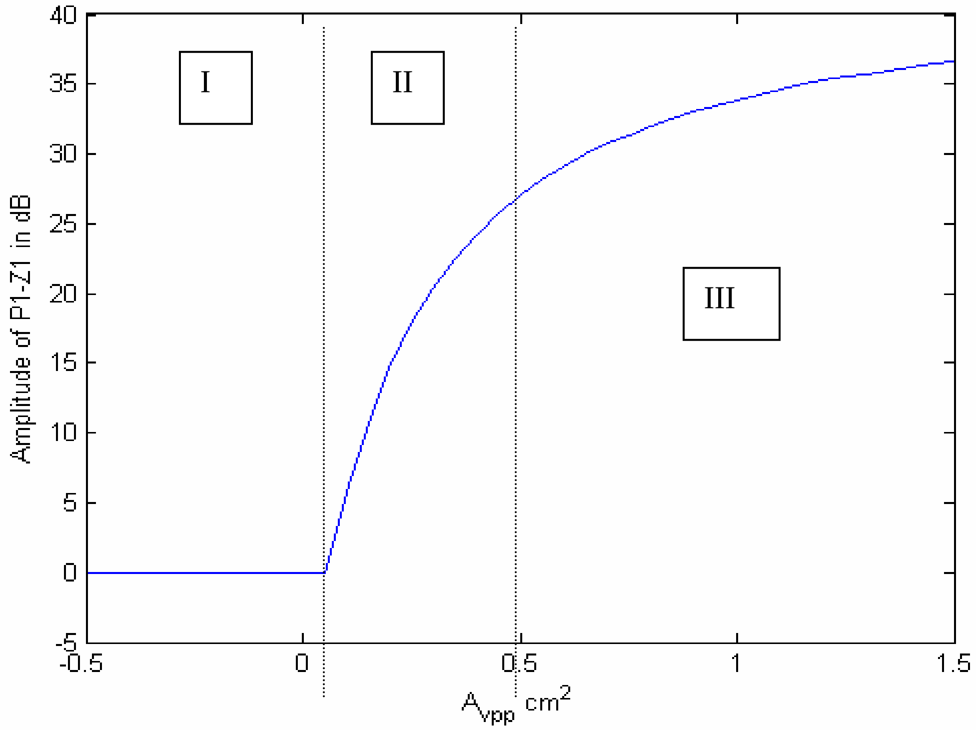

Figure 2.15 Amplitude of the maximum difference between the amplitudes of the pole and zero pairs for each opening of the velopharyngeal port.

The curve of P1-Z1 plotted against the cross-sectional area of the velopharngeal port shows a non-linear relationship. Within region I, the velpharyngeal port is essentially closed and the pole-zero ride on top of each other and there is no pole prominence in the acoustics that would indicate the presence of a nasal sound. Within region II, the amplitude of P1-Z1 is most affected by small changes in the size of the velopharyngeal opening. In region III, just as in region I, moderate changes in the size of the velopharyngeal opening do not lead to large changes in the amplitude of P1-Z1.
The opening of the velopharyngeal port above 0.5 cm^2 does not produce the steep effect on the acoustics as it did under 0.5 cm^2. This seems to indicate that the quantal effect of nasalization is most readily observed when the velopharyngeal port is opened to about $0.5 cm^2$.

2.6 Summary

In this chapter, we have provided an account of the earlier models of nasalization and data from sweep-tone measurements used to construct graphs for finding the poles and zeros of a nasalized vowel and nasal consonant. We have also provided a simple method for estimating the poles and zeros by using simple tubes to model the nasal and oral cavities and solving the susceptance equations.

The presence of the nasal pole at about 1 kHz in a vowel is one acoustic correlate of nasality. The precise location of a nasal pole would depend at least in part on the opening of the velopharyngeal port. Figure 2.13 provides a range within which to search for this acoustic correlate of nasalization, given that acoustics indicates the presence of a nasal.

We have related increases in the velopharyngeal opening, which is considered the articulator responsible for the perception of nasality, to the acoustics, in terms of the amplitude difference between the nasal pole-zero pair. We have found that the maximum amplitude difference between the nasal pole-zero pair corresponds to a range in cross-sectional area of the velopharyngeal port of up to 0.5 cm^2. That is, the opening of the velopharyngeal port leads to the greatest change in the acoustics only up to about 0.5 cm^2. Further opening does not contribute more to the perception of nasality. Outside of this range, moderate changes in the size of the velopharyngeal opening do not affect very much the acoustics. This nonlinear relationship between the articulator and the acoustics is precisely what captures the underlying quantal nature of speech perception.

This page is intentionally left blank.

Chapter 3. Distribution of acoustic correlates of nasalization

In this chapter, we explore how language-specific phonotactics might impose constraints on the enhancing gestures that are selected for a universal feature by examining for possible differences in the distribution of the acoustic correlates for the place of nasal articulation in Chinese vs. English nasal codas. The nasal codas in English and SC may have similar acoustic attributes but the attributes may differ in their distribution due to differences in the phonotactic constraints. We are interested in how much information, in terms of the acoustic cues related to the nasal place of articulation, is contained in the vowel transition region versus the consonant murmur region in English vs. SC.

First, let us review some differences in the phonotatic constraints in SC and English. We have discussed in Chapter 1 that there is only one underlying low vowel, [a], in Standard Chinese. This low vowel is unspecified for the feature [back]. We hypothesize that the articulation of this low vowel can therefore be modified by the nasal coda that follows it. For example, the low vowel can be produced with a more backed tongue body before a velar [ŋ], where the constriction in the oral tract for [ŋ] is made near the soft palate. We will represent this more backed low vowel as [ɑ]. The low vowel is produced with a more fronted tongue body before an alveolar [n], where the constriction in the oral tract is made near the hard palate. We will represent this more fronted low vowel as [a]. Thus, when the low vowel is followed by a nasal coda in SC, we have two possible combinations, either [an] or [ɑŋ]. In English, on the other hand, there are two underlying low vowels: a fronted [æ], as in the word “dad”, as well as a backed [ɑ], as in the word “pop”. These two vowels are contrastive in English where as the more backed and fronted versions of the low vowel in SC are not contrastive. In English, however, when the low vowels are followed by the nasal coda, we have four possible combinations, [æn], [ɑn], [æŋ], and [ɑŋ] .

Because the phonotactic constraints differ in the two languages, SC and English speakers will encode the contrast between the coda places differently, and SC and English listeners will interpret the cues in the two languages in different ways. In this perception experiment, we ask native speakers of Standard Chinese to listen to English consonant-vowel-nasal (CVN) stimuli and pick from two corresponding SC CVN stimuli the one that sounds the closest to the English stimuli. This experiment is prompted by a loanword study where words from English are mapped to SC in real-time (Hsieh, Kenstowicz, & Mou, 2005). Real-time online loan-word adaptation experiments (Kenstowicz & Suchato, 2004, Broselow, 2000) are thought to potentially to reveal constraints in a native phonology, in this case SC, with foreign input, in this case English.

The vowel that is the focus of this experiment is the low vowel, although included in the discussion as a comparison will be the high vowel /i/, which is specified for the feature [back] and therefore has presumably less freedom in terms of the forward and backward movements of the tongue body. Table 3.1 shows the possible mappings from English to SC. For every English sound, there are two possible mappings to SC.

Table 3.1 English to SC VN choices

English stimuli	SC choices	
	an	ɑŋ
æn		
ɑn		
æŋ		
ɑŋ		

We are interested in the distribution of the acoustic cues for the place of articulation of the nasal coda. If the perceptual choice is based on vowel similarity, then more acoustic information about the place of articulation of the nasal coda must be contained in the vowel region. In this case, we would expect the listeners to hear the English /æŋ/ and the SC /an/ as the same place of articulation for the nasal coda. Conversely, if the perceptual choice is based on nasal consonant similarity, then more acoustic information about the nasal place of articulation must be contained in the murmur region. In this case, we would expect the listeners to hear the English /æŋ/ and the SC /ɑŋ/ as the same place of articulation for the nasal coda. Either conclusion must be further checked against the acoustic evidence contained in the recorded speech. Through acoustic analysis of the recorded speech stimuli, we can then quantify the acoustic cues that are contained in the vowel transition region vs. in the nasal murmur region.

3.1 Subjects and stimuli

A male native English speaker, who is a student of phonology at MIT, was recorded reading a list of 36 non-sense English words that are legal syllables, and 36 fillers for control purposes, each embedded in the sentence “Please say ___ for me.” Recordings are made using the MARSHA digital recording software tool. A native speaker of Standard Chinese, who is currently a professor in linguistics, was recorded reading a counterpart list of 36 SC target words and 36 fillers embedded in the SC sentence “wo^{11} s̢^{w}o^{55} ___ ts̢ɤ^{51} kɤ^{11} tsɨ^{51} ‘I say ___ this word’.” The English and Chinese recordings were converted into two Microsoft wave files and the individual words were extracted using Praat, a speech analysis and synthesis computer program.

The extracted words were then organized into sequences of AXB, where A, B were Chinese syllables and X an English syllable. The stimuli for the experiment were constructed with MATLAB, a high-level computing language. Each of the AXB utterances was repeated so that for every AXB sequence, there was a BXA sequence, for a total of 144 target utterances. The stimuli were randomized by means of a randomizing algorithm in MATLAB. A practice session preceded the actual experiment and consisted of 5 AXB stimuli. The practice session was meant to familiarize the subjects with the stimuli and task. There was no feedback given.

Fourteen subjects were recruited to participate in the listening experiment. The listeners, who are all students or affiliates at MIT, were paid a nominal fee for the 30-minute task. They are all speakers of Standard Mandarin.

3.2 Description of analysis tool

This section briefly describes XKL, an analysis tool for extracting acoustic cues from the speech waveforms. XKL is an X-windows port of the interactive speech analysis package originally developed by Klatt (1980). The XKL program makes a spectral representation of the waveform by computing a 512-point discrete Fourier transform on a length of waveform that is first differenced, and multiplied by a Hamming window. In this thesis, a long window length of 25.6 ms is used to measure the acoustic cues. The longer time window corresponds to a higher frequency resolution and is used to better capture harmonics.

The results from the perceptual experiment may indicate that more acoustic correlates for the place of articulation of the nasal coda are contained in the vowel region than in the murmur region or vice versa. On the assumption that production and perception are highly related, we should find acoustic evidence in the production of the stimuli to support the results from the perceptual experiment. The second component of this experiment involves looking for acoustic patterns that correlate with the results from the perceptual experiment. If the result of the experiment shows that the perceptual mapping from English to Chinese stimuli by Chinese listeners were based mainly on vowel rather than nasal coda similarity, we would expect more cues for the place of nasal articulation to be contained in the vowel-nasal transition region than in the nasal murmur region for Chinese. That is, we would infer that the cues to the place of articulation that distinguish the final nasal are located largely in the vowel. We would expect a salient change in the formant frequencies of the SC vowels preceding the nasal coda, which provides the listener with information about the identity of the upcoming nasal. This change in the formant frequencies would not be expected in a SC vowel that is not followed by a nasal coda. The change in the vowel acoustics can be considered as the result of an enhancing gesture of the tongue that helps to distinguish the upcoming nasal place of articulation. We will test the change in vowel quality quantitatively by tracking the change in the second formant frequency, F2, which is a measure of the backness vs. frontness of the tongue body. That is, we would expect the tongue body to become more fronted when the low vowel is followed by an alveolar nasal, whose place of articulation is toward the anterior region of the palate, and more backed when the low vowel is followed by a velar nasal, whose place of articulation is toward the posterior region of the palate. In addition, if more cues for the place of nasal articulation are contained in the vowel, in terms of differences in formant movement, we would expect the other cues for nasalization to be less salient in Chinese than in English; for example, we might expect the Chinese vowels to be less nasalized. This is because the evidence for an upcoming place of articulation will allow the listener to infer the presence of a coda consonant, and since this consonant can only be a nasal in SC, less evidence for this feature is required. We will quantify the extent of nasalization in the vowel region by measuring the difference in amplitudes of the first formant frequency and the first nasal pole (A1-P0) for the low vowel and in amplitudes of the first formant frequency and the second nasal pole (A1-P1) for the high vowel.

3.3 Hypothesis

In the two cases where the vowel and the nasal coda stimuli are matched in English and SC, as in En [æṇ] and SC [an], both vowels with relatively high F2̣, En [ɑŋ] and SC [ɑŋ], both vowels with relatively low F2, we would expect the subjects to map En [æṇ] to SC [an], both with high F2, and En [ɑŋ] to SC [ɑŋ], both with low F2, based on similarities in the tongue position of the vowel and the place of articulation of the nasal coda. In the remaining two cases, the vowel and the nasal coda are mismatched, as in En [æŋ], to either SC [an] or [ɑŋ], and En [ɑn] to either SC [an] or [ɑŋ]. If the vowel-nasal transition region contains more acoustic correlates for the place of articulation of the nasal coda than the does the murmur region, then we would expect the subjects to map the En [æŋ] to SC [an] and the En [ɑn] to the SC [ɑŋ], because the vowel pairs are more similar. On the other hand, if the murmur region contains more acoustic correlates for the place of articulation than does the vowel-nasal transition region, then we would expect the subjects to map the En [æŋ] to SC [ɑŋ] and the En [ɑn] to the SC [an], because these pairs have similar murmurs.

We will concentrate mainly on the low vowel in this study because the SC low vowel is unspecified for the feature [back], but will also bring into discussion the high vowel, which is specified for the feature [back]. The tongue body can move to a more fronted or backed position during the production of the low vowel, depending on the place of articulation of the following nasal coda. We have suggested that more acoustic cues for the nasal place of articulation can be found in the SC low vowel, simply as a result of the backward and forward movement of the tongue body during the vowel production.

On the other hand, the high vowel is constrained by the feature [back] and so the tongue body does not have the freedom to move forward and back. Thus the acoustic cues for the nasal place of articulation would not be expected to be found in the SC high vowel.

That is to say, when the distinctive features that specify the SC vowel allow for modifications during its production as influenced by the place of articulation of the nasal coda, there can be expected differences in the distribution of the acoustic cues for the place of articulation of the nasal coda between English and SC. We would expect the SC vowel to bear the burden of containing more acoustic cues for the nasal place of articulation than the nasal murmur itself. Since the two English low vowels are contrastive, [æ] is specified for non-back while [ɑ] is specified for back, there is very little room for the tongue body to be modified by the upcoming nasal coda. We would not expect the English vowel to bear much of the burden of containing the acoustic cues for distinguishing the nasal coda. When the distinctive features that specify the SC vowel do not allow for modifications during its production as influenced by the place of articulation of the nasal coda, it can be expected that there is not much difference in the distribution of the acoustic cues for the place of articulation of the nasal coda between English and SC. As a result, we hypothesize that the SC low vowel, will move to a more fronted or backed position, depending on the nasal coda, while the SC high vowel will not. The movement of the tongue body will be reflected in the second formant trajectory in the acoustics. In sum, we would expect the SC low vowel to contain more cues for the following nasal place in its F2 movement than the SC high vowel does, and more place cues than the English low vowels.

3.4 Results

The core results for the low vowel /a/ are shown in Table 3.2. The numbers indicate how often subjects chose each SC sound to be the most similar to each English sound.

Table 3.2 Results for /a/ mapping

English	Standard Chinese	
	an	ɑŋ
æn	88.0%	12.0%
ɑn	49.8%	50.2%
æŋ	69.0%	31.0%
ɑŋ	19.4%	80.6%

When vowel quality and place of articulation are similar in English and SC, we observe the following mappings: the English [æn], with a high F2, is chosen to be most like the SC [an], with a high F2, and the English [ɑŋ], with a low F2 is chosen to be most like the SC [ɑŋ], with a low F2, to a significant extent, as predicted.

When the experiment was repeated with eight new Mandarin subjects, with one important modification in the design, the results of the experiment were augmented. Instead of introducing the A and B Standard Chinese CVN stimuli aurally, we presented the subjects with Chinese characters on the computer screen. The English stimuli were still presented aurally. The subjects were instructed to choose the SC character on the basis of its similarity to the English stimulus in sound rather than in meaning. Under this design we can be reasonably assured that the experiment is making contact with the grammar because the experimental task requires the subjects to extract the stimuli from their own mental lexicons. The results for this experiment are shown in Table 3.3.

Table 3.3 Results for /a/ characters

English	Standard Chinese	
	an	ɑŋ
æn	96.6%	3.4%
ɑn	52.7%	47.3%
æŋ	92.2%	7.8%
ɑŋ	13.9%	86.1%

The results from the mapping of the English [ɑn] are essentially the same from the first experiment, with a slight increase of (2.9%), from 49.8% to 52.7%, for the mapping of the English [ɑn] to the SC [an]. There is also a small increase of (8.6%), from 88% to 96.6%, for the mapping of the English [æn] to the SC [an] and a large increase of 23.2% for the mapping of the English [æŋ] to the SC [an]. The English vowel /æ/ has the feature [-back]. The significance of the increase in both of the mappings of the non-back English vowel /æ/, regardless of whether it is followed by the alveolar or velar nasal, to the SC [an], is that the SC subjects seem to put the SC low vowel /a/ in the SC syllable [an] in the same category as the English /æ/. This is consistent with the prediction that the tongue body will move to a more forward position during the production of the SC vowel /a/ in the SC syllable [an], and the SC listeners will interpret the high F2 in the resulting acoustics to be similar to the F2 in the English /æ/.

For comparison purposes, we show the results for the high vowel /i/ in Table 3.4. The results indicate that the subjects seem to have trouble distinguishing SC [in] and [iŋ] in this experiment, even though [in] and [iŋ] are contrastive rimes in SC.

Table 3.4 Results for /i/ mapping

English	Standard Chinese	
	in	iŋ
in	52.4%	47.6%
iŋ	42.3%	57.7%

The results seem to support the hypothesis that the high vowel /i/, which is specified for the feature [-back], does not have much room to move the tongue body, and therefore does not provide additional cues for the place of articulation for the nasal coda. Most of the cues to distinguish the nasal coda would have to be contained near the nasal landmark or in the nasal murmur. There is one possible strategy that the SC speakers can use to differentiate the nasal codas: inserting a schwa in [iŋ], resulting in a transitional schwa: [i^{ə}ŋ] (Duanmu, 2002). Since there is no transitional schwa for [in] in SC, the contrast between the velar and alveolar nasal coda, although minimal, can be maintained by the presence vs. absence of the schwa. This could explain why the subjects had difficulty in mapping the English [iŋ], as shown by the confusion between the SC stimuli [in] and [iŋ] in Table 3.4.

3.5 Discussion

The results are consistent with the hypothesis that, because of its high F2, the English non-back low vowel /æ/, regardless of its following nasal coda, is interpreted as the non-back version of the underlying low vowel in SC, which also has a high F2. In contrast, the English back low vowel /ɑ/, with its low F2, when followed by an alveolar nasal, is mapped to either SC [an], with high F2, or [ɑŋ], with low F2, and when followed by a velar nasal, is mapped to the SC [ɑŋ], with low F2.

For the two cases that are not matched in vowel quality and nasal place of articulation, [ɑn] and [æŋ], we observe that, based on the mapping observed for the English [æŋ] (92.2% mapped to SC [an]), listeners seem to match on the basis of F2 in the vowel, suggesting that the acoustic cues for the place of articulation seem to be dominant in the vowel region. However, this cannot account for the mapping of the English [ɑn] to either the Chinese [ɑŋ] (52.7%) or [an] (47.3%). There were 22 English [ɑn] utterances and out of the 14 subjects, four consistently mapped them to the Chinese [ɑŋ] (i.e., about two thirds of the time), three consistently mapped it to the [an] (i.e., about two thirds of the time), and the rest of the subjects mapped to either [ɑŋ] or [an] equally. The results of this mapping seem to be highly dependent on the individual.

In order to explain why the subjects may be mapping based on both the place of articulation and vowel quality or neither for /a/, we examine the acoustics of the nasal codas and the extent of vowel nasalization in the vowels preceding the codas.

3.6 Acoustic analysis of English and Chinese nasal codas

3.6.1 F2 movement

Figure 3.1 below shows the male English speaker's production of 'dɑn' on the top and 'dɑŋ' on the bottom of the left panel and the male Chinese speaker's production of 'dan' on the top and 'dɑŋ' on the bottom of the right panel. The F2 values are similar for both the English words, but F2 is higher in the Chinese 'dan' (about 1500 Hz), than in 'dɑŋ' (about 1100 Hz). That is to say, the vowels are unchanged in English regardless of the place of articulation of the nasal coda, except near the very end of the vowel, during the transition fro vowel to nasal murmur. On the other hand, in SC, the vowel /a/ is fronted throughout most its duration when it occurs next to a more fronted alveolar nasal coda.

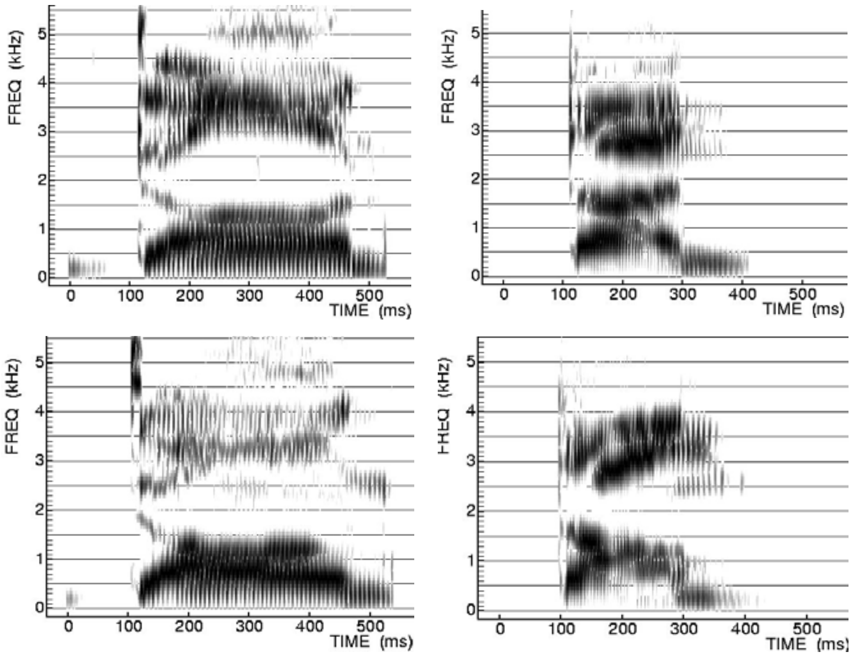

Figure 3.1 Spectrograms showing F2 movement. The left panel shows English male's production of 'dɑn' (top) and 'dɑŋ' (bottom). The right panel shows SC male's 'dan' (top) and 'dɑŋ' (bottom).

Averaged values of F2 in Hz 100 msec before the vowel-to-nasal landmark and 30 msec after the landmark are shown separately for the English speaker (left) and Chinese speaker (right) in Figure 3.2. The 18 CVN utterances began with either a vowel or /d g/, for example, [an], [dan], and [gan]. The nasal landmark was determined by looking at abrupt changes in the waveform from the vowel to the nasal murmur. The nasal landmark is indicated in the graphs at 0 ms, at the point in time where the vowel region (in white) meets the nasal murmur region (in grey). Each F2 value was obtained with a 25.6ms window every 10 milliseconds with the speech analysis program XKL. The lines marked by –x- represent data points for the velar nasals while the lines marked by –o- represent data points for the alveolar nasals.

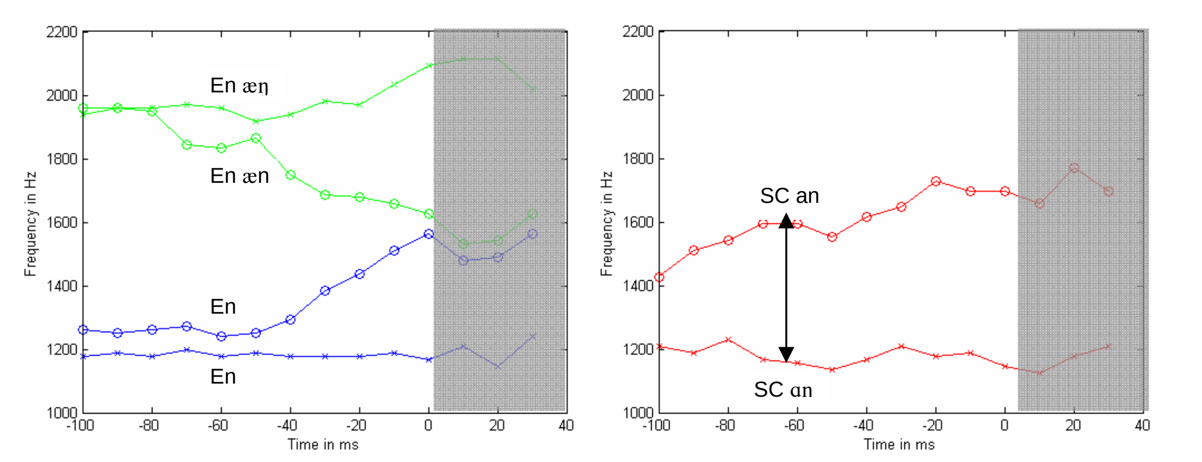

Figure 3.2 Averaged values of F2 movement for 18 vowels from 100 msec prior to the onset of nasal murmur to 30 msec into the nasal murmur. English F2 movement is displayed on left graph in green and blue and indicated by "En". SC data is displayed on right graph in red and indicated by "SC". –x- indicates velar nasal coda and –o- indicates alveolar nasal coda.

We are interested in the reasons behind two mappings: the mapping of the English [æŋ] to the Chinese [an] and the mapping of the English [ɑn] to either the Chinese [an] or [ɑŋ]. The F2 movement for the English [ɑn] is denoted in the figure by [En ɑn], and the

English [ɑŋ] is, denoted by [En ɑŋ]; both are shown in blue on the left panel of the figure. Note that the English speaker produces the vowel /ɑ/, as a back vowel with a low F2, regardless of the nasal coda. On the other hand, the Chinese speaker fronts the normally backed /a/ when it occurs next to the alveolar /n/, (as shown by the higher F2), but keeps the /a/ as a backed vowel when it occurs next to a velar /ŋ/, (as shown by the lower F2) in the right panel of the figure.

The F2 movement for [En æŋ], shown by the top line in green in Figure 3.2, is closer to the F2 movement for [SC an] than to that of [SC ɑŋ] both in the vowel and nasal region. Similarly, the F2 movement for [En æn], indicated by the next line in green, is closer to the F2 movement for [SC an] both in the vowel and the nasal region. The F2 movement for [En ɑŋ] follows the F2 movement for [SC ɑŋ] faithfully. We can conclude then that the perceptual mappings are based on both vowel and consonant similarity for [En ɑŋ] and [En æn], while the mapping of [En æŋ] is based on only vowel similarity.

The F2 movement for [En ɑn] follows the F2 movement for [SC ɑŋ] in the vowel region but follows the F2 movement for [SC an] in the nasal murmur. The F2 movements indicate that the subjects are paying attention to either the vowel quality or the nasal place of articulation when the task is to map the [En ɑn]. Neither seems to take precedence over the other. The [En an] seems to match the vowel in the [SC ɑŋ] and to match the consonant in the [SC an]. Therefore, the subjects chose either the [SC an] or [SC ɑŋ] with equal likelihood. That is to say, the SC subjects are looking for a best match to the English stimuli and there does not seem to be a single dominant acoustic cue. What this suggests is that although there are definitely acoustic cues for contrasting the coda place in the murmur region, there are also additional acoustic cues for contrasting the coda place already in the vowel region. At 60 ms prior to the landmark, there is a difference of more than 400 Hz between the [SC an] and [SC ɑŋ], whereas the difference is only about 50 Hz between the [En ɑn] and [En ɑŋ]. Regardless of the duration of the utterances in English and SC, we expect that the SC low vowel will cue for the place of articulation of the nasal coda with its F2 shift earlier than the English low vowel does. Although the F2 movement shows separation between the back and non-back low vowels in the region near the nasal landmark, the reality is that the English low vowel cannot shift its F2 as early as the SC low vowel because the cues to its back or non-back feature must be maintained as long as possible to ensure perception of this contrast.

3.6.2 Measurements of A1-P0 as an estimate of the degree of nasalization

We have hypothesized that because the nasal codas in Chinese have no other consonants to compete with for perceptual contrast, they may be produced in a more relaxed manner, possibly with a less-than-complete constriction in the oral tract. Consequently, we would expect them to be less nasalized than their English counterparts.

The measure we used to estimate the degree of nasalization is A1-P0, based on earlier studies of the acoustic correlates for nasalization (Chen, 1997). This measure was chosen because the first formant for a low vowel is approximately 750 Hz while the first nasal pole is approximately 250 Hz so there is enough of a separation between the two prominences to extract meaningful measurements. More discussions on this measure are provided in Section 1.5.2. Figure 3.3 shows the averaged A1-P0 values in dB in the vowel region, 70 msec to 20 msec prior to the nasal landmark. The smaller the A1-P0 values are, the more nasal the vowels are. This is a reflection of the physiological opening of the velopharyngeal port, which dampens the formant formant, lowering its amplitude, A1, and at the same time gives rise to a nasal pole, with amplitude P0. The English /ae/ in [aeŋ] and [aen], shown in green, have the smallest (mostly negative) A1-P0 values and are therefore the most nasalized vowels. The Chinese [an] and [aŋ], shown in red, have the largest A1-P0 values and are therefore the least nasalized vowels. This is a reflection of a less open velopharyngeal port, which does not dampen as much the first formant amplitude, A1, and does not give rise to a prominent nasal pole with amplitude P0. The A1-P0 values for the English [an] and [aŋ] fall somewhere in between. This seems to imply that SC nasal codas require less nasalization in the vowel than English nasal codas in order to be perceived as nasal.

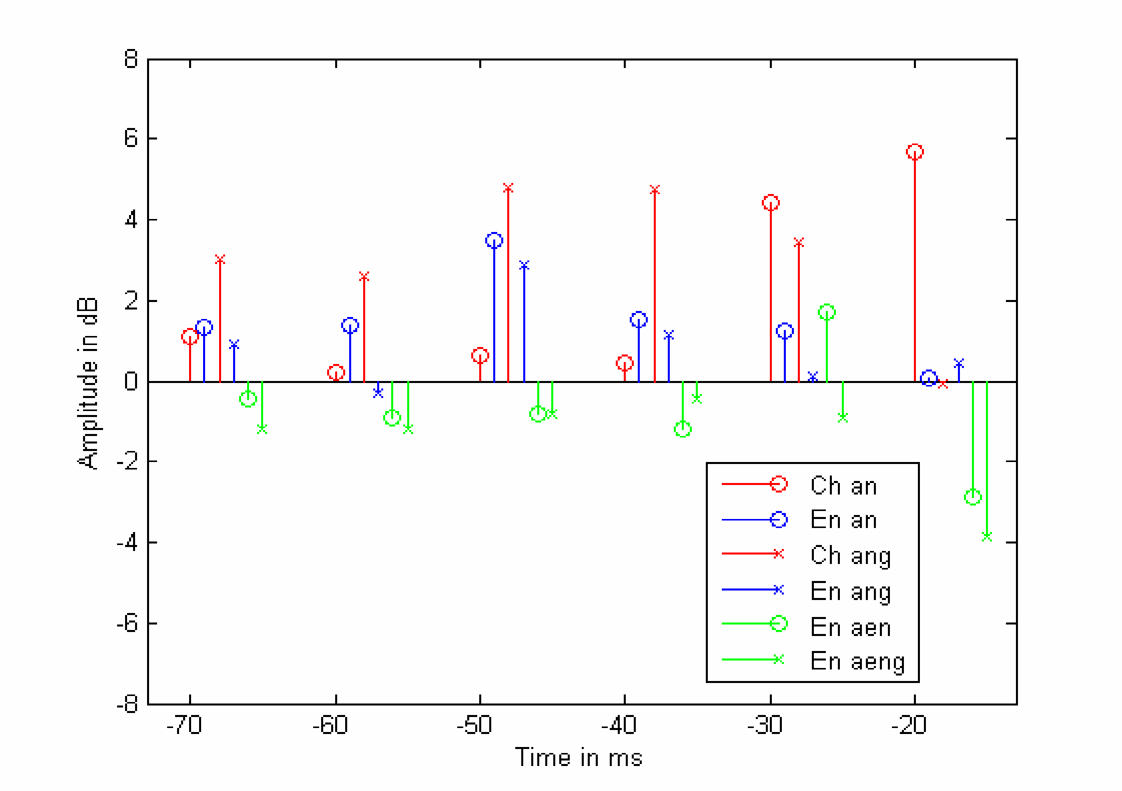

Figure 3.3 Averaged A1-P0 values for the Chinese [an] and [aŋ] in red, the English [an] and [aŋ] in blue, and the English [aen] and [aeŋ] in green, measured 70 to 20 msecs prior to the onset of nasal murmur. –x- indicates velar nasal coda and –o- indicates alveolar nasal coda.

3.7. Summary

The goal of this experiment was to determine whether there are differences in the distribution of the acoustic cues for the place of articulation of nasal codas in English and Standard Chinese, as predicted by phonotactic differences such as the non-specification of the feature [back] for low vowels and the restriction of coda consonants to the alveolar and velar nasals. Results showed that the tongue body is able to move to a more fronted or backed position during the production of the low vowel, to help contrast the alveolar /n/ and the velar /ŋ/. The results can be interpreted as support for the hypothesis that more cues for the [nasal] coda were present in the vowel region in the Chinese nasal-final syllables than in the English nasal-final syllables. That is, the

English vowels did not change, regardless of the nasal coda that followed it, but the Chinese vowels did. This implies that, where phonotactics allow it, vowel modification in Chinese might arise as an enhancing gesture by the tongue body, to increase the salience of cues to the alveolar-velar distinction in the nasal place of articulation.

These observations lead to the question what is systematic in Standard Chinese about modifications on the vowel that are induced by the place of articulation of the nasal coda, which act to contrast the consonants. In the next chapter, we describe a study that examines the other Standard Chinese vowels that occur in the vowel-nasal environment, for possible similar modifications.

This page is intentionally left blank.

Chapter 4. Acoustic analysis of CVN and NV environments

We saw in Chapter 3 that changes in the second formant of the SC low vowel /a/ serve as important and additional acoustic cues for the place of nasalization in the nasal coda. We argued that the tongue body moved forward or backward during the production of /a/, depending on the place of articulation of the nasal coda. That is, the tongue body moves forward when /a/ is followed by an alveolar nasal, and in the acoustic signal we observe an increase in F2. The tongue body moves somewhat back when /a/ is followed by a velar nasal, and in the acoustics we observe a decreasing F2.

In this chapter, we explore if this change in the acoustics is also evident when the mid vowel /ə/ and high vowel /i/ precede a nasal coda, in CVN environments. We will focus the acoustic analysis on these two vowels, as they are the remaining vowels, beside /a/, that are allowed to precede both the velar and the alveolar nasal. The second half of this chapter is devoted to exploring whether this change in the acoustics exists when the three vowels occur in the NV environment, where they follow an onset nasal.

4.1 Other CVN environments

4.1.1 Hypothesis

The results from Chapter 3 suggest that the vowel region contains more acoustic cues for the nasal place of articulation than does the nasal murmur region itself, when the vowel is the low vowel /a/, which is unspecified for the feature [back]. This conclusion is derived from perception tests in which SC subjects mapped the English stimuli to SC sounds based on vowel similarity rather than agreement of the place of articulation in the nasal coda.

The SC mid vowel, /ə/, like the low vowel /a/, is also unspecified for the feature [back]. Therefore, we hypothesize that the mid vowel will also be produced with the tongue body free to move forward or backward, depending on the place of articulation of the nasal coda. The high vowel, /i/, on the other hand, is specified for the feature [back], and thus we hypothesize that it will not be produced with the tongue body free to move forward and backward, regardless of the place of articulation of the nasal coda. Thus, we predict that the non-high vowels will carry more acoustic cues for the place of articulation of the nasal coda than the high vowel does.

4.1.2 Subjects and material

Four native speakers of Standard Chinese (two female and two male), who are students or affiliates at MIT, were recorded reading from a list of CVN syllables in a sound attenuated chamber in the MIT Speech Communication Group. The list of CVN syllables, where the V is a low, mid, or high vowel, appeared on the computer screen one after another as Chinese characters. The characters were chosen from the "Modern Chinese Word Dictionary" and care was taken to pick the most commonly encountered character, whenever there was a choice of allophones. Each character was embedded in the carrier phrase "I say __ this word", so that the entire phrase appeared as Chinese characters on the computer screen. The CVN syllables were presented on the screen so that minimal pairs occurred one after another. For example, /bin/ was followed by /biŋ/ and /ban/ was followed by /baŋ/. The speakers were informed of this pattern and read each phrase as they appeared on screen. The presentation was designed to force the speakers to make the greatest effort to maximally contrast the minimal pairs.

The same native speaker of English who was recruited for the mapping experiment was recorded reading from a list of counterpart CVN English nonsense syllables.

The CVN syllables were then extracted using, PRAAT, a speech analysis program. Each syllable was analyzed using the XKL program, which we described in Chapter 3. The acoustic correlates of vowel nasalization include difference measurements between the amplitudes of the first formant and the first nasal pole, A1-P0, in the low and mid vowels; difference measurements between the amplitudes of the first formant and the second nasal pole, A1-P1, in the high vowel; the first formant frequency, F1; the second formant frequency, F2; and their amplitudes. For each syllable, the nasal landmark was determined as the point in the acoustics where the waveform showed an abrupt discontinuity, where the amplitudes of the first three formants that characterized the vowel preceding the nasal decreased abruptly as a result of the onset of the nasal zero in the adjacent nasal murmur. Values of F1, F2, F3, A1, A2, and A3, and the two nasal poles, P0, and P1, were taken every 10 ms, 70 ms prior to the nasal landmark, and 30 ms after the nasal landmark, for a total of 11 data points.

4.1.3 Results

In Standard Chinese, the only consonants C that are allowed in the syllable-initial position in a CVN environment, where the N can be an alveolar or velar nasal, are the labials /b/, /p/, /m/ and /n/. For example, /jun/ is allowed, but /jung/ is not. Out of these four consonants, the nasals are eliminated from analysis because we want to isolate the effect of the nasal coda on the vowel in the middle. We also eliminated /p/ from the analysis because the burst release of the unvoiced consonant interfered with the tracking of the vowel formants. The only consonant left, therefore, was /b/. In our analysis, we focus only on /bin/, /bing/, /ban/, /bang/, /ben/, and /beng/.

4.1.3.1 Formant shift in SC and English /a/

The averaged values of the first three formants for the SC speakers and for the English speaker during the utterances /ban/ and /baŋng/ are displayed separately in Figures 4.1 and 4.2. The behavior of the second formant, F2, is highlighted in both figures. For the SC speakers, F1 and F3 for both /ban/ and /baŋ/ follow similar trajectories, while F2 of /baŋ/ is about 500 Hz lower than the F2 of /ban/, even during the 70 ms before the nasal landmark. A higher F2 for /ban/ is a reflection of a more fronted tongue body during the production of /a/ when it is followed by an alveolar nasal. This is consistent with the findings of the acoustic analysis in the previous chapter. These results are also consistent with the findings of Lin & Yan (1991), where the end-points of F2 in the low vowel /a/ in [an] and [aŋ] showed significant differences (T-value=22.708 and P<0.001, see Appendix A).

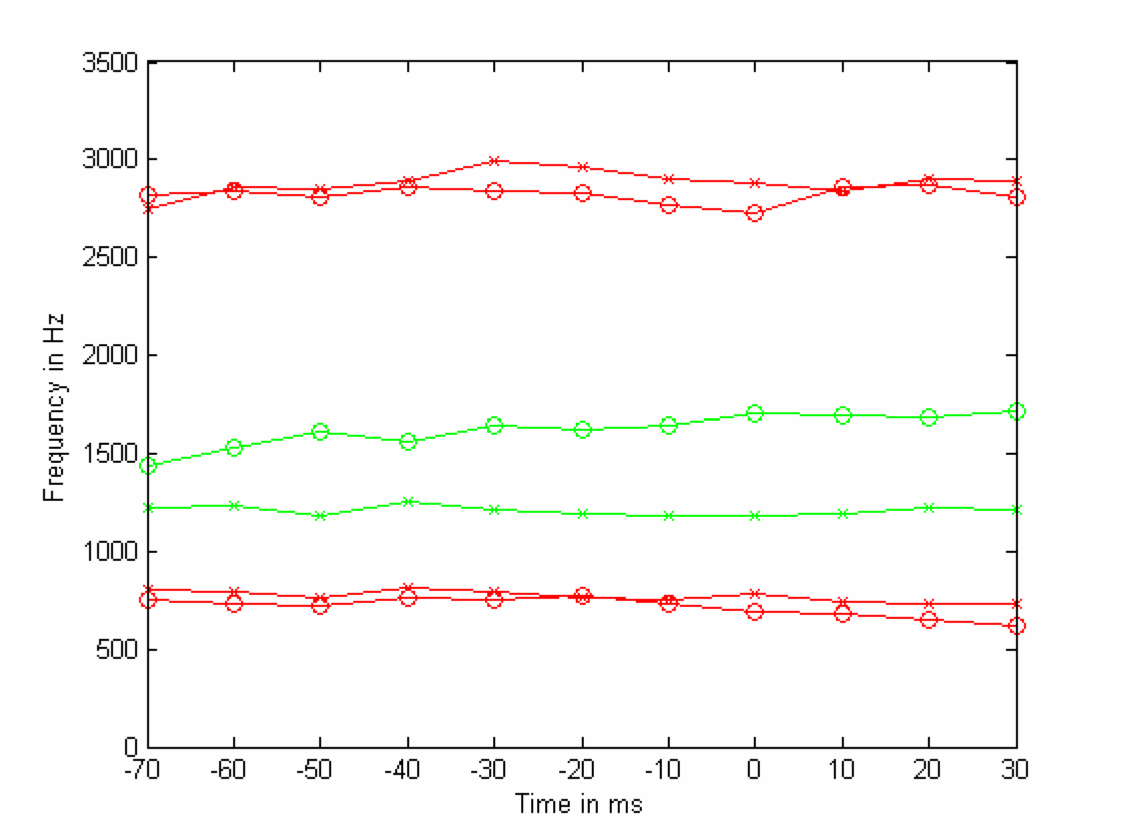

Figure 4.1 Averaged values of the first three formants for the SC speakers. –x- indicates /baŋ/ and –o- indicates /ban/.

On the other hand, figure 4.2 shows that the native English speaker makes a greater effort to keep the tongue back throughout the production of /ban/, although F3 is lower for the /ban/ than it is for /baŋ/. The English speaker also makes a velar pinch at the nasal landmark, as indicated by 0 ms in the graphs, where the second and third formants come closer together. There is no noticeable velar pinch in the production of the SC /baŋ/.

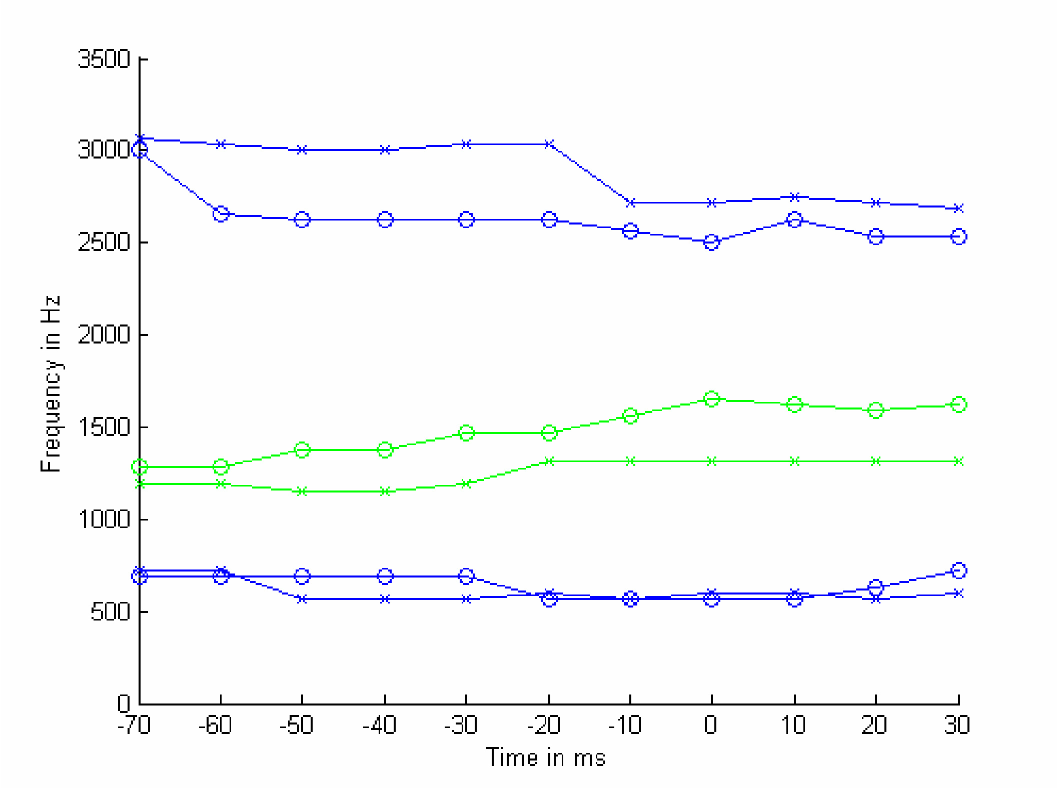

Figure 4.2 Averaged values of the first three formants for the English speaker. –x- indicates /baŋ/ and –o- indicates /ban/.

4.1.3.2 Formant shift in SC and English /e/

The averaged values of the first three formants for the SC speakers and for the English speaker during his utterances of /ben/ and /beŋ/ are displayed separately in Figures 4.3 and 4.4. The behavior of the second formant, F2, is highlighted in both figures. We see the same patterns in the formant trajectories as for /ban/ and /baŋ/, only more pronounced. For the SC speakers, F3 for both /ben/ and /beŋ/ follow similar trajectories, while F2 of /beŋ/ is again about 500 Hz lower than the F2 of /ben/. In addition, we also observe that the SC F1 for /ben/ is nearly 200 Hz lower than that for /beŋ/. The lowering of F1 and raising of F2 renders the vowel /ə/ in /ben/ much more like a front vowel. Again, the results confirm the hypothesis that the mid vowel /ə/, which is also unspecified for the feature [back], behaves like that low vowel /a/, in that the tongue body is free to move forward and back, depending on the nasal coda. The results of the acoustic analysis are also consistent with the findings of Lin & Yan (1991), where the end-points of F2 in the mid vowel /ə/ in [ən] and [əŋ] showed significant differences (T-value=20.624 and P<0.001) (See Appendix A).

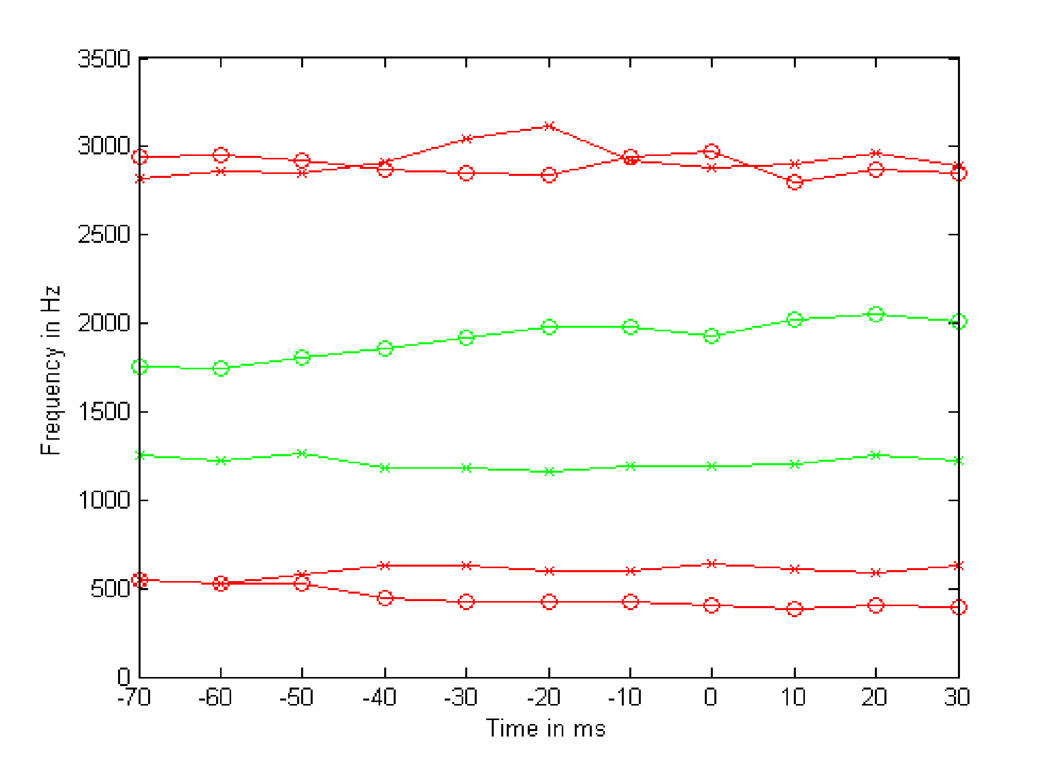

Figure 4.3 Averaged values of the first three formants for the SC speakers. –x- indicates /beŋ/ and –o- indicates /ben/.

On the other hand, the first three formants do not show much difference in their trajectories between the native English speaker’s productions of /ben/ and /beŋ/.

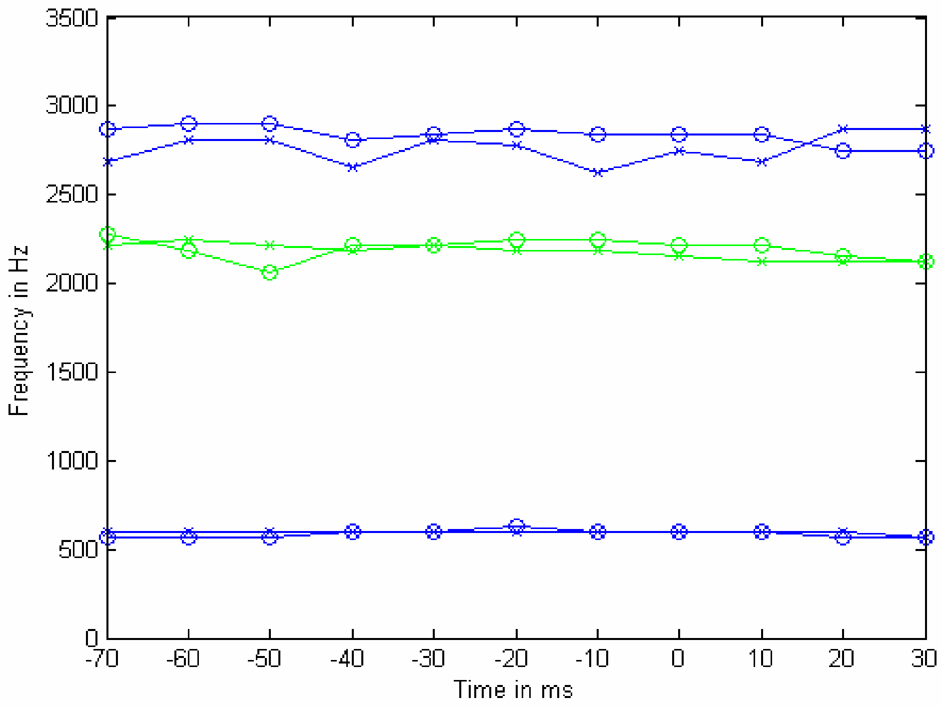

Figure 4.4 Averaged values of the first three formants for the English speaker. –x- indicates /beŋ/ and –o- indicates /ben/.

Although the English speaker does not front or back the tongue body during the vowel to help contrast the nasal coda, he may be manipulating some parameter other than the feature [back] to enhance that place contrast. Figure 4.4 shows only the movement of F2, which is an important acoustic correlate for the feature [back] in the vowel region, but it does not give any information about the oral closure made for the nasal murmur. The acoustic correlates of the nasal murmur include the presence of nasal poles, whose placement and amplitude reflect the places of articulation of the nasal coda.

Figure 4.5 shows the spectrograms for /ben/ (left) and /beŋ/ (right) as produced by the SC speaker XS (top panels) and by the English speaker (bottom panels). The F2 of the SC /ben/ rises from about 1450 Hz to about 1750 Hz while the F2 of the SC /beŋ/ rises and falls around 1 kHz. The F2 of the English /ben/ and /beŋ/, on the other hand, stays approximately constant at 2100 Hz, as is appropriate for a mid vowel. The spectrograms for the SC /ben/ and /beŋ/ show that aside from differences in the trajectories of F2, there is not much difference elsewhere in the acoustics. The spectrograms for the English /ben/ and /beŋ/, however, show that while F2 during the vowel is not an important factor in distinguishing the place of articulation in the nasal coda, there is the presence of a strong resonance near 3000 Hz during the nasal murmur of /beŋ/ that is nearly absent in the nasal murmur of /ben/, which corresponds to third nasal pole for a for a velar nasal consonant. Thus the contrast between the velar and alveolar nasal coda is maintained with acoustic cues found in the nasal murmur region rather than in the vowel region.

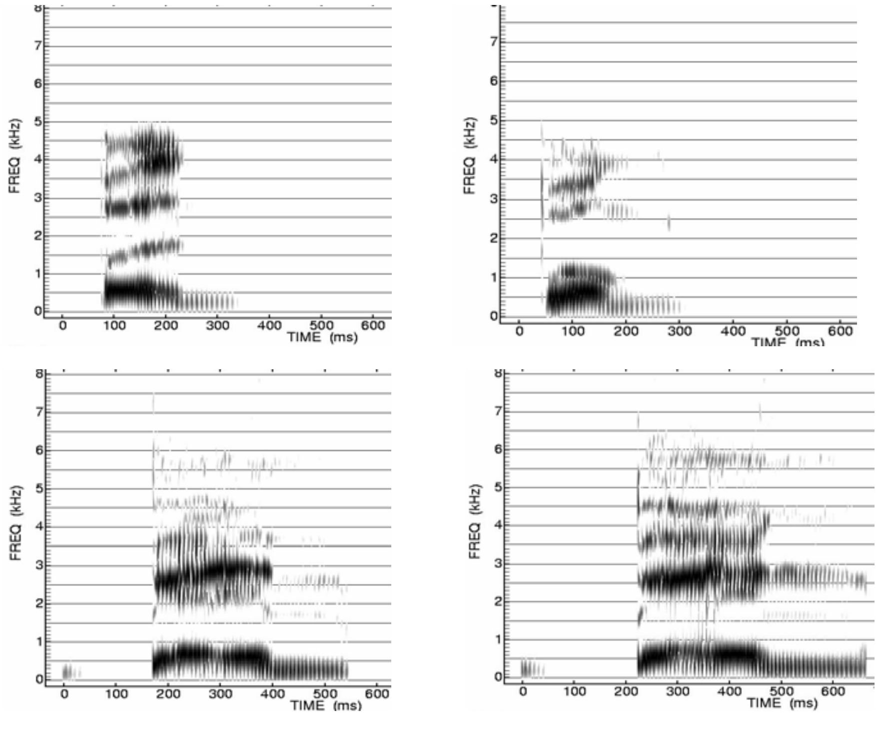

Figure 4.5 Spectrograms of /ben/ (left) and /beŋ/ (right) as produced by the SC speaker XS (top panels) and by the English speaker (bottom panels).

4.1.3.3 Formant shift in SC and English /i/

The averaged values of the first three formants during the utterances of /bin/ and /biŋ/ are shown separately for the SC speakers in Figure 4.6 and for the English speaker in Figure 4.7. The first two formants for /bin/ and /biŋ/ do not differ much for both the English and the SC speakers. We have discussed that the SC /i/ is specified for the feature [-back]. We do not expect the speaker to be able to manipulate his tongue to make it more fronted or backed, and we do not expect the acoustic correlates for the place of articulation of the nasal coda to fall in the region of the high vowel. The results are also consistent with the findings of Lin & Yan (1991), where the end-points of F2 in the high vowel /i/ in [in] and [iŋ] did not show much significant differences (T-value = 6.703 and $P < 0.001$) compared with those values for the non-high vowels (See Appendix A).

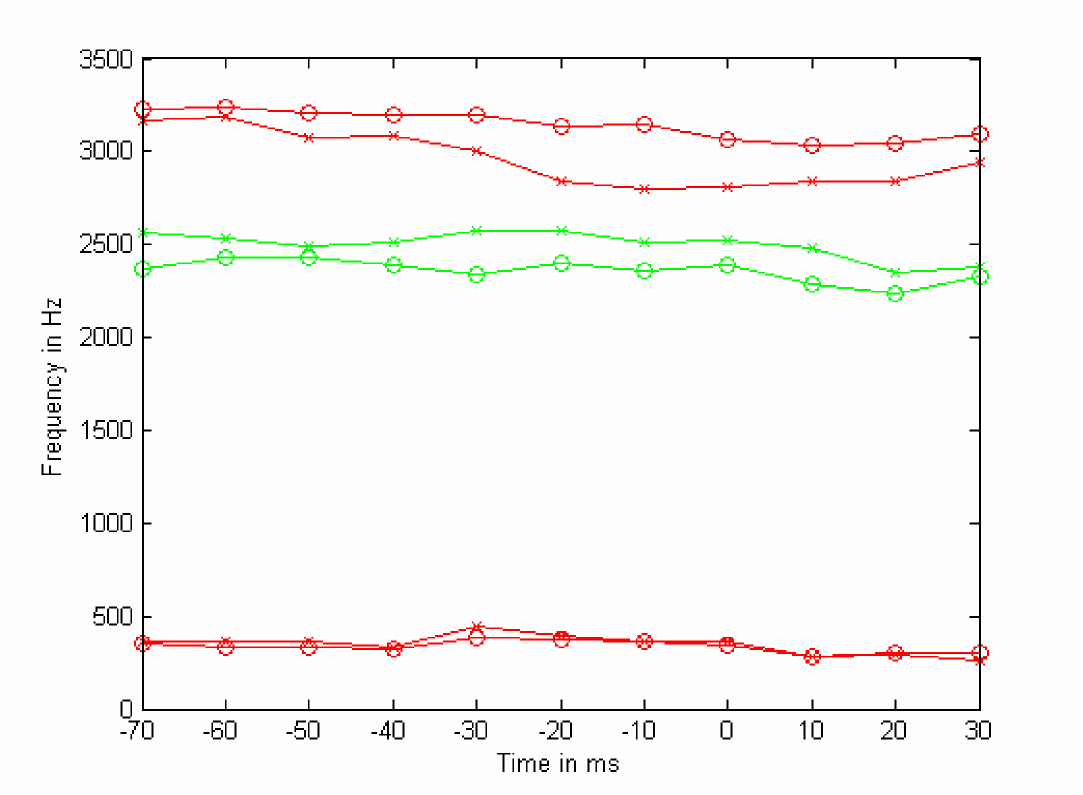

Figure 4.6 Averaged values of the first three formants for the SC speakers. –o- indicates /bin/ and –x- indicates /biŋ/.

What seems to distinguish the alveolar nasal coda (indicated by –o- in figure) from the velar nasal coda (indicated by –x- in figure) is the coming together of the second and third formants that indicates a velar pinch. That is, the acoustic correlate for the place of articulation of the nasal coda is not found in the vowel region, but rather close to the nasal landmark. This is also consistent with the findings of Lin & Yan (1991), where the end-points of F3 in the high vowel /i/ in [in] and [iŋ] showed significant differences (T-value = 10.008 and P < 0.001) compared with those values for the low vowel (T-value = 0.349 and P > 0.05) (See Appendix A). A similar velar pinch is observed in the English speaker's production of the velar nasal, as shown in Figure 4.7.

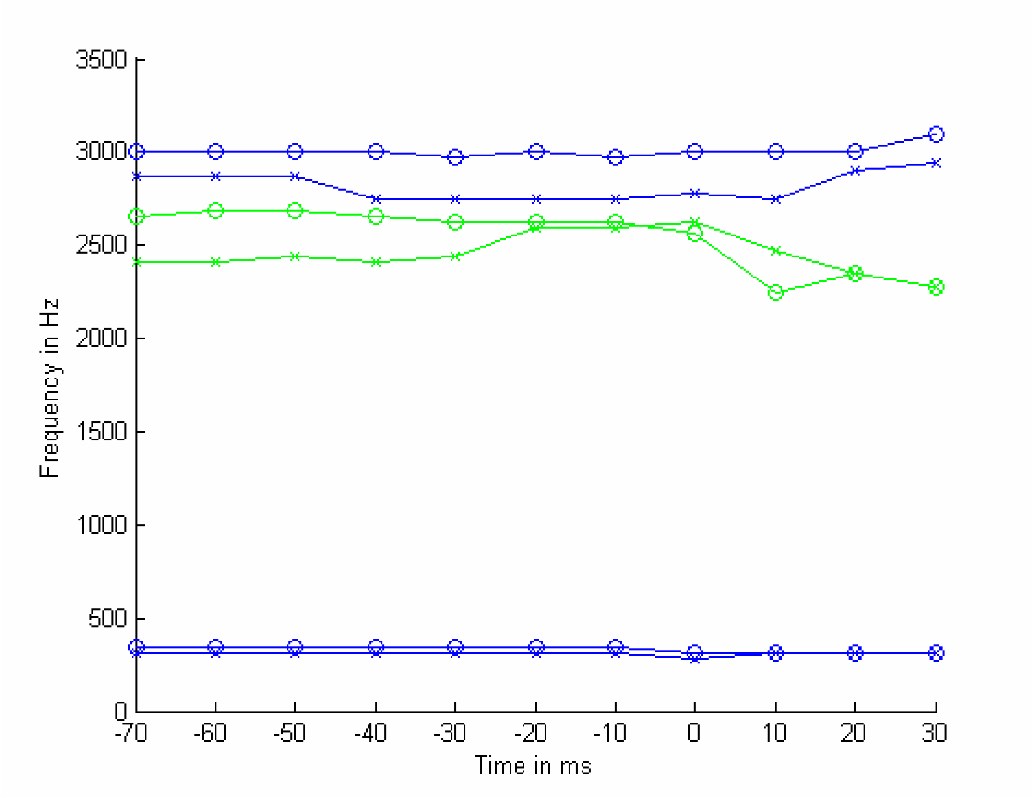

Figure 4.7 Averaged values of the first three formants for the English speaker. –o- indicates /bin/ and –x- indicates /biŋ/.

The F3 values are in general higher for the SC speakers than for the English speaker. This is to be expected since our SC speakers include two females and their formant frequencies will tend to be higher as females have a shorter vocal tract compared to male counterparts. The F3 trajectories of the English and SC production of /bin/ and /biŋ/ are otherwise quite similar. For /biŋ/, both the English and SC productions exhibit a dip in F3 to about 2800 Hz at the nasal landmark, which is 0 ms in the figures. Thus both the English and SC speakers are making the velar pinch to contrast the alveolar coda from the velar coda.

4.1.3.4 Nasalization measures of A1-P0 and A1-P1

Measures of the extent of vowel nasalization between the period 70 ms and 30 ms prior to the nasal landmark are shown in Figure 4.8. The top panel shows the difference measurements of A1-P0 for the low vowel /a/. The middle panel shows the difference measurements of A1-P0 for the mid vowel /ə/. The bottom panel shows the difference measurements of A1-P1 for the high vowel /i/. Measures from the production of SC speakers are indicated in red while those of the English speaker are highlighted in blue. The lower the difference measurements of A1-P0 or A1-P1, the more nasalized the vowel. For the vowel /a/, the SC speakers nasalize to a lesser extent than the English speaker. For the vowel /ə/, the extent of nasalization is comparable between the SC and English speakers. For the vowel /i/, it seems that the SC speakers nasalize to a lesser extent than the English speaker.

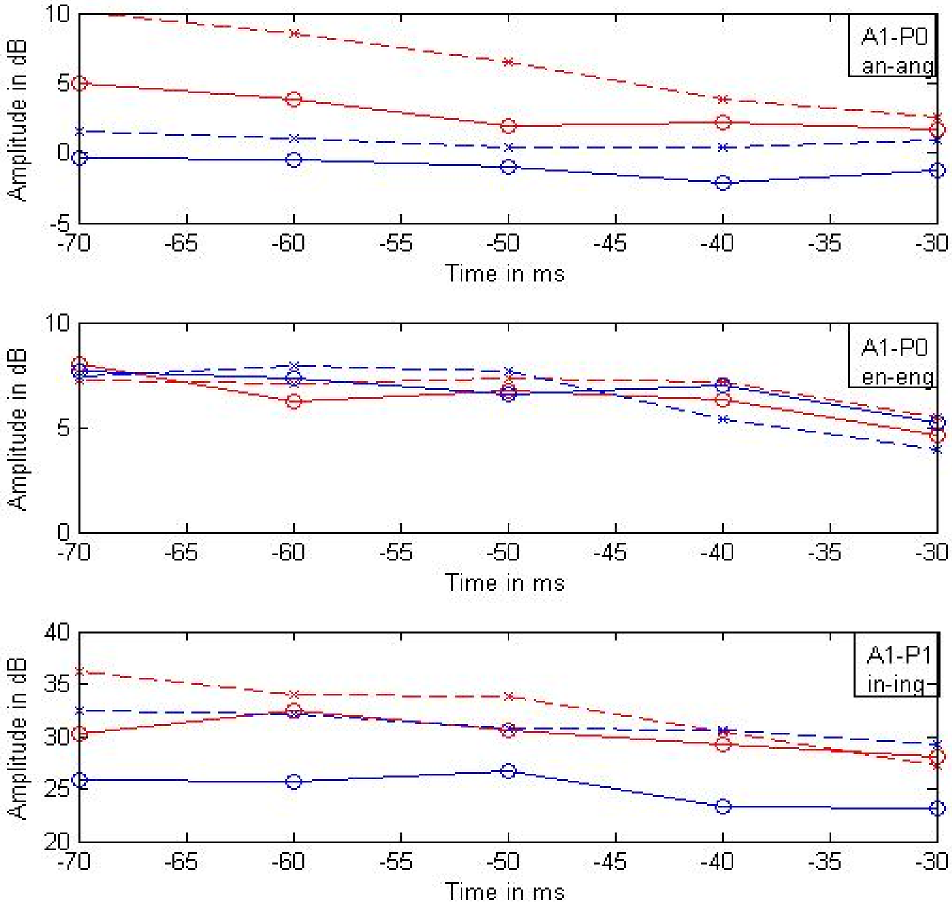

Figure 4.8 Measures of nasalization shown for /a/ (top panel), /ə/ (middle panel), and /i/ (bottom panel), red for SC speakers, and blue for the English speaker. The velar coda is indicated by –x- and the alveolar coda is indicated by –o-.

4.1.4 Discussion

The results presented in the previous section seem to confirm the hypothesis that whenever the phonotactics allow for it, SC speakers manipulate their tongue bodies during the production of the vowel to help contrast the place of articulation of the upcoming nasal coda. The phonotactics of SC allows the low and mid vowel to have the freedom to change their backness, depending on the nasal coda, as they are both unspecified for the feature [back]. The fronting of the tongue body during the production of a vowel that precedes an alveolar nasal coda could be considered an enhancing gesture for contrasting the place of articulation in the nasal coda in SC and the acoustic manifestation of such an articulatory event in vowel formant transition can be considered an additional acoustic correlate for nasalization in SC. Although fronting and backing the tongue body enhances the place of articulation, since only nasals posses these features in SC, they implicate nasalization.

The high vowel, however, is specified for [-back]. The SC speaker must take care to produce the VN combination, where V is the high vowel /i/, so that certain universal features of nasalization are maintained. There is not much the speakers can do during the vowel region. Instead, both the SC and English speakers make the velar pinch near the nasal landmark to help contrast the velar nasal from the alveolar nasal.

Figure 4.8 shows that the SC /a/ in the VN combination is not as nasalized as the English /a/. This does not conflict with our hypothesis that since most of the burden for contrasting the place of articulation of the nasal coda falls within the vowel region by differences in the trajectories of the second formant, /a/ does need to be as nasalized as the English counterpart. The moment a listener detects some vowel shift in /a/, he can already expect a nasal coda.

Similarly, we would expect the SC vowel /ə/ to be less nasalized than the English /e/ in the VN combination. Figure 4.3 showed that the formants F1 and F2 come together for the backed /ə/ and that they are further apart for the fronted /i/. A1-P0 measurements, however, indicate that the extent of nasalization for /ə/ is comparable between the English and SC productions. We know that the A1-P0 measurement is robust when the vowel is a low vowel because F1 and F_{P0} are far apart enough that the poles do not obscure each other. The first formant of the low vowel /a/ is approximately 750 Hz while the first nasal pole is approximately 250 Hz. The first formant of the mid vowel, however, is approximately 500 Hz. When the SC speaker produces a more fronted /ə/ in /ben/, the F1 is furthered lowered, to about 400 Hz. This brings the first formant much closer to the first nasal pole, and they may obscure each other so the measure of their amplitudes, A1-P0, is lower than expected.

On the other hand, we do not expect the SC /i/ in the CVN combination to be less nasalized than the English /i/. Since the SC speakers cannot manipulate their tongues to produce the vowel differently, depending on the nasal coda, we expect the vowel to be produced with stronger adherence to the universal acoustic correlates of nasalization, i.e., more nasalization, so that a CVN syllable can be readily distinguished from a CV syllable. We note here that the measure for nasalization for the high vowel is A1-P1, where P1 is the amplitude of second nasal pole. We have discussed in chapter 2 that the presence of this pole, around 1 kHz, is considered a robust acoustic correlate for vowel nasalization. We have also explored what happens when the cross-sectional area of the velopharyngeal port is varied, so that this nasal pole can actually take on a range of values, namely between approximately 900 Hz and 1300 Hz. We have also shown with a simplified nasal model consisting of three tubes coupled at the velopharyngeal port, a quasi-quantal relationship between the articulator responsible for the perception of nasality, and the acoustic manifestation of its movements. That is, when the velopharyngeal port is opened beyond 0.3 cm^2, the nasal pole increases, but does not increase nearly as much as it does when the port is opened from about 0.1 to 0.2 cm^2. If the SC speakers do not need to be so precise in their production of the nasal, then they could be more relaxed about the control of the velopharyngeal port. A wider opening would lead to a higher F_{P1}. Since the measures of P1 were taken at a fixed F_{P1}, at around 1 kHz, they may be actually lower than the actual P1, which correspond to a higher F_{P1} that was not considered when taking the measurements.

Figure 4.9 shows the spectrograms and spectra taken in the middle of /bi/ and /bin/, as produced by the English speaker. Notice the prominent nasal pole at around 1000 Hz when the utterance is a nasal /bin/. It is not prominent in /bi/.

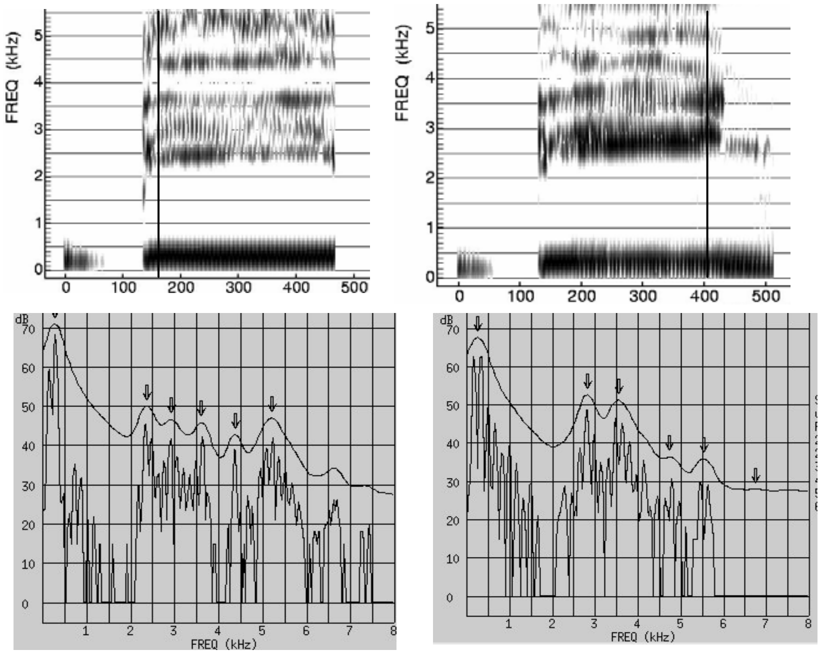

Figure 4.9 Spectrograms of /bi/ (left) and /bin/ (right), as produced by the English speaker, are shown on the top panel and their spectra taken in the middle of the utterances are shown on the bottom panel.

Figure 4.10 shows the spectrograms and spectra taken in the middle of /bi/ and /bin/, as produced by the SC speaker JZ. Notice the extra peak at around 1300 Hz when the utterance is a nasal /bin/. It is not present in /bi/. This could be the actual nasal pole, which would correspond to a more open velopharyngeal port.

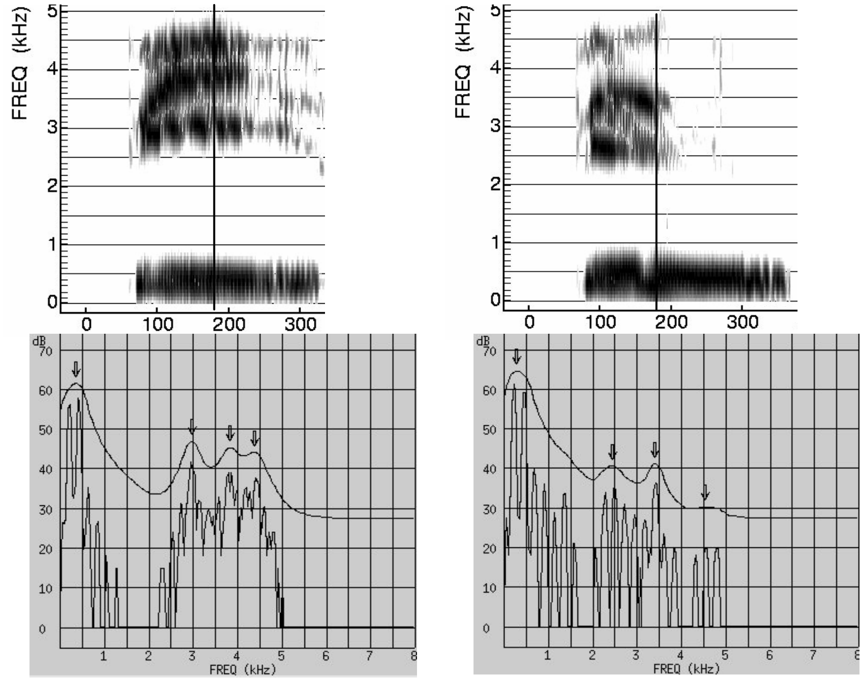

Figure 4.10 Spectrograms of /bi/ (left) and /bin/ (right), as produced by the SC speaker JZ, are shown on the top panel and their spectra taken in the middle of the utterances are shown on the bottom panel.

Further examination for individual speaker differences in the spectra of /bin/ reveals that in one speaker, there is consistently a pole at approximately 1600 Hz that is more prominent than the pole at 1 kHz. This seems an anomaly, as our nasal model in Chapter 2 indicates that the nasal pole can only increase so much, and actually levels off after 1300 Hz. Figure 4.11 shows the spectrograms and spectra taken in the middle of /bi/ and /bin/, as produced by the speaker YP. Notice the extra peak at above 1500

Hz when the utterance is a nasal /bin/. It is not present in /bi/. If this is not a nasal pole, then what is responsible for this prominence in the spectrum? Could this peak at 1500 Hz be the acoustic correlate of an event related to nasalization that is not nasalization itself? That is, could this peak be due to some other event that contributes to the perception of nasalization but is not a direct consequence of the primary gesture correlated with nasalization, namely, the opening of the velopharyngeal port?

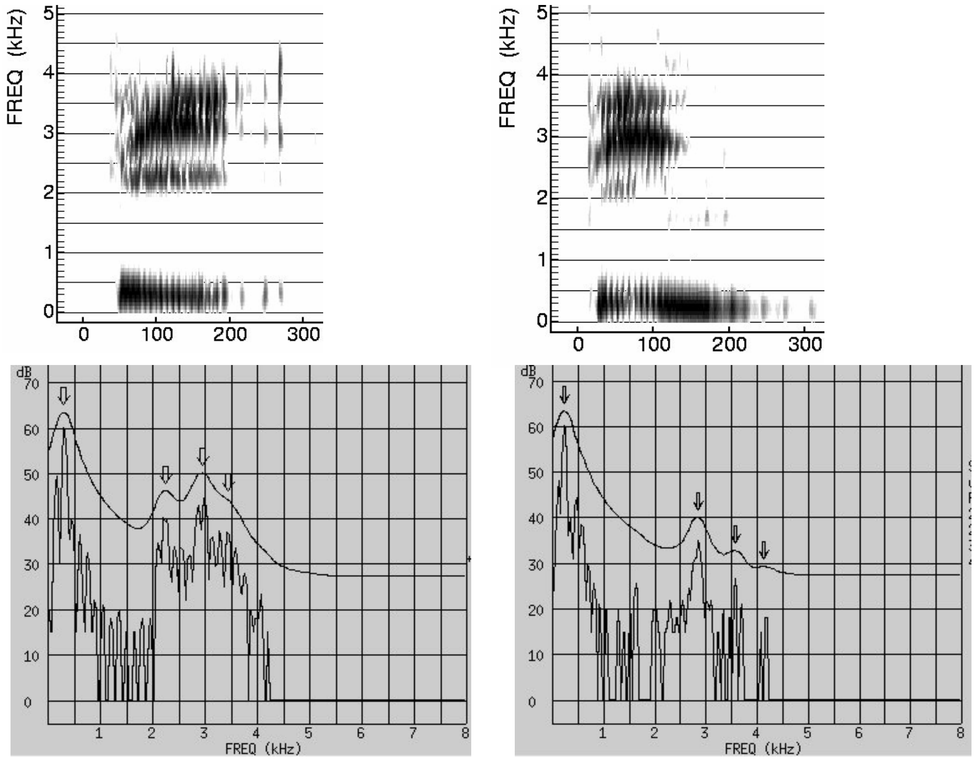

Figure 4.11 Spectrograms of /bi/ (left) and /bin/ (right), as produced by the SC speaker YP, are shown on the top panel and their spectra taken in the middle of the utterances are shown on the bottom panel.

On closer examination, the spectrum of the /bin/, produced by the male speaker YP, shows more loss of energy at higher frequencies than the spectra of /bin/ from the production of other speakers. When the glottis is spread or when it is not closed entirely, this leads to loss of energy at higher frequencies.

The principal consequence of the acoustic coupling between the main vocal tract and the nasal cavity is a flattening of the spectrum in the vicinity of F1. This decreased low-frequency prominence is due to increased acoustic losses on the extensive surface area of the nasal cavity as well as to the introduction of additional spectral peaks in the vicinity of F1. An increased F1 bandwidth and a general flattening of the spectrum at low frequencies also occur when there is vocal fold vibration with a spread glottis (Hanson, 1997). The glottal opening contributes acoustic losses to the lowest vocal tract resonance, and this also enhances the amplitude of the first harmonic. Some of the acoustic consequences of spreading the glottis, therefore, are similar to the primary acoustic correlates of nasalization. The SC speaker YP may be spreading his glottis to further enhance the perception of nasality in his speech. This is consistent with the acoustic theory of speech production, where speech production is driven by the acoustic goal. The second subglottal resonance has been shown to be approximately at 1555 Hz (Sonderegger & Chi, 2004).

4.2 NV environments

We have examined some of the acoustic correlates of nasalization in the CVN context and the primary gesture associated with nasalization, which involves the opening of the velopharyngeal port, as well as some enhancing gestures that speakers of SC may use to help contrast the place of articulation of the nasal coda. We have attributed the changes in the second formant, which is an acoustic manifestation of the changes in the tongue position, in the SC productions CVNs, to the differences in the phonotactic constraints between SC and English.

4.2.1 Hypothesis

In the syllable-initial position, the nasal are not the only consonants allowed in Standard Chinese, so the phonotactic constraint that may have been responsible for the differences observed in the syllable-final position would not apply here. We would expect more similarity between the production of syllable-initial nasals in SC and in English. We would not expect F2 change in the vowel to be an enhancing gesture to contrast the place of articulation of the syllable-initial nasal. We would not expect much difference in the distribution of the acoustic correlates for nasalization between English and Chinese.

4.2.2 Task and materials

We compare the English and SC productions of NVs, where N is either the labial nasal /m/ or the alveolar nasal /n/. /m/ and /n/ are the only two nasals allowed in the syllable-initial position in both English and SC.

Four native speakers of Standard Chinese (two female and two male), who are students or affiliates at MIT, were recorded reading from a list of NV syllables in a sound attenuated chamber in the MIT Speech Communication Group. The procedure is exactly the same as that described in section 4.1.3. The NV syllables, where the N is either the labial /m/ or the alveolar /n/, and V is a low or high vowel, appeared on the computer screen one after another as Chinese characters. The mid vowel was eliminated because in the NV combination, /ne/ and /me/ are often pronounced as /na/ and /ma/. In fact, the mid vowel cannot follow many consonants. For example, /be/ and /fe/ do not occur. Each character was embedded in the carrier phrase “I say __ this word”, so that the entire phrase appeared as Chinese characters on the computer screen. The same native speaker of English recruited to record nasal codas was recorded reading from a list of counterpart NV syllables.

The NV syllables were then extracted using, PRAAT, a speech analysis program. Each syllable was analyzed using the XKL program, which we described in Chapter 3. The acoustic correlates of vowel nasalization include difference measurements between the amplitudes of the first formant and the first nasal pole, A1-P0, in the low vowel; difference measurements between the amplitudes of the first formant and the second nasal pole, A1-P1, in the high vowel; the first formant frequency, F1; the second formant frequency, F2; and their amplitudes. For each syllable, after the nasal landmark was handpicked, values of F1, F2, F3, A1, A2, and A3, and the two nasal poles, P0, and P1, were taken every 10 ms, from 30 ms prior to the nasal landmark, to 70 ms after the nasal landmark, for a total of 11 data points. The landmark for nasal release is set at the 0 ms mark. At 0 ms, the nasal murmur is released into a vowel.

4.2.3 Results

4.2.3.1 English and SC production of /ma/ and /na/

The trajectories of the first three formants for both the English and the averaged SC productions of /ma/ are shown in figure 4.12. The English and the SC productions of the low vowel are very similar. Speakers of both languages produce the low vowel so that F1 is about 800 Hz, F2 about 1250 Hz, and F3 about 2600 Hz. F3 is slightly higher for the SC speakers than for the English speaker. This could be the result of averaging 4 SC speakers, two of whom are females, who are expected to have slightly higher formants. Here the nasal is the labial /m/ and since the constriction is made with the lips, the tongue is free to assume the position required for the back vowel, so that we would expect a relatively constant trajectory, even right from the release of the nasal.

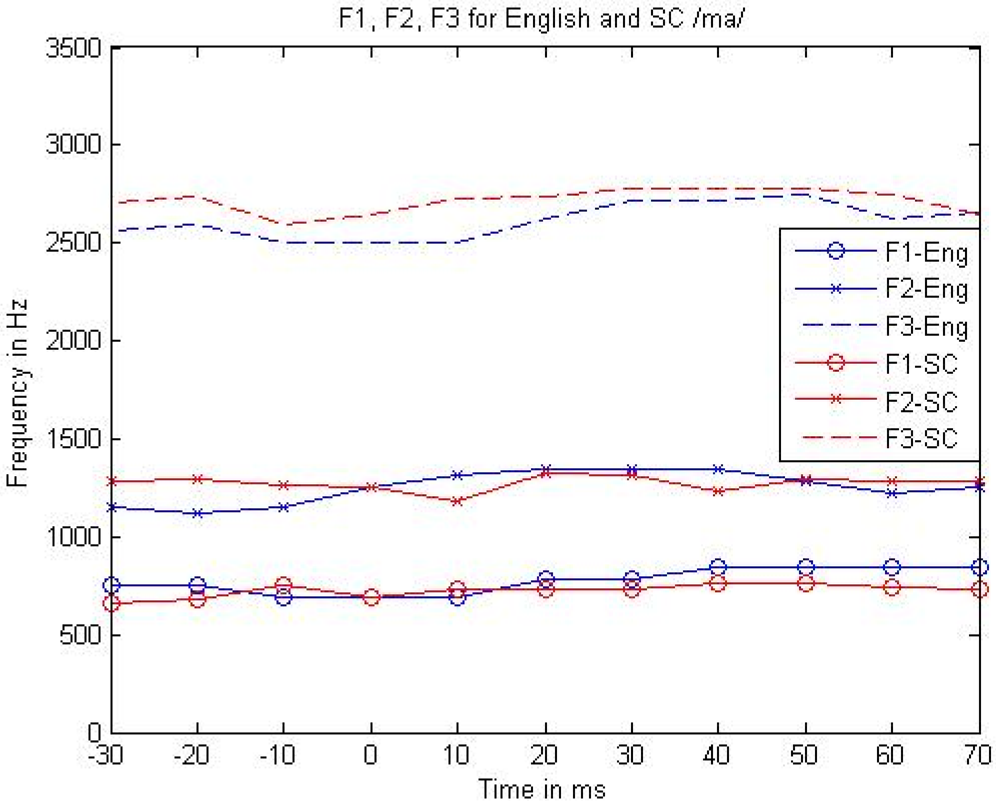

Figure 4.12 Averaged values of the first three formants for the production of /ma/ by the SC speakers (shown in red) and by the English (shown in blue) speaker. –o- indicates F1, –x- indicates F2, and -- indicates F3.

The trajectories of the first three formants for both the English and the averaged SC productions of /na/ are shown in figure 4.13. Again, the English and the SC productions of the low vowel, which follows the alveolar nasal, are very similar. Unlike the production of /ma/, where the oral closure is made with the lips, the closure for /na/ is made with the tongue blade. The tongue has to move forward to make the constriction first for /n/ before it can assume its more backed position for the vowel /a/. Speakers of both languages produce the low vowel so that F1 and F2 transition smoothly from those of a more neutral vowel, from about 500 Hz and 1500 Hz, to those of a low vowel, at 750 Hz and 1250 Hz.

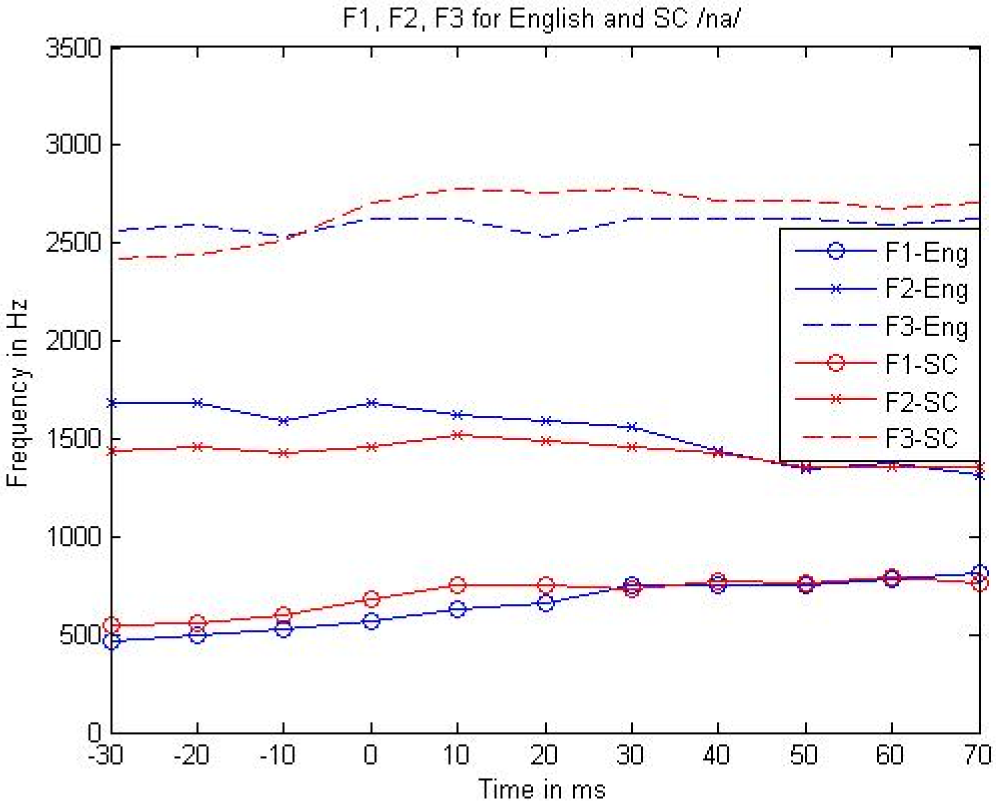

Figure 4.13 Averaged values of the first three formants for the production of /na/ by the SC speakers (shown in red) and by the English (shown in blue) speaker. –o- indicates F1, –x- indicates F2, and -- indicates F3.

4.2.3.2 English and SC production of /mi/ and /ni/

The trajectories of the first three formants for both the English and the averaged SC productions of /mi/ are shown in figure 4.14. The English and the SC productions of the high vowel are very similar. The first three formants of /mi/, as produced by the English speaker, are about 300 Hz, 2300 Hz, and 3000 Hz, respectively. The first three formants of /mi/, as produced by the SC speakers, are slightly higher, again, presumably because two of the speakers are females. They are at approximately 400 Hz, 2500 Hz, and 3200 Hz respectively.

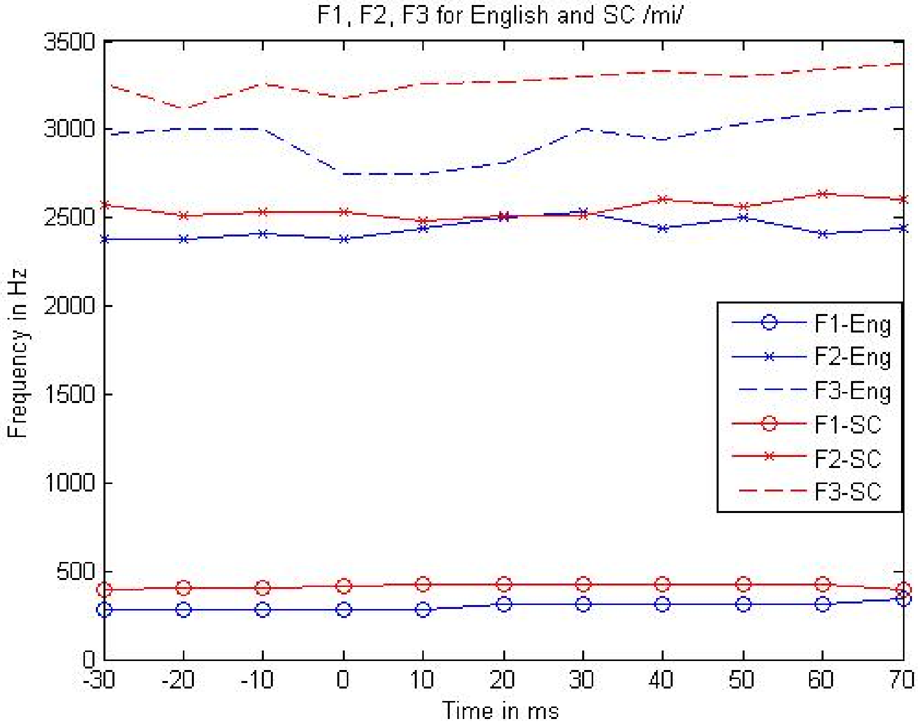

Figure 4.14 Averaged values of the first three formants for the production of /mi/ by the SC speakers (shown in red) and by the English (shown in blue) speaker. –o- indicates F1, –x- indicates F2, and -- indicates F3.

Here the nasal is the labial /m/ and since the constriction is made with the lips, the tongue is free to assume the position required for the high vowel, so that we would expect a relatively constant trajectory, even right from the release of the nasal.

The trajectories of the first three formants for both the English and the averaged SC production of /ni/ are shown in figure 4.15. The English and the SC productions of the high vowel are very similar. The average first formant of /mi/, as produced by the SC speakers, is slightly higher than the first formant of /mi/, as produced by the English speaker, again, because two of the SC speakers are females.

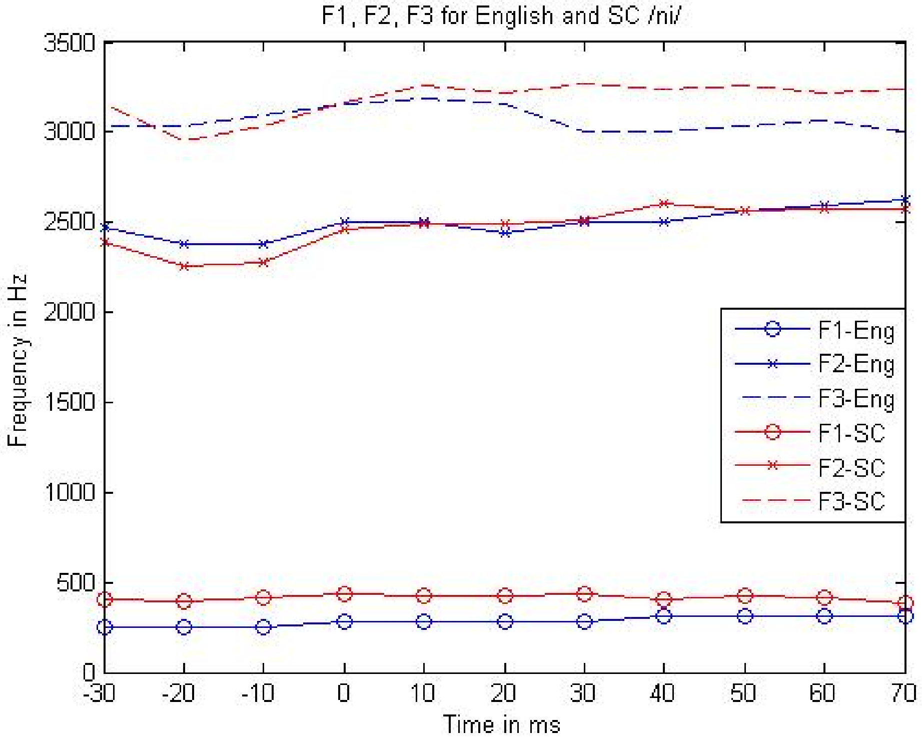

Figure 4.15 Averaged values of the first three formants for the production of /ni/ by the SC speakers (shown in red) and by the English (shown in blue) speaker. –o- indicates F1, –x- indicates F2, and -- indicates F3.

Here the nasal closure for /ni/ is made with tongue. The tongue has to move forward to make the constriction first for /n/ before it can assume its more backed position for the vowel /a/. However, since the vowel that follows it is a high vowel, the trajectories of the formants are expected to stay constant. Speakers of both languages produce the high vowel so that F1, F2, and F3 maintain values that are appropriate for a high vowel throughout the utterance. The contrast between the /m/ and /n/ seems to be maintained by differences in the second and third formant frequencies in the murmur region.

4.2.3.3 Nasalization measures of A1-P0 and A1-P1

Measures of the extent of vowel nasalization between the period of 30 ms and 70 ms after the nasal landmark are shown in Figure 4.16. The nasal landmark, which indicates the release from the nasal murmur into the vowel, is not shown because it is set at 0 ms. The top panel of Figure 4.16 shows the difference measurements of A1-P0 for the low vowel /a/. The bottom panel shows the difference measurements of A1-P1 for the high vowel /i/. Measures from the production of SC speakers are indicated in red while those of the English speaker are highlighted in blue. The lower the difference measurements of A1-P0 or A1-P1, the more nasalized the vowel. For the vowel /a/, the SC speakers nasalize to a greater extent than does the English speaker. A1-P0 decreases in the SC /a/ as we move further away from the nasal landmark and to the vowel. For the vowel /i/, the SC speakers nasalize to approximately the same extent as does the English speaker.

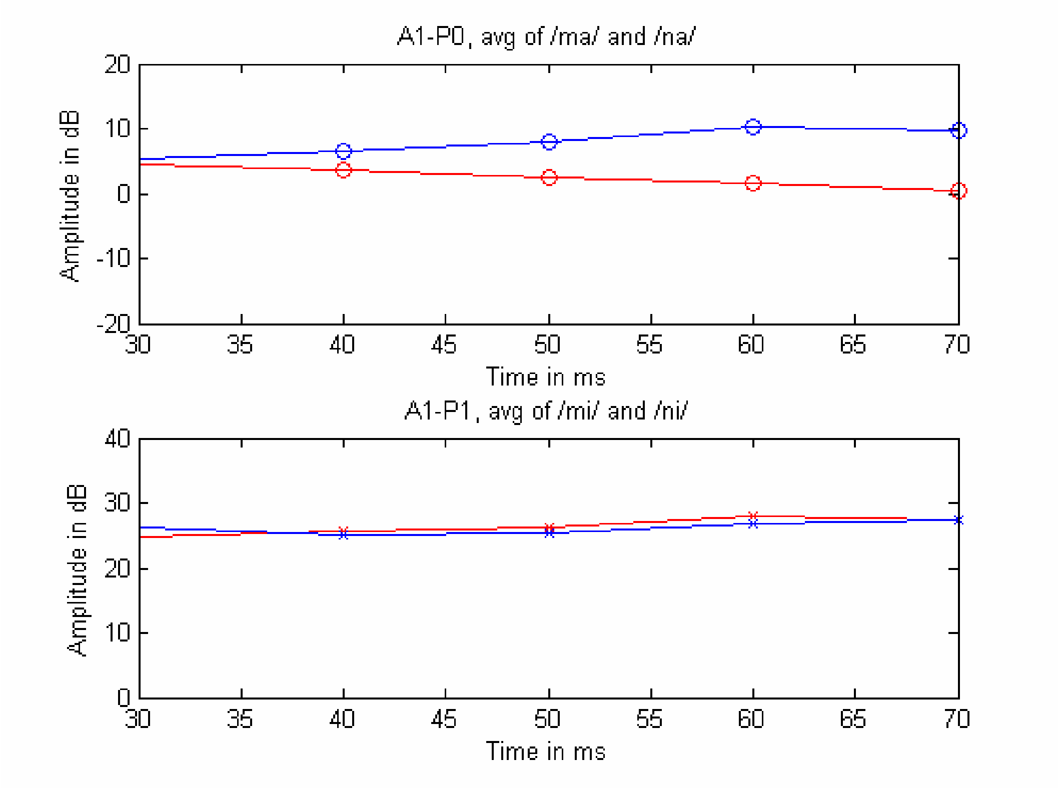

Figure 4.16 Measures of nasalization shown for /a/ (top panel) and /i/ (bottom panel), red for SC speakers, and blue for the English speaker.

The results are consistent with the prediction that SC nasals in the onset position should be produced similarly to the English nasal onset. The acoustic correlates for the place of articulation must be contained in the nasal onset as nasals are no longer the only consonants allowed in the syllable-initial position and must be distinguished from other consonants. The SC onset nasal is not a nasal approximant as it seemed to have been in the coda position; instead, it is made with an oral closure and there is evidence of a strong nasal murmur with formants that are indicative of an alveolar or a labial nasal in the acoustic signal.

4.3 Summary

In this chapter, we examined the shift in the second formant frequency of a nasalized vowel in both the CVN and NV environments, where the nasal appears in both the coda and syllable-initial positions, as an additional acoustic correlate for the contrast between the velar and alveolar nasal coda in Standard Chinese. F2 is assumed to be the acoustic correlate for the feature [back], and is directly related to the position of the tongue. Acoustic analysis of CVN syllables show that for the SC low and mid vowel, which are unspecified for the feature [back], F2 increases when the vowel is followed by an alveolar nasal and is kept low when it is followed by a velar nasal. The English speaker, however, maintains a lower F2 for the low vowel /a/ for a longer period, even when the vowel is followed by an alveolar nasal. The difference between the English speaker and SC speakers is largest in the production of the mid vowel. In this case, the English speaker produces the vowel exactly the same, regardless whether it is followed by the velar or alveolar nasal. The SC speakers, however, produce a more fronted /ə/ when it is followed by the alveolar nasal and a more backed /ə/ when it is followed by the velar nasal.

For the high vowel, F2 is specified for the feature [back] and acoustic analysis of CVN syllables show that indeed, F2 does not change, and we do not find differences in the production of /i/ in the CVN context between the English and SC speakers. F2 is therefore not an additional acoustic correlate for the contrast between the velar and alveolar nasal coda in SC. The SC speaker must be faithful to the primary acoustic correlates that contrast the place of articulation of the nasal codas, namely, producing the velar pinch when the coda is the velar /ŋ/. Measures of A1-P0 and A1-P1 also indicate that SC vowels are in general less nasalized than English vowels. This is consistent with the view that if SC vowels already shift the second formant to cue for the place of articulation of nasal codas, then there is less of a need for vowel nasalization as a place cue for the nasal coda. We have also suggested that the second nasal pole at 1 kHz may be shifted as a result of differences in the opening of the velopharyngeal port. The nasal pole in the spectra of some of the SC speakers may be above 1 kHz, and that these speakers may be producing the nasals with a more open velopharyngeal port, at least in the region of the vowel. In addition, we have found that some SC speakers may even spread the glottis, in order to enhance the perception of nasality, as spreading the glottis magnifies the effect of dampening of the first formant and widening of the first formant bandwidth.

Differences in the phonotactic constraints between the English and Standard Chinese vowels and nasal codas could potentially allow the speakers of SC to differently manipulate the opening of the velopharyngeal port, which is considered the primary articulator responsible for the perception of nasality. In addition to the primary articulator, the SC speakers may use the spreading of the glottis to enhance the perception of nasality. Most importantly, they take advantage of the unspecification for the feature [back] for SC mid and low vowels to move their tongues forward or backward to accommodate the place of articulation of the nasal coda. These different manipulations of the tongue body appear in the acoustic signal as different trajectories of the F2 in a vowel that precedes a velar or alveolar nasal coda. Whenever the phonotactic constraints do not allow for such changes in the acoustics, as in the high vowel /i/, whose feature is specified for [-back], the tongue body cannot be manipulated, and the additional acoustic correlate in the form of F2 shift does not appear in the acoustic signal, so the speakers must adhere to the universal acoustic correlates of nasalization.

In the second section of this chapter, we showed, through acoustic analysis, that the production of the English and SC vowels in the NV context, where the nasal is the labial /m/ or the alveolar /n/, are very similar to each other. The extent of vowel nasalization is also very similar. As nasals are not the only consonants allowed in the syllable-initial position in SC, their production may require a more precise effort at adhering to the universal acoustic correlates of nasalization.

We have seen that although the perception of nasality is similar in English and Standard Chinese, the interactions of the language-independent distinctive features and the language-dependent enhancing gestures could lead to differences in the acoustic manifestation of the same feature in two different languages.

Chapter 5. Vowel change as a perceptual cue for the nasal place of articulation

5.1 Introduction

Results from the perceptual experiment presented in Chapters 3 and 4 suggest that there are differences in the weighting of the acoustic correlates for the place of articulation of the nasal coda in English and Standard Chinese. In SC, more acoustic correlates of nasalization seem to be contained in the vowel region, when the vowel is a non-high vowel. The SC non-high vowels are unspecified for the feature [back], and are produced so that they are fronted when followed by an alveolar nasal and backed when followed by a velar nasal. The English non-high vowels, however, do not exhibit this kind of F2 behavior. The SC high vowel, /i/, which is specified for the feature [-back], also does not exhibit this kind of F2 behavior. Analysis of the SC NV environments shows that when the nasal appears in the syllable-initial position, the following vowel behaves very much like its English counterpart. The special status of the nasal as the only consonant allowed the in the coda position in SC, combined with the unspecification of the SC non-high vowels for the feature [back], seems to give rise to some additional acoustic correlates for the place of articulation of the nasal coda. The phonotactic constraint that is particular to the SC mid and low vowels does not apply to the SC high vowel, /i/, which is specified for the feature [-back], so the burden of contrasting the place of articulation of the nasal coda falls in the murmur region.

In this chapter, we present a perceptual experiment, which examines how SC listeners actually make use of the primary acoustic correlates associated with the universal features of nasalization and the additional acoustic correlates associated with language-specific enhancing gestures to make judgments of the place of articulation for the nasal coda. The experiment is a gating study (Grosjean, 1996, 1997, Lahiri & Marslen-Wilson, 1990) that explores to what extent nasality is perceived as listeners receive more information about a syllable presented incrementally in time.

5.2 Hypothesis

If the nasal codas induce changes in vowel quality in SC in different degrees depending on the vowel, then we would expect the vowel that undergoes the most change to signal most readily the nasal place of articulation in an earlier region. The vowel that undergoes the least change would require many more portions in time for subjects to make correct judgments of the place of nasal articulation.

We predict that early portions of the low and mid vowel, /a/ and /ə/, whose F2 trajectories vary depending on the place of articulation of the following nasal, would contain more information than early portions of the vowel /i/, whose F2 does not change much according to the place of articulation of the nasal. In particular, acoustic analysis of CVN environments in Chapter 4 show that the mid vowel /ə/ contrasts even more significantly than the low vowel /a/ the place of articulation of the nasal coda. We would therefore expect listeners to identify the place of articulation of the nasal coda most readily when the vowel is the mid vowel /ə/, to be closely followed by the low vowel /a/, and then least readily when the vowel is the high vowel /i/.

5.3 Materials and design

The stimuli consists of 6 CVN syllables “ban”, “bang”, “ben”, “beng”, “bin”, “bing”, and 2 CV syllables “ba” and “bi”. These syllables were chosen because out of all of the possible CVN combinations, “b” and “p” were the only non-nasal consonants that can occur before all three of the vowels. “p” is an unvoiced consonant and at the burst release we would expect some noise during the first gate that may interfere with perception. The CV syllables “ba”, “bi”, and “be” served as foils. Although “be” is a missing form in SC, our subjects were told to treat each stimulus as simply a sound, or nonsense syllable, and try as much as possible to not associate any meaning with it, and had the choice to pick the category of non-nasals.

Four SC speakers were recorded in a sound attenuated chamber. The same speakers were recruited for experiments in this thesis. The syllables were embedded in a carrier phrase: SC sentence “wo^{11} ʂ̡^{w}o^{55} ___ tʂ̡ɤ^{51} kɤ^{11} tsɨ^{51} ‘I say ___ this word’.” The utterances were digitized at a sampling rate of 16 KHz. The recordings were converted into two Microsoft wave files and the individual words were extracted using Praat. Each sound was gated from the onset of the ‘b’ burst. The first gate was set at the beginning of the ‘b’ burst release and it was 40 ms in duration. The gating sequence continued through the vowel across the nasal landmark in 40 ms increments toward the end of the syllable. Each gate ends at the waveform zero-crossing just before the initial rise at the beginning of the pitch period. This setup is shown in Figure 5.1.

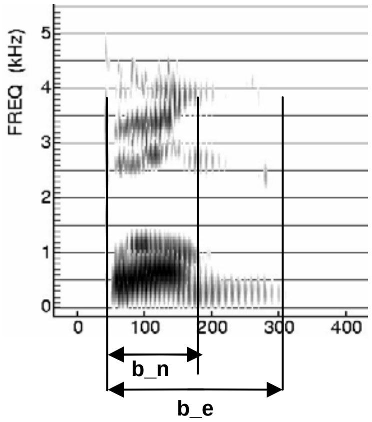

Figure 5.1 Spectrogram of /beŋ/, produced by the SC speaker XS. The duration from the burst release to the nasal landmark is denoted by b_n, the duration from the burst release to the end of the utterance is denoted by b_e.

The gating experiment was limited to the first 200 ms from the burst release, or the equivalent of 5 gates. All of the landmarks occur within the first 200 ms of the “b” burst release, as shown in Table 5.1. Beyond the landmark, the listener should have unambiguous information about the final consonant.

The first number in each cell of Table 5.1 indicates the time in ms from the release of the “b” burst to the nasal landmark, and this is duration is indicated in Figure 5.1 by b_n. The second number in Table 5.1 is a ratio of the duration from the burst to the nasal landmark, b-n, to the duration from the burst to the end of the utterance, b_e.

Table 5.1 Duration from burst release to nasal landmark, b_n, and ratio of duration from burst release to nasal landmark and of duration from burst release to end of utterance, (b_n/b_e) by syllable and speaker.

SC syllables	Speaker JZ	Speaker WC	Speaker YP	Speaker XS
ban	105 (.40)	188 **(.83)**	135 **(.81)**	170 (.63)
baŋ	129 (.48)	129 (.53)	123 (.48)	153 (.61)
ben	143 (.54)	87 (.45)	82 (.43)	130 (.53)
beŋ	97 (.37)	77 (.37)	86 (.42)	109 (.44)
bin	87 (.32)	96 (.39)	98 (.42)	141 (.60)
biŋ	128 (.46)	84 (.34)	90 (.38)	200 **(.80)**

The three numbers that are shown in bold in Table 5.1, the ratios of b_n to b_e, indicate that the corresponding syllables, WC's /ban/, YP's /ban/ and XS' /biŋ/, were produced with nasal landmarks that appear well beyond the landmarks of the other sounds. These syllables have been eliminated from the analysis because we want to control for the point at which the nasal landmark appears relative to the duration of the entire utterance. We want to be fairly confident that the acoustic information available at each successive gate is not lacking as a result of a nasal landmark that appears much later in the utterance or do not appear even in the latest gate. We want to eliminate the possibility that listeners might categorize the nasal stimuli as non-nasals because the landmark appears relatively late in the syllable.

5.4 Stimuli and task

The stimuli consisted of six CVN syllables gated 5 times and two CV syllables gated 3 times; yielding 36 syllables. There were 4 speakers and 6 repetitions, yielding a total of

864 stimuli. Each stimulus was presented twice, with a 75 ms pause in between, through headphones connected to a PC to the subjects.

Ten native speakers of Standard Chinese (five female and five male) were paid for their participation. In each trial of this experiment, they heard a cutout portion of a SC syllable X that is either a CVN or CV. They were asked to try as much as possible to pay attention to the sound, and not to associate it with any possible meaning. Their task was to categorize it in one of three groups: “_n”, “_ng”, or “_”, which indicates the absence of a nasal perceived. After they made their decision, they were not able to go back. There were ten trials before the start of the experiment.

5.5 Results

5.5.1 Correct identification of the nasal place of articulation

Figure 5.2 shows the averaged responses from the identification task by the ten listeners. Results from the identification task corresponding to the production of /ban/ by the speakers YP and WC, and of /biŋ/ by the speaker XS are not included in the following discussion. In place of the responses to these stimuli, a red X is marked in the graphs. Acoustic analysis from chapter 4 showed that there is speaker variability in the production of the nasal murmur and differences in the location of the nasal pole in the acoustics of the vowel preceding the nasal murmur. Under the assumption that production is acoustically and perceptually driven, differences in production could lead to differences in subject responses. The results from the identification task are therefore broken down by speakers.

The results are displayed for each of the six identification tasks: an_an, ang_ang, en_en, eng_eng, in_in, and ing_ing. Within each identification task, the responses are shown for each of the five successive gates. The responses indicate how often the subjects correctly judged the nasal coda based on the limited information contained in each successive gate. For instance, listeners responded with the correct place of articulation for /ben/ nearly 100% of the time, beginning at the 3rd gate at 80 ms, for all 4 speakers. In particular, listeners correctly identified the place of articulation for the nasal coda when the stimulus containing the mid vowel was produced by the speaker YP, as early as the 2nd gate, with nearly 90% rate. The information contained in the first 80 ms of a stimulus containing the mid vowel is already enough for listeners to correctly infer the rest of the syllable.

Figure 5.2 shows that subjects identified the place of articulation for the nasal coda most correctly when the stimulus contains the mid vowel, /ə/, followed by the low vowel, /a/. Subjects were poorest at identifying the place of articulation when the nasal coda is preceded by the high vowel /i/. The identification rate involving a high vowel was best with speaker YP, but even this does not exceed 90%.

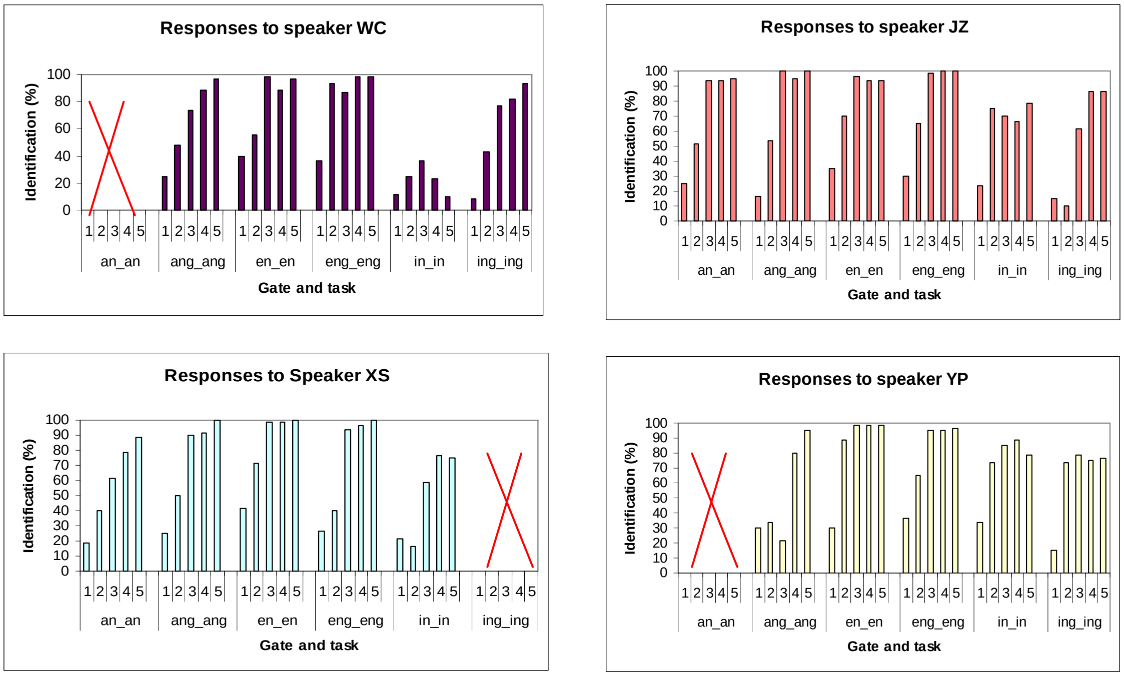

Figure 5.2 Speaker-specific correct identification of the place of nasal articulation with incremental gates.

The identification rate was lowest when the stimulus containing the high vowel was produced by the speaker WC (upper left panel). Even by the 3rd gate, listeners identified the actual stimuli “in” as “in” only 35% of the time. The subjects seem to identify even less consistently with increasing acoustic information as the gates were lengthened. Since the task involves categorizing the stimuli into three groups, “_n”, “_ng”, or “nil”, there are two possible incorrect choices for every correct choice. It is important to distinguish the response that mistakes a nasal stimulus as a non-nasal and the response that mistakes the nasal place of articulation. For the stimuli containing the high vowel produced by speaker WC, we are interested in the incorrect choices the listeners made when the actual stimulus was “in”.

5.5.2 Incorrect identification of the place of nasal articulation

Figure 5.3 shows how often the subjects incorrectly identified the place of articulation of a nasal coda. For stimuli containing an alveolar nasal, the subjects reported hearing a velar nasal and for stimuli containing a velar nasal, the subjects reported an alveolar nasal. Listeners especially confused WC’s production of “bin” (shown in the upper left panel) with “biŋ”. Even at the final gate, confusion was 90%.

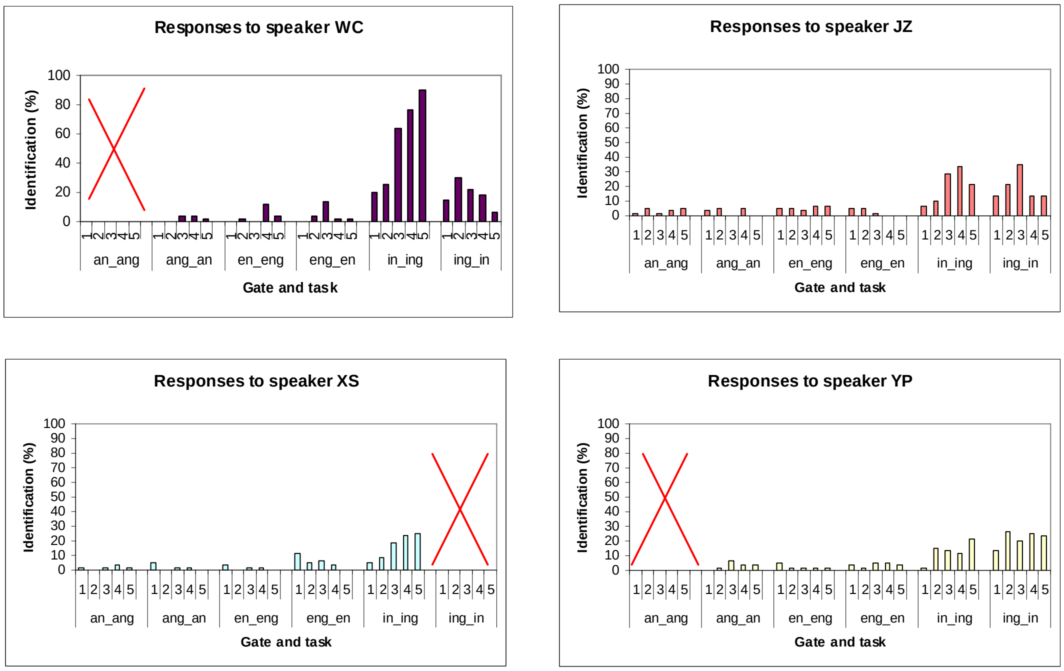

Figure 5.3 Speaker-specific incorrect identification of the place of nasal articulation with incremental gates.

Figure 5.3 shows that over all, all subjects had some trouble identifying the place of articulation of the nasal coda when the coda is preceded by the high vowel, /i/, regardless of which of the four speakers produced it. Listeners were also likely to report an alveolar nasal when the stimulus contained a velar nasal. They are likely to confuse “biŋ” with “bin”.

5.5.3 Incorrect identification of nasals for non-nasals

Figure 5.4 shows how often the subjects reported a non-nasal syllable when the stimulus contained a nasal coda. For the nasal stimuli containing the low vowel /a/, listeners most frequently judged it as non-nasal, even up to the 3rd gate. Beginning at the 4th gate, however, subjects were able to correctly identify the place of articulation. That is, once the subjects decided the stimulus contained a nasal coda, they had little trouble distinguishing its place of articulation. On the other hand, for the nasal stimuli containing the high vowel /i/, subjects were able to identify correctly the sound as nasal, even though they often incorrectly identified the place of the nasal articulation.

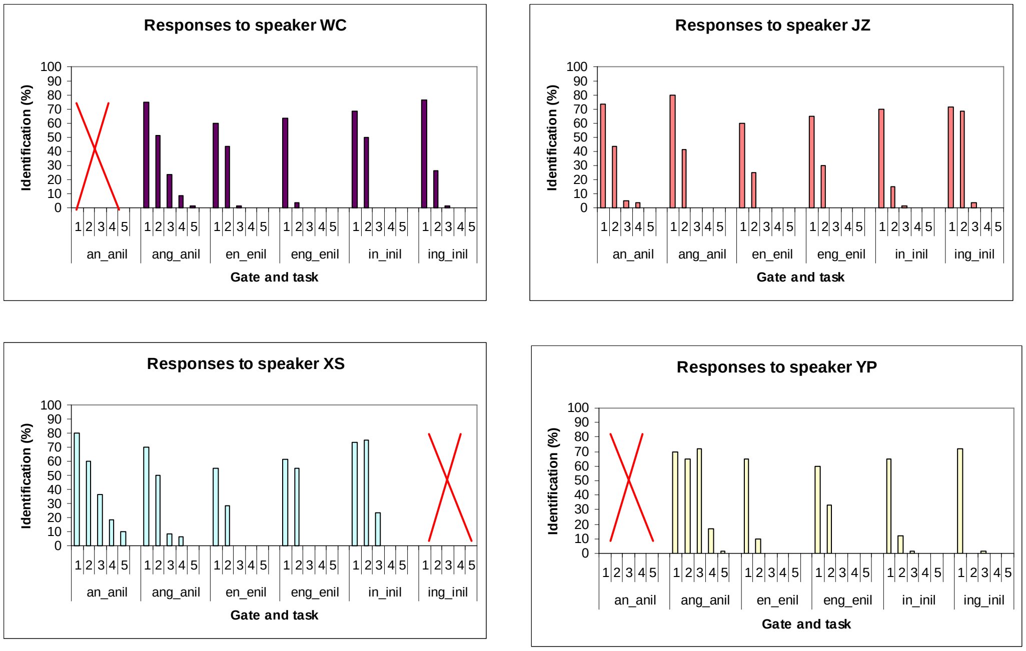

Figure 5.4 Speaker-specific incorrect identification of nasals for non-nasals with incremental gates.

5.6 Acoustic analysis

This section presents some acoustic evidence for the choices that the subjects made in the gating experiment. The speakers are able to produce the low and mid vowels by fronting or backing the tongue body, depending on the nasal coda that follow them. The spectrograms in Figure 5.5 are shown for /a/ and /ə/ in a CVN environment, where the consonant is /b/ and the nasal either an alveolar [n], or velar [ŋ]. The syllables are spoken by one male SC speaker. The top panels show the spectrograms for [ban] and [bɑŋ], and the bottom panels show the spectrograms for [ben] and [beŋ]. The vowels become more fronted next to [n], manifested in a more separated F1 and F2, and more backed next to [ŋ], manifested in the coming together of F1 and F2.

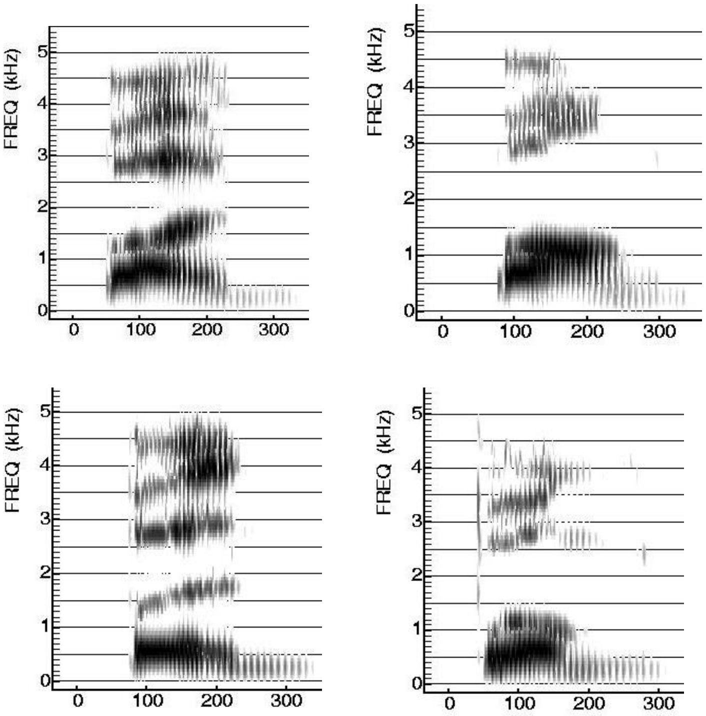

Figure 5.5. Spectrograms for the low and mid vowels produced by a SC male speaker (XS). Top panels show the spectrograms for [ban] (left) and [bɑŋ] (right). Bottom panels show the spectrograms for [ben] (left) and [beŋ] (right).

The subjects are able to make use of the additional acoustic cue provided by the transition of the second formant frequency to make judgments of place of articulation.

On the other hand, the high vowel is specified for the feature [-back] and the speakers cannot change the position of the tongue body to accommodate the place of articulation of the following nasal. Figure 5.6 shows the spectrogram of [bin] and [biŋ], as produced by a female SC speaker. In the spectrogram for [biŋ], between 265ms and 285 ms, there is evidence of a higher F1 during the vowel just before the vowel-nasal landmark, which occurs at around 315ms. The inserted vowel has a F1 that is higher than the F1 of /i/ (300 Hz) and it is at about 500 Hz, at the frequency that we would expect of a schwa. As discussed in Chapter 3, inserting a schwa in [biŋ] so that it becomes [bi$^{\varepsilon}$ŋ] might be a strategy that a speaker uses to help contrast it from [bin].

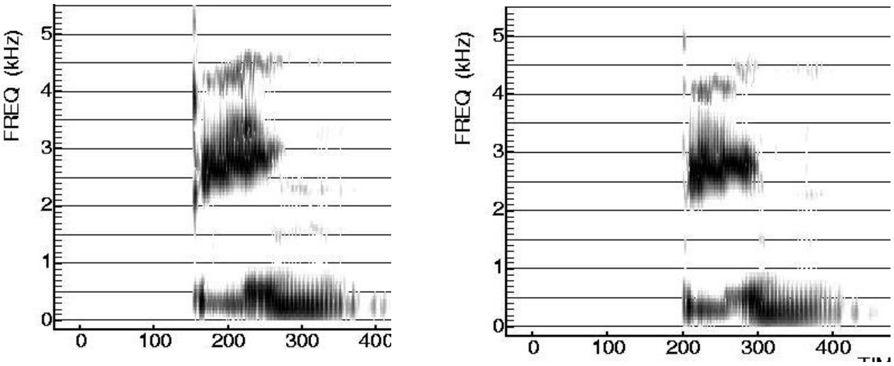

Figure 5.6 Spectrograms of a female SC speaker's (WC) production of [bin] (left) and [biŋ] (right). Possible evidence of schwa insertion in [biŋ] at about 265 ms to 285 ms.

Despite this production strategy, however, the acoustics of the two syllables appear to be very similar and there is no significant difference in the first three formants in the vowel region to cue for the place of articulation of the nasal coda.

5.7 Summary

The results of the gating experiment show that the subjects were able to identify correctly the place of the articulation of the nasal coda, as early as the 3rd gate, in particular when the stimulus contained the mid vowel. Given a stimulus containing the mid or low vowel, once the subjects decide that the stimulus contains a nasal coda, they are able to make the correct judgment of the place of nasal articulation. On the other hand, when the stimulus contains a high vowel, even if the subjects detect a nasal coda, they have trouble identifying the nasal place of articulation.

These results seem to support the findings from section 4.1.3, which report a higher extent of vowel nasalization in the high vowel /i/, than in the low vowel /a/ and mid vowel /ə/. The speakers are able to manipulate the tongue body position during the production of the low or mid vowel, depending on the place of articulation of the nasal coda. Changes in the tongue body position during the vowel thus become an important enhancing gesture for contrasting the nasal place of articulation. Listeners perceive the corresponding changes in the acoustics of the vowel, namely F2, and use its trajectory to help distinguish the place of articulation of the nasal coda, even when they have not been given the entire nasal utterance. It seems that the shift in F2 alone in the mid and low vowels can be responsible for the perception of nasality and that the direction of that shift can help determine the place of articulation for the nasal coda. We have suggested in the previous chapter that the mid and low vowels do not need to be as nasalized as the high vowel. The high vowel is specified for the feature [back] and its

F2 cannot shift to signal a contrast between the places of articulation for the nasal coda. Without this additional acoustic correlate to utilize, a speaker must produce the vowel /i/ with more nasalization than the mid and low vowels if it precedes a nasal coda. Furthermore, even when a listener correctly perceives this utterance as a nasal sound, the place of articulation of the nasal coda can still be highly confused.

Overall, the subjects were best at identifying the place of articulation of the nasal coda when the stimuli contained the mid vowel /ə/, followed by the low vowel /a/, and were worst at identifying the place of articulation when the stimuli contained the high vowel /i/. The results of the gating experiment confirm the hypothesis that the listeners are able to identify most correctly the place of articulation of the nasal coda when the stimulus contained the vowels that are able to carry the weight of the acoustic correlates of the place of articulation for the nasal coda. These vowels are the mid and the low vowels. When the stimulus contained the high vowel, which does not carry much of the weight of the correlates for the place of articulation of the nasal coda, listeners had trouble in the identification task.

Chapter 6. Summary

The goal of this thesis is to quantify the language-independent acoustic correlates for the distinctive feature [nasal] and the language-dependent enhancing gestures that stem from the influence of the nasal feature on the adjacent vowel, which may help to contrast the place of articulation for nasal codas in SC. We chose to study the nasal in SC as it occupies a special position in the language, which can lead to differences in the acoustic manifestation of the place cues for the nasal. The nasal coda is the only consonant allowed in the coda position in SC, so it has no other consonants to compete with for perceptual saliency. The need for contrast lies only within the nasal codas itself, that is, whether the coda is an alveolar nasal vs. velar nasal. The production of the nasal coda often does not involve complete closure in the oral tract and thus the primary gestures that are associated with the place cues for the nasal are not necessarily made. In this environment, the perceptual saliency might be at risk in the nasal murmur region, because the incomplete articulatory gestures would not give sufficient information for the recognition of the nasal coda. We might expect some compensatory acoustics earlier in the syllable in the vowel region that can provide information about the identity of the nasal coda. As a result, much of the discussion in the thesis focuses on the enhancement for the place of articulation of the nasal codas, and examines the acoustics near the nasal landmark and in the adjacent vowel. The SC listeners are capable of making the [an] vs. [ɑŋ] distinction, so the question becomes where, when, and how? That is, at what point in time do they recognize the nasal coda? Where in the syllable do these acoustic cues reside, in particular, the vowel region or the nasal murmur region? How do the listeners make use of the acoustic cues which reflect the enhancing gestures to differentiate the nasal coda?

In order to answer these questions, the research in this thesis is organized into four main sections. The first section focuses on quantifying the acoustic correlates and articulatory gestures of the nasal through acoustic modeling, and establishing the defining acoustic and articulatory attributes of the nasal consonant. The second section observes for acoustic patterns that are particular to the production of nasal consonants in Standard Chinese. The third section obtains acoustic data that compare nasal production in English and SC. The fourth section examines the time course of perception of the nasal coda in SC. The overall goal is to interpret all of these findings in terms of the distinctive feature theory and the theory of enhancement.

We began with a simple model of vowel nasalization based on the one-dimensional wave propagation theory and varied the parameter thought to be the primary acoustic correlate for the feature [nasal], namely the area of the velopharyngeal opening, and observed the resulting acoustics as this articulator undergoes a continuous movement. The parametric model of nasalization highlighted a region in the spectrum within which to search for the presence of the nasal pole, thought to be the primary acoustic correlate of the feature [nasal]. Furthermore, the separation of the nasal pole from the nasal zero, which is reflected in the amplitude difference between the spectrum prominence and minimum for the nasal pole and zero, shows a quantal relation with the opening of the velopharyngeal port. We found a region within which small changes in the cross-sectional area of the velopharyngeal port (between 0.1 cm^2 to 0.5 cm^2) lead to large changes in the acoustics. There are also two other regions on either side of this region of rapid acoustic change: when the area of the velopharyngeal opening is small (between 0 and 0.1 cm^2), the pole and zero cancel initially and they begin to separate as the opening becomes larger; when the area of the veloparyngeal port is increased to beyond 0.5 cm^2, the acoustic attribute is stable again and there is not much movement in the acoustics in response to an even larger velopharyngeal opening. The three regions together describe a nonlinear relationship between the articulator and the acoustics, which gives rise to the quantal nature of nasality.

The observation of the acoustic patterns in SC focused on the differences in the distribution of acoustic cues for the nasal place of articulation between English and Chinese nasal codas. Results from the mapping experiment involving the low vowel /a/ showed that more cues for the place of articulation for the nasal coda were contained in the vowel transition region in the Chinese syllables with nasal codas than in the English syllables with nasal codas. This suggested that vowel modification in Chinese might be an enhancing gesture for contrasting the nasal place of articulation.

Further acoustic analysis of the distribution of the acoustic correlates of nasalization in VN environments focused on the high vowel /i/ and the non-high vowels /ə/ and /a/ as well as nasals in the syllable-initial position. Results suggest that speakers take advantage of the unspecification for the feature [back] of SC mid and low vowels by moving their tongues forward or backward to accommodate the place of articulation of the nasal coda. These different manipulations of the tongue body appear in the acoustic signal as different trajectories of the F2 in a vowel that precedes a velar or alveolar nasal coda. Whenever the phonotactic constraints do not allow for such changes in the acoustics, as in the high vowel /i/, which is specified for [-back], the speakers cannot manipulate their tongue bodies and must instead fall back onto the universal acoustic correlates of nasalization to contrast the nasal codas. In addition, the location of the nasal poles in the spectra is found to vary by speaker, indicating that speakers may control the size of the velopharyngeal opening differently. The SC speakers may also be spreading their glottis during the production of the nasal utterances, which leads to a dampening of the first formant, and thereby further enhancing the perception of nasality.

Acoustic analysis of the production of the English and SC vowels in the NV context, where the nasal is the labial /m/ or the alveolar /n/, showed that the F2 movements during the vowel are very similar to each other. What distinguishes /m/ from /n/ lies mostly in the murmur regions, in the differences between the movement of the formants as well as in the duration of the nasal murmur itself. A flat F2 in the high vowel indicates a tongue body position that is kept more constant from nasal onset to the end of the utterance than for the non-high vowels. A decreasing F2 in the low vowel indicates that the tongue body position is somewhat adjusted to accommodate the place of articulation of the nasal onset. Both patterns are observed in both SC and English. As nasals are not the only consonants allowed in the syllable-initial position in SC, their production is thought to require a more precise effort to adhere to the universal acoustic correlates of nasalization.

Finally, the perceptual experiment sought to correlate nasal production with nasal perception by investigating how SC subjects make use of the acoustic attributes and enhancing gestures for contrasting the nasal place of articulation to identify the place of articulation of nasal codas. The results confirm that the change in the tongue body position during the vowel can be an important enhancing gesture to increase the perceptual salience of the contrast between the places of articulation. Listeners are able to identify most correctly the place of articulation of the nasal coda when the stimulus contained the vowels that are able to carry the weight of the acoustic correlates of the place of articulation for the nasal coda. These vowels are the mid and the low vowels. When the stimulus contained the high vowel, which does not carry much of the weight of the correlates for the place of articulation of the nasal coda, listeners had trouble in the identification task.

The movement of F2 in the mid and low vowels seems to be responsible for the perception of nasality and the direction of that shift seems to determine the place of articulation for the nasal coda. In the case of the high vowel /i/, the listeners do not have at their disposal the additional acoustic correlate of F2 shift to contrast the place of articulation for the nasal coda. Even when a listener correctly perceives an utterance as a nasal sound, the place of articulation of the nasal coda is often highly confused.

The phenomenon of modification of the vowel acoustics to accommodate the place of articulation of the nasal coda that we observed in Standard Chinese is not unique to SC. Studies of Marshallese (Choi, 1992, 1995), which has a small vowel inventory like that of SC, have shown that vowel modification also occurs. The mid, low, and high vowels of Marshallese are all unspecified for the feature [back]. The places of the articulation of the neighboring consonants determine the tongue body position during the vowel production.

Within SC, vowel modification, as used to accommodate the place of articulation of the neighboring consonant, is also not unique to the nasal coda. It is also observed in vowels following fricatives. There is a three way acoustic distinction for the Standard Chinese strident fricatives: the alveolar [s], the flat postalveolar [ʃ], and the palatalized postalveolar [ɕ] (Ladefoged and Maddieson, 1996). Stevens, Li, Lee & Keyser (2004) studied the distinctive features that characterize the place distinction of the SC fricatives and how the place contrast is manifested acoustically. Their study shows that the main difference among the three sounds lie in the shape of the channel between the tongue and the roof of the mouth. The alveolar is produced with the tip of the tongue forming the constriction. The flat postalveolar is produced with the upper surface of the tongue tip. The palatalized postaveolar is produced with the tongue body. The point of maximal constriction for [ɕ] is in between that for [s] and [ʃ]. The vowel [i] occurs only after the palatalized post-alveolar fricative. After the alveolar fricative, the vowel is produced with a more backed tongue body; and after the flat post-alveolar fricative, the vowel is retroflexed.

These three phonetically distinct vowels are considered allophones. Under the view that the implementation of distinctive features may be enhanced by non-contrastive articulatory gestures to maintain perceptual distinctiveness (Keyser and Stevens, 2003), they proposed that the three vowels are derived from an underlying [i] and are modified to enhance the defining articulatory and acoustic attributes of the strident fricatives. They found that the articulatory and acoustic properties of the palatalized fricative is most compatible with those of [i], whereas the tongue body backing gesture in the alveolar and palatoalveolar fricatives are consistent with lowered F2 frequency in the apical and retroflex vowels. Acoustic analysis also showed that the starting frequency and the end frequency of the vowel [i] in [ʃi] and [ɕi] are very similar, while they differ by nearly 700 Hz when the vowel is [a]. When the fricatives are followed by [i], they are produced by slightly lowering the tongue tip, without further adjustments to the tongue body position during the production of the vowel. When the fricatives are followed by [a], they are produced with a lower tongue body.

In our analysis of NV utterances in Chapter 4, we also observed that the tongue body position is held much more constant throughout the vowel /i/ than it is in the vowel /a/, which varies depending on the place of articulation of the nasal coda. This may explain why an absolute value of F2 in [i] is not sufficient to help determine the place of the articulation of the preceding consonant, while a relative value of F2 in [a] or [ə] , or the movement of F2 throughout the vowel, can help determine the place of articulation of the preceding consonant.

The F2 transition during the SC vowel preceding a nasal coda is an important additional acoustic correlate for the contrast between the places of articulation of the codas when the vowel is unspecified for the feature [back]. The F2 transition is an acoustic manifestation of the tongue body manipulation; it is lower for a more backed vowel and high for a more fronted vowel. The fronting and backing of the tongue body position to accommodate the place of articulation of the nasal coda can be considered an enhancing gesture.

This enhancing gesture sometimes takes precedence over the primary acoustic correlates for the place of articulation. The gating experiment results show that the transition of the F2 in these SC vowels alone can signal for the presence of a nasal coda and the direction of that transition can help contrast the place of its articulation. When this F2 enhancement is not available in the high vowel /i/, the primary acoustic correlates for the place of articulation are preserved. Speakers tend to make the appropriate constriction in the nasal murmur to help contrast the place of articulation between the velar and alveolar nasal. The nasal poles in the acoustics are manifestations of the locations of the oral closures.

Standard Chinese has been the focus of this thesis and the experimental results seem to suggest that we may expect similar modifications on the vowel by the place of articulation of the adjacent consonant in other languages with small vowel inventories, in particular, where the vowels are unspecified for certain features. This kind of conclusion would certainly require further work in other languages with small vowel inventories. We might begin by examining a similar kind of consonant-induced vowel modification in Cantonese, a language that is related to SC, which contains a few more vowels and tones and allows /p/, /t/, /k/, and /m/ in the coda position, in addition to the velar and alveolar nasals. For example, since /t/ constriction is made at the same place as the alveolar /n/, we would expect the non-high vowels preceding /t/ to be produced with a more fronted tongue body. Since the /k/ constriction is made at the same place as the velar /ŋ/, we would expect the non-high vowels preceding /k/ to be produced with a more backed tongue body.

Before we make such further studies, it would also be important to conduct an additional experiment in SC where listeners are asked to make judgments of the nasal place of articulation with synthetic speech stimuli. The naturally produced non-high vowels can be adjusted in their F2 formant movements to simulate the more fronted and backed tongue body position that we observed to be helpful in enhancing the place of articulation of the nasal coda.

The distinctive feature theory and the theory of enhancement are useful in providing a possible explanation for the patterns of articulation and acoustics that are observed for the nasal coda in Standard Chinese and English. Enhancement theory assumes that there are defining acoustic and articulatory correlates of distinctive features and that enhancing gestures may be called upon when there is a need to preserve or increase the perceptual saliency of the contrasts defined by the features. In the distinctive feature theory, the vowels and nasals are specified in terms of their primary articulatory gestures and acoustic correlates. In SC, the non-high vowels are unspecified for the feature [back] while the high vowel is specified for the feature [back]. The distinctive feature theory and the theory of enhancement would predict that the tongue body can move to a more fronted or backed positions during the production of the non-high vowels, depending on the place of articulation of the nasal coda. In addition, the theories would also predict that the tongue body cannot move as freely during the production of the high vowel, regardless of the place of articulation of the nasal coda. The theories would also predict that in a time-course investigation of the recognition of the nasal place, listeners would make use of the enhancing gestures in the tongue body movement of the non-high vowels to recognize the nasal coda. When the tongue body cannot move in the case of the high vowel, recognition occurs later into the gating experiment, and at any given time, recognition is poorer compared to the non-high vowels. We would have been very surprised, and indeed the theories would have been falsified, if the results showed the opposite trend: the F2 movements in the high vowel are very different when it is followed by the velar vs. alveolar nasal, or that perception of nasal place of articulation was more accurate when the preceding vowel was the high vowel than when it was the non-high vowel. Indeed, all the observations and results of perceptual experiments are consistent with the predictions of the distinctive feature theory and the theory of enhancement.

At the basis of the distinctive feature theory is the quantal theory of speech, which we discussed in Chapter 1. The quantal nature of speech is grounded in the observation that the relationship between the articulatory gestures and the acoustic patterns is non-monotonic. The quantal theory (QT) gives us a way to represent each sound with a combination of discrete units, or features, that are specified in terms of associated articulatory gestures. It is based on the physiology of sound production and acoustic modeling of the vocal tract configurations. Would the results we have obtained still make sense if the studies were conducted in a framework other than the distinctive theory? Another theory that provides an account for the vowel systems of the world's languages is Lindblom's Theory of Adaptive Dispersion (TAD) (Liljencrans & Lindblom, 1972, Lindblom, 2003), where the distinctive sounds of a language are positioned in the phonetic space so as to maximize perceptual contrast, and optimal vowel inventories of different sizes have been predicted. However, the TAD does not specify features, or the direction and distance of the dispersion each vowel can take. If this thesis work were to be carried out within the framework of TAD, the direction and the distance of the dispersion for neighboring vowels would have to be specified in order to make sense of the differences in the productions of the high and non-high vowels, when they precede the nasal coda. That is, the distance the high vowel is allowed in the vowel space should be much shorter than that allowed for the non-high vowels.

The QT gives a phonetic basis for the distinctive features, in terms of articulatory and aerodynamic processes in the vocal tract. The distinctive feature theory and the theory of enhancement together give an articulatory and physiological explanation of how neighboring sounds might influence each other. The findings in this thesis also raise the possibility that this type of interaction might provide a way of looking at context-dependent changes observed in other features in other languages. These interactions may explain the introduction of allophones based on patterns of the articulatory and aerodynamic processes. Linguistic theories such as the Optimality Theory and the rule-based approach to phonological systems derive allophones from an underlying representation, through a series of constraints and rules that capture the systematic regularities in sound-combinations, which have motivated the development of phonological rules in the first place. Can the surface form be thought of as a result of the interaction of the distinctive features and the enhancement process that are grounded in physiology? Can these theories, which seem to be based on different entities, rules and constraints vs. articulatory processes, be reconciled to help explain the same phenomenon, though from different perspectives? Regardless, the results of this thesis work suggest that language-specific constraints play an important role in determining the enhancing attributes for language-universal features and that the interactions of the distinctive features and the enhancing gestures could lead to differences in the acoustic manifestation of the same feature in different languages.

Appendix

A. Results of t-test The significance of F1, F2, F3 at end-point of the vocalic portion in SC vn/vŋ pairs (from Lin & Yan, 1991).

vn/vŋ	formants	T value	Significance level
an/aŋ	F1	t(86)=-5.071	P<0.001
	F2	t(86)=22.708	P<0.001
	F3	t(86)=-0.349	P>0.05
ən/ ŋə	F1	t(43)=-0.888	P>0.05
	F2	t(43)=20.624	P<0.001
	F3	t(43)=-4.295	P>0.05
in/iŋ	F1	t(115)=-8.676	P<0.001
	F2	t(115)=6.703	P<0.001
	F3	t(115)=10.008	P<0.001

B. SC Syllables used in the acoustic analysis and perceptual tests spelled in Pinyin. The first eight syllables were used in the perceptual gating experiment.

ban, bang, ben, beng, bin, bing, ba, bi, be, bu, pan, pang, pen, peng, pin, ping, man, mang, men, meng, min, ming, nan, nang, nen, neng, nin, ning, dan, dang, gan, gang, san, sang, wan, wang, shan, shang, an, ang, fan, fang, guan, guang, xin, xing, lin, ling, jin, jing, in, ing, qin, qing, han, hang, lan, lang, zan, zang, tan, tang, en, deng, hen, heng, ken, keng, leng, sen, seng, ten, teng, zen, zeng, ding, ting, dong, gong, hong, kong, kan, kang, gen, geng, xun, shun, jun, zun, da, di, du, ma, mi, mu, la, li, lu, na, ni, nu.

References

Broselow, E. (2003). Language contact phonology: richness of the stimulus, poverty of the base. NELS 34.

Chen, M. (1997). Acoustic correlates of English and French nasalized vowels. J. Acoust. Soc. Am., 102 (4).

Chen, M. (2000). Acoustic analysis of simple vowels preceding a nasal in Standard Chinese. Journal of phonetics, 28, 43-67.

Choi, J.D. (1995). An acoustic-phonetic underspecification account of Marshallese vowel allophony, Journal of Phonetics, 23, 323-327.

Choi, J.D. (1992). Phonetic underspecification and target-interpolation: An acoustic study of Marshallese. vowel allophony. UCLA Working Papers in Phonetics, 82.

Duanmu, S. (2002). The phonology of standard Chinese. Oxford University Press.

Fujimura, O. (1962). Analysis of nasal consonants. J. Acoust. Soc. Am., 34 (12), 1865-1875.

Fujimura, O. & Lindqvist, J. (1971). Sweep-tone measurements of vocal-tract characteristics, Journal of the Acoustical Society of America. 49 (2), 541-558.

Grosjean, F. (1996, 1997). Gating. Language and Cognitive Processes, 1996, 11 (6), 597-604.

Halle, M (2002). From memory to speech and back. Papers on Phonetics and Phonology 1954-2002. Edited by Lahiri, A. Mouton de Gruyter. New York.

Halle, M. Stevens, K. N. (1991). Knowledge of language and the sounds of speech. Proceedings on the International symposium on music, language, speech and hearing, held at the Wenner-Gren Center, Stockholm, Sweden, Sept. 5-8, 1990. Sundberg, Nord and Carlson (ed), 1-19.

Harrington, J. (1994). The contribution of the murmur and vowel to the place of articulation distinction in nasal consonants. J. Acoust. Soc. Am., 96, 19-32.

House, A. S., and Stevens, K.N. (1956). “Analog studies of the nasalization of vowels,” J. Speech Hear. Dis, 21, 218-232.

Hsieh, F., Kenstowicz, M., Mou, X. (2005). Mandarin adaptations of coda nasals in English loanwords. Submitted to Lingua. Currently under review.

Johnson, K., Ladefoged, P., Lindau M. (1993). Individual differences in vowel production**.** J. Acoust. Soc. Am., 94 (2), 701-714.

Kenstowicz, M., Suchato, A. (2004). Issues in loanword adaptation: a case study from Thai. MIT.

Klatt, D.H. (1980). Software for a cascade/parallel formant synthesizer. J. Acoust. Soc. Am., 67, 971-995.

Krakow R., Huffman M.K. (1993). Nasals, nasalization, and the velum. Academic Press Inc.

Kurowski, K., and Blumstein, S.E. (1987). Acoustic properties for place of articulation in nasal consonants. J. Acoust. Soc. Am., 81, 1917-1927

Lahiri, A., Marslen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. Cognition, 38 (3), 245-294.

Lee, C., Stevens, K. N. (2006). Strident fricatives in Mandarin Chinese: from acoustics to articulation and features. Submitted to Phonetica.

Liljencrants & Lindblom. (1972) Numerical simulation of vowel quality systems: the role of perceptual contrast, Language, 48, 839-862.

Lin, M.C. & Yan, J. Z. (1991) Coarticulation in the zero-initial syllable with nasal ending in Standard Chinese. Report of Phonetic Research, Institute of Linguistics, (CASS), Beijing, China, pp. 68-86.

Lindblom, B., (1983). Economy of speech gestures. In The Production of Speech, edited by P. F. MacNeilage. Springer-Verlag, NY, 213-245

Lindblom, B., (1990). Models of phonetic variation and selection. PERILUS (Phonetic Experimental Research, Institute of Linguistics, University of Stockholm) 11, 65-100.

Lindblom, B., (2003). Patterns of phonetic contrast: towards a unified explanatory framework, 15th ICPhS, Barcelona, 39-42.

Manuel, S. Y. (1990). The role of contrast in limiting vowel-to-vowel coarticulation in different languages. Haskins Laboratories status on speech research. SR-103/104, 1-20.

Mermelstein, P.(1977). On detecting nasals in continuous speech. J. Acoust. Soc. Am., 61 (2), 581-587.

Recasens, D. (1983). "Place cues for nasal consonants with special reference to Catalan," J. Acoust. Soc. Am., 73, 1346–1353

Repp, B.H., and Svastikula, K. (1988) Perception of the [m]-[n] distinction in VC syllables. J. Acoust. Soc. Am., 83, 237-247.

Sonderegger, M., and Chi, X. (2004). Subglottal coupling and vowel space. J. Acoust. Soc. Am., 115, pg 2540.

Stevens, K.N. (1972). The Quantal nature of speech: evidence from articulatory-acoustic Data. In E.E. David, Jr. and P.B. Denes, (eds), Human communication: a unified view, Inter-University electronics series, 15 (3), 51-66. McGraw-Hill Book Company, New York (NY).

Stevens, K.N. (1989). On the quantal nature of speech. Journal of Phonetics. 17, 3-46.

Stevens, K.N. (1998). Acoustic phonetics. Boston. The MIT Press.

Stevens, K.N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. J. Acoust. Soc. Am., 111, 1872-1891.

Stevens, K. N., Li, Z., Lee, C., Keyser. S. (2004). A note on Mandarin fricatives and enhancement. From traditional phonology to modern speech processing. Fant, G., Fujisaki, H., Cao, J., Xu, Y. (eds).

Wang, J.Z. (1993). The geometry of segmental features in Beijing Mandarin. Ph.D. Dissertation. The University of Delaware, Newark.

Warren, D.W. (1976). Aerodynamics of speech production. Lass, N.J.(Ed.), Contemporary Issues in Experimental Sciences, 105-137, Academic Press, New-York.

Warren, D. W., T. M. Dalston, et al. (1993). Aerodynamics of nasalization. Phonetics and phonology. Nasals, nasalization and the velum. M. K. Huffman, Rena. San Diego, Academic Press, 5, 119-146.