Journal of Phonetics (2000) **28,** 43–67
doi:10.006/jpho.2000.0106

Acoustic analysis of simple vowels preceding a nasal in Standard Chinese

Marilyn Y. Chen*

Institute of Linguistics, Chinese Academy of Social Sciences, 5 Jianguomennei Dajie, Beijing 100732, People's Republic of China

Received 29th January 1999, and in revised form 12th April 2000

The aims of this study are (1) to determine the acoustic pattern signaling that a syllable in Standard Chinese, SC, contains a nasal coda, and (2) specify the acoustic attributes that distinguish between the two places of articulation for a nasal coda. The first syllable in SC disyllabic words with simple vowels [a], [i], and [ə] was examined in different contexts: (1) vowel with a nasal coda [n] or [ŋ] followed by a syllable beginning with a stop or vowel, and (2) vowel without a nasal coda followed by a syllable beginning with a stop or [n]. It was found that a nasal coda can be detected by the presence of a vowel–nasal consonant boundary using the maximum first difference of the first four formant amplitudes or by the amount of vowel nasalization using the time contours of normalized $A1-P0$ and $A1-P1$ (amplitude differences between the first formant and nasal peaks). The places of articulation of nasal codas can be distinguished by the formant frequencies in the vowel and by acoustic attributes at the vowel–nasal consonant boundary.

1. Introduction

By definition, nasal consonants are produced by an oral closure and a velopharyngeal opening (House, 1957; Stevens, 1998). Nasalized vowels are produced with an oral opening and a velopharyngeal opening. The goal of this study of Standard Chinese disyllabic words is to determine the acoustic pattern which signals that a syllable contains a nasal coda and to specify the acoustic attributes that distinguish between the two places of articulation for a nasal coda. The vowel preceding the nasal coda was examined for nasalization to confirm the presence of a nasal coda. In addition, in cases where oral closure is not realized in the articulation of the nasal consonant, the nasalized vowel is the only cue for place of the intended consonant. The results of this study may be useful to research into speech synthesis and speech modeling of vowels in nasal contexts and is intended to deepen our understanding of the perceptual cues for nasal place in Standard Chinese. With further testing, it may lead to the enhancement of speech recognition for nasal codas in SC. Extension of the findings from this study to other languages may be possible with careful examination of those languages using similar techniques.

---

*Now at Research Laboratory of Electronics, Massachusetts Institute of Technology, MA, USA.

0095–4470/00/010043 + 26 $35.00/0

The different vocal tract configuration for a vowel and a nasal consonant introduces an abrupt change in the spectrum at the transition between the vowel and the consonant (House, 1957). To predict the change at the vowel–nasal consonant boundary, V : N, it is necessary to compare the articulation and acoustic attributes of the vowel and the nasal murmur. Previous theoretical work has been based largely on English, and this work may be extended to Chinese and other languages. The production alterations from oral to nasal vowel involve coupling to the nasal cavity so that the natural frequencies in the volume–velocity transfer function are shifted and additional poles and zeroes are present (Stevens, 1998). The shifted formants are mainly dependent on the vocal-tract configuration determined by vowel type and by the velopharyngeal (v–p) opening size, while the nasal peaks, around 230–950 Hz on average, and zeros are mainly dependent on the nasal tract coupling (Chen, 1997). Switching of the nose and mouth output to nose alone for the nasal consonant introduces zeros to the transfer function, based on the distance between the v–p opening and the oral closure, and eliminates the spectral peak corresponding to the resonance of the cavity anterior to the oral constriction (Stevens, 1998).

For an alveolar nasal consonant, the oral closure is formed with the tongue tip 5–6 cm from the v–p port, introducing zeros around 1600–1900 Hz and 5 kHz. The cavity in front of the oral constriction is not excited during the nasal murmur so that the amplitude above 4500 Hz, which is the lowest natural frequency of the cavity, decreases from the vowel to [n]. The poles for the alveolar murmur occur roughly at 250, 1000, 1700, 2000, 3200, and 4000 Hz (Stevens, 1998). A velar nasal consonant is formed with the dorsum of the tongue body contacting the hard or soft palate about 3–7 cm from the lips, introducing zeros above 3 kHz in the murmur and eliminating the excitation of the front cavity with natural frequency in the range 1300–3000 Hz. The velar murmur in the context of a neutral vowel has poles at about 250, 1200, 2100, and 3200 Hz. Based on calculations and measurements, Stevens found that the amplitudes of the second and higher formant peaks of a neutral vowel decrease more than 20 dB over a 30–40 ms time interval from the vowel to the beginning of alveolar or velar closure. Over a 7–11 ms interval before the complete oral closure for the nasal consonant, the amplitude drop of the second-formant prominence averages around 10 dB for an alveolar murmur with different vowel types. About 60–70% of the amplitude change (in dB) occurs during this short time interval. For a velar consonant, most of the change occurs in the estimated time period of 10–15 ms before the complete oral closure, which is the time required for the acoustic mass of the nasal cavity to dominate that of the oral cavity, based on the estimated rate of change of the cross-sectional areas in the oral and nasal cavities. This time period is slightly greater than that for the alveolar consonant. The first-formant amplitude, on the other hand, does not show as much change as the amplitude of the second formant at the V : N boundary.

In cases where the nasal coda is realized as a murmur, the V : N boundary may be determined. Estimates of this boundary location have been made for English based on the waveform, the spectrogram, and changes in the spectra. Kurowski & Blumstein (1987) determined the point of release between the nasal murmur for [m] and [n] and the transition into the vowel by visually identifying in the waveform a break in the murmur pulse pattern with the beginning of high-frequency components, detecting abrupt spectral changes via LPC analysis, and perceptual testing. Based on the fact that the vowel has more high-frequency energy than nasal murmur, Seitz, McCormick, Watson, & Bladon (1990) used a manuo-visual estimate of the greatest change of spectral slope in the nasal murmur and 100 ms of the adjacent vowel. The boundary between the nasal and vowel was also determined based on abrupt change in energy and in formant frequencies on spectrogram displays (Harrington, 1994). The above-mentioned techniques are subjective with no attempt to quantify the parameter used for boundary detection. Compared to NV, there is less boundary identification accuracy for VN due to a less abrupt spectral change between the vowel and the murmur (Repp, 1986; Repp & Svastikula, 1988), not to mention finding the beginning of nasalization within the vowel.

In addition, oral closure may sometimes be eliminated (although a partial constriction may occur), so that no V : N boundary exists. Previous work using acoustic analysis has not distinguished these cases from nasal consonants made with an oral closure which introduces a murmur. Nasal absorption, which is the reduction of the vowel and nasal consonant sequence to a nasal vowel, has been observed phonologically in several languages including Chinese (Trigo Ferré, 1988). Based on the assumption that the nasal feature is stable, Trigo Ferré believes that nasal absorption is caused by the oral occlusion of a nasal stop being removed or weakened to a glide. She suggests that the nasal consonant is either first debuccalized so that it loses its place feature or it undergoes other weakening processes such as spirantization and gliding before absorption. As a result, nasalization spreads onto the neighboring vowels. According to Chen (1973), longitudinal studies of Chinese dialects show debuccalization and absorption for target nasals in the coda position. On the other hand, Hajek (1997) argues strongly that the phonological process of vowel nasalization is independent of nasal attenuation or deletion, based on substantial evidence and evaluation of cross-linguistic data.

Independent of oral closure realization for the nasal consonant, nasal anticipation affects the degree and timing of velum and pharyngeal wall movement in the preceding vowel (Ohala, 1975). From cinefluorographic, photoelectric, and electromyographic studies, it has been suggested that high vowels exhibit significantly greater velar height than low vowels (Chen & Wang, 1975; Clumeck, 1976; Bell-Berti, 1993). Furthermore, vowels that precede a nasal consonant have a greater velum–pharynx distance than vowels following a nasal consonant. In fact, according to Moll & Daniloff (1970), velar opening coarticulates over as many as two vowels preceding the nasal, even in cases where there is a word boundary between the two vowels. By using Velotrace to track the vertical velic movements, Krakow (1993) concluded that a vowel preceding a syllable-final [m], whether in word-medial or word-final position, has lower velic height than a vowel preceding a syllable-initial [m]. However, Hajek (1997) cautioned that there is no sufficient cross-linguistic evidence to support the velic opening characteristics as universal.

Nasalization can be quantified indirectly by using frequency domain analysis of speech and by inferring the size of the opening to the nasal cavity from this analysis (Fujimura, 1960). The spectral alterations due to vowel nasalization often occur in the first formant region (Ohala, 1975). Maeda (1993) utilized spectral spread in the low frequency as an indication of the degree of vowel nasalization. However, his measure worked well for only two out of the three vowels that he examined. Vowel nasalization introduces spectral prominences and reduction of the first formant amplitude. Acoustic correlates distinguishing non-nasal and nasal vowels were developed (Chen, 1997) with nasal peak amplitude $P0$ (dB) below the first formant, spectral prominence with an amplitude $P1$ (dB) between the first two formants, and the first formant amplitude $A1$ (dB). The difference $A1 - P0$ shows a statistically significant variation between non-nasal and nasal vowels, especially in the non-high vowels; the variation of the $A1 - P1$ difference is especially evident in the non-low vowels, with an inverse relationship between the acoustic correlates and nasalization.

Once the existence of a nasal coda is detected by the presence of a V : N boundary or by the amount of vowel nasalization, the place of articulation for the nasal consonant can be determined. Based on perceptual studies, some researchers believe that only the vowel transition is a place cue, while the nasal murmur is a manner cue (Malécot, 1956; Pickett, 1965; Delattre, 1968; Mermelstein, 1977; Miller, 1977; Larkey, Wald & Strange, 1978). On the other hand, other researchers have found that combining the nasal murmur with the vowel transition enhances the perception of nasal place for English (Kurowski & Blumstein, 1984; Repp & Svastikula, 1988; Ohde, 1994) and Catalan (Recasens, 1983), since the discontinuity at the boundary may provide the relevant cue. Attempts have been made to develop an algorithm for identifying the place of articulation for nasals. Kurowski & Blumstein (1987) used energy changes between murmur and transition to label labials and alveolars in syllable-final position, with an accuracy of 56–75%. Seitz *et al.* (1990) used the maximum and minimum in the difference spectrum of the murmur and the vowel, achieving 51% correct classification for syllable final nasals [m, n, ŋ]. Harrington (1994) found that combinations of murmur and vowel spectra rather than difference spectra achieved the best results for the place distinction, with 82% correct for syllable-final nasals. This finding suggests that the murmur and vowel contribute independently to the place of articulation distinction.

This study examines simple vowels preceding a nasal in Standard Chinese (SC). Chinese characters are uttered as monosyllables with simple, diphthongal, or triphthongal vocalic patterns. In Standard Chinese, the numbers of rimes with [n] or [ŋ] coda are equal; simple vowels with nasal coda include words with [an, ən, in yn, aŋ, əŋ, iŋ, uŋ] (Hanyu Fangyin Zihui, 1989). Variations are formed by adding consonant initials, including the zero initial for which the consonant is not present, and by assigning tones: tone-1 (˥), tone-2 (˧˥), tone-3 (˨˩˦), or tone-4 (˥˩). Ren (1988) used the second-formant frequency F_2 of the vowel to examine SC monosyllables in the form (V)VN, where N is [n] or [ŋ], with tone-1 (˥). He found that in general, the beginning and end of the F_2 transition of [a, i, ə] have a higher frequency with [n] as the nasal coda rather than [ŋ]. In another study of monosyllabic (V)VN in SC, Lin & Yan (1991) also found that, independent of tone, for vowels before [n] as oppose to [ŋ], F_2 is higher at the end of (V)V for [a, i, ə]; for [a, i], the first formant-frequency F_1 is lower; and for [i], the third formant-frequency F_3 is higher. Although in Chinese about 70% of the words with full meaning are disyllabic, a similar study on vowel formant frequencies preceding the nasal coda in a disyllabic setting has not been carried out.

In this study, disyllabic words in SC were examined. The analytical techniques that were utilized in this study were: (1) vowel–nasal consonant boundary detection using the maximum first difference of the first four formant amplitudes, (2) vowel nasalization using time contours of normalized $A1 - P0$ and $A1 - P1$ (amplitude differences between the first formant and nasal peaks), $A1 - P0n$ and $A1 - P1n$, and (3) formant frequencies at the end of the vowel and formant frequencies averaged over time in the vowel. Normalization of $A1 - P0$ and $A1 - P1$ was needed to correct for the effect of formant frequencies on the amplitudes of the nasal peaks $P0$ and P1.

2. Method

2.1. *Corpus*

The speech material was based on a preliminary study of disyllabic words consisting of simple vowels with nasal codas followed by all phonetically possible syllable initials in

Standard Chinese spoken by one male speaker. The study showed that when the initial consonant in the second syllable was a stop consonant, the preceding nasal coda was produced with an oral closure. Such a closure often did not happen when the second syllable began with a zero-initial.

Based on preliminary results, the corpus of the present experiment consists of SC disyllabic words with simple vowels [a, i, ə] with nasal coda, [n] or [ŋ], in the first syllable in a carrier phrase "wo^3 shuo^1 ____ zhe^4 ge^4 ci^2", (in English "I say ____ this word"). The numbers indicate the tones. The second syllable may begin with a stop or vowel. For the purpose of comparison, vowels without a nasal coda followed by a syllable beginning with a stop or [n] were also analyzed. In order to study how the acoustic parameters change with time, simple vowels rather than diphthongs or triphthongs were examined, since a given vowel in a monophthong has a longer duration (Ren, 1988). For [a], the disyllabic words are "da^4-ta^3", "da^4-nao^3", "dan^1-da^3", "shan^1-ao^4", "fang^4-da^4", and "chang^2-ao^3"; for [i], they are "bi^3-ti^3", "bi^3-ni^3", "xin^1-pi^2", "xin^4-yi^4", "ping^2-bi^3", and "ding^4-yi^4"; and for [ə], they are "ke^3-pa^4", "ke^1-na^4", "fen^1-pi^1", "fen^4-e^2", "feng^1-pi^2", and "deng^1-e^2". The test phrases were presented in Chinese characters. All of the phrases were read once and the process was repeated four more times by five male and five female speakers of Standard Chinese.

The test tokens were chosen to reflect plausible words in Standard Chinese. Therefore, the tones are not consistent since it is difficult to have meaningful words with the same tonal patterns. Tone was found to have little effect on the vowel formant frequencies except for F_1 at the end of the vowels preceding [n] (Lin & Yan, 1991). By averaging the F_1 values over time throughout the vowels, it is likely that the effect of tone on the first formant frequency is lessened. Furthermore, in SC there is no voice/voiceless distinction for stops; there is only an aspirated/unaspirated distinction. Therefore, orthographically, the initial consonant of the second syllable in "da-ta" and "bi-ti" is [t^h]; in "dan-da" and "fang-da" is [t]; in "xin-pi", "ke-pa", "fen-pi", and "feng-pi" is [p^h]; and in "ping-bi" is [p]. Since it is the voice/voiceless distinction that affects the duration of the preceding vowel, the duration of the nasal consonant, and nasal reduction (Raphael, Dorman, Freeman & Tobin, 1975; Hajek, 1997), the different types of stop consonant at the initial position of the second syllable may not be an issue. In addition, the only duration considered in this study is that of vowel nasalization, which was not shown to be affected by the following stop consonant (Fujimura & Lovins, 1978).

The corpus was chosen so that comparisons could be made between vowels (1) in a nasal context with [n] or [ŋ] as nasal coda (a) before a stop-initial syllable (i.e., "dan-da", "fang-da", "xin-pi", "ping-bi", "fen-pi", "feng-pi") or (b) before a zero-initial syllable (i.e., "shan-ao", "chang-ao", "xin-yi", "ding-yi", "fen-e", "deng-e"), (2) in a nasal context with [n] across a syllable boundary (i.e., "da-nao", "bi-ni", "ke-na"), and (3) in a stop consonant context (i.e., "da-ta", "bi-ti", "ke-pa"). The first set allows comparison of vowels with different place for the nasal coda whether or not the oral closure was realized. The variation of the number of tokens that do not have oral closure formed for the nasal coda was found to be large, depending on the vowel type and the nasal coda. In order to reduce the variation, more tokens with no oral closure were generated by having the words with zero-initial in the second syllable spoken at a fast rate in addition to the normal rate. The second set allows a comparison of coarticulation in the vowel due to syllable-final *vs.* syllable-initial [n]. The last set is used to examine vowels in non-nasal context, for comparison with vowels in nasal contexts.

2.2. *Recording*

The recording was done with an omnidirectional microphone on a stand about 3 in from the speakers' lips and a JVC model TD-W318 recorder in a sound-attenuated room. Digitization and data analysis were done on a Compaq personal computer using the Speech Station Version 3.0 (Sensimetrics Corporation). The signal was digitized at 10 kHz, and spectra without pre-emphasis of the vowel and the nasal consonant were calculated by using a 25.6 ms Hamming window and 512-pt FFT.

3. Boundary detection

3.1. *Analysis*

Based on the acoustic model of nasal consonant production, a nasal murmur can only be produced with an oral closure and a velopharyngeal opening. Theoretically, because the vocal tract configuration varies between a vowel and a nasal consonant with closure, a nasal murmur is distinguished from the preceding vowel by an abrupt change in the amplitudes of the formants. The change occurs especially for the first four formants, which have the greatest formant amplitudes in the vowel. The V : N boundary, with N as the nasal coda of the first syllable or nasal initial of the second syllable, was determined by examining the time course of the amplitudes for the first four formant peaks in the spectra of the vowel and nasal consonant. Fig. 1 shows the spectrogram of [ən] in "fen-pi" for one of the utterances. The arrows at the bottom indicate the times at which a window was centered in order to calculate the FFT spectra, which were 10 ms apart from the beginning of the first glottal vibration of the vowel to the end of the nasal consonant. Based on the expectation that the large formant amplitude changes occur within a time interval of just 7–15 ms in the vicinity of the V : N boundary, 10 ms intervals were used for boundary location. The data collection starts by placing the center of the window 12.5 ms after the onset of voicing according to the waveform and spectrogram, so that the left edge of the window is placed at the beginning of the voicing onset. It ends before more than half of the window moves past the last glottal pulse of the vowel or of the murmur if it exists. By way of illustration, four of the spectra are shown in the figure with the amplitudes of the first four formants, labeled as $A1$, $A2$, $A3$, and $A4$. The formant amplitude in the vowel was defined as the maximum harmonic amplitude near the formant peak in the spectrum; in the nasal consonant, the maximum harmonic amplitude closest to the formant frequency of the preceding vowel (± one harmonic) was used. For each of the four formants, the amplitude at a given time minus the corresponding amplitude sampled 10 ms later (the first-difference amplitude) was obtained, its time contour was plotted, and the maximum first difference was found. In Fig. 1, the maximum differences for all of the four formants were obtained from the right-most two spectra. This is called the maximum first difference for each formant amplitude.

A threshold may be used to detect the actualization of nasal consonants. The process of detecting the V : N boundary objectively was guided by visual inspection of the spectrogram and the waveform. From subjective observation, vowels with a nasal coda followed by a stop (VN-C_s), where N is [n] or [ŋ], as shown in Fig. 2(a) and (b), or vowels followed by [n] at the beginning of the second syllable (V-n), as shown in Fig. 3(a), all had a distinct V : N boundary due to the presence of the nasal murmur. Therefore, measurements were first made in VN-C_s and V-n utterances to determine the maximum first

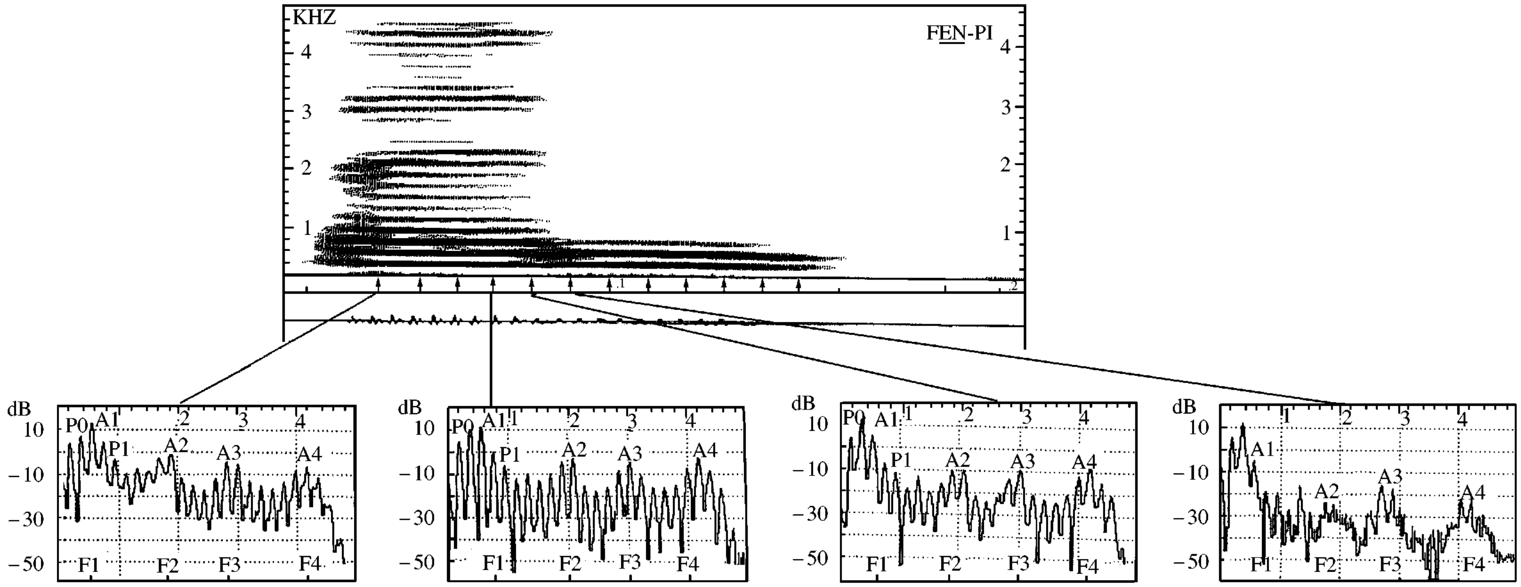

Figure 1. Spectrogram of [ən] in “fen-pi” is shown. Four of the FFT spectra calculated 10 ms apart from the beginning of the first glottal vibration of the vowel to the end of the nasal consonant are also shown.

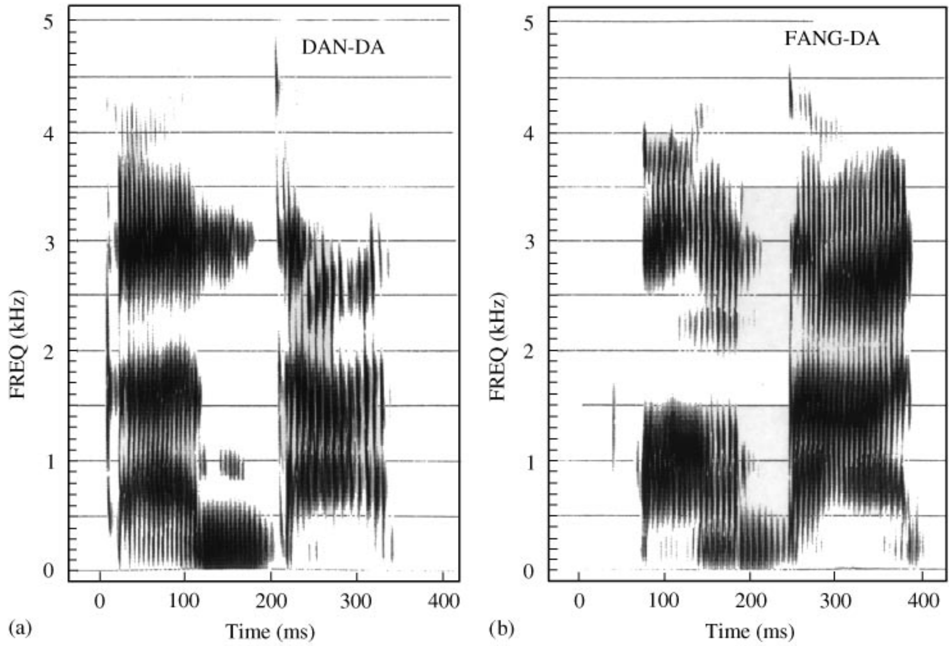

Figure 2. Spectrograms of (a) "dan-da" and (b) "fang-da" are shown to compare the effect of nasal coda [n] *vs.* [ŋ] on the preceding vowel [a].

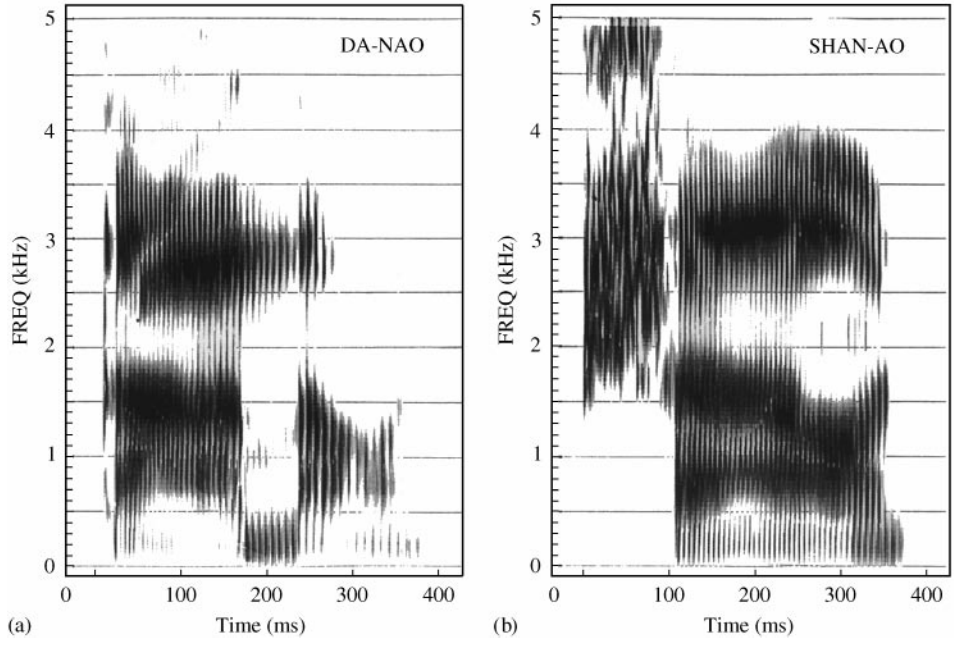

Figure 3. The spectrograms of (a) "da-nao" and (b) "shan-ao" are shown to compare the effect of syllable-initial [n] across syllable boundary and nasal coda [n] without oral closure.

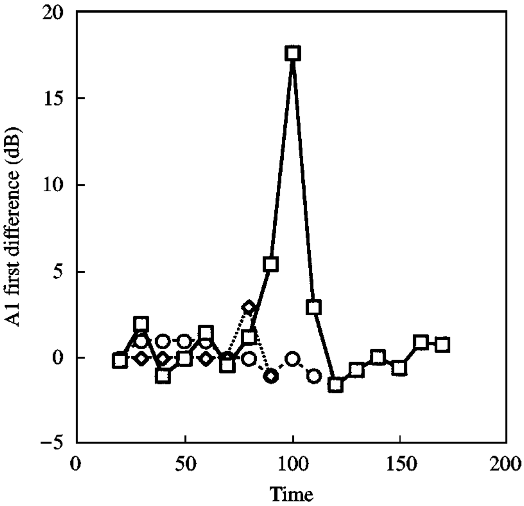

Figure 4. An example of *A*1 first difference over time of a given speaker for [a] in three contexts: an-C_s (□); a-C_s (◇); an-a (○).

difference for each of the four formant amplitudes (*A*1, *A*2, *A*3, *A*4) in three vowel types ([a, i, ə]), and three nasal contexts (Vn-C_s, Vŋ-C_s, V-n) for a given speaker. The effects of the three nasal contexts on the maximum first difference for the four formant amplitudes were observed and compared with the maximum first difference amplitude values within the vowel of corresponding vowel type in V-C_s, where no V : N boundary is expected. Because the vocal tract configuration varies more between a vowel and a nasal consonant with closure than throughout the vowel, the maximum first differences ought to be the greatest at the V : N boundary. As the two spectra adjacent to the V : N boundary in Fig. 1 indicate, there was a steep drop for the four formant amplitudes. On the other hand, vowels with nasal coda followed by a zero-initial in the second syllable (VN-V) often did not have a clear V : N boundary, as shown in the spectrogram in Fig. 3(b).

In order to objectively determine the existence of a V : N boundary, a threshold was implemented. If the formant amplitude changes showed a significant statistical difference ($p < 0.001$ from a one-tailed test) between VN-C_s and V-C_s for a given vowel across repetitions and speakers, the maximum first difference averaged across repetitions from VN-C_s was used as a threshold for V : N boundary detection. For each vowel, nasal coda, and speaker, there would be four thresholds, one for each of the four formant amplitudes. A test token was then examined to see if the maximum first difference of any one of its first four formant amplitudes reached the threshold described above. If this threshold is reached in the VN portion of VN-V, then it is stipulated that a V : N boundary due to oral closure is detected. Only when the threshold is not reached for any of the formant amplitude differences, is it assumed that no oral closure is made for the nasal consonant. This stipulation of an unoccluded nasal consonant is a stringent one in order to avoid identifying tokens with oral closure as a non-occluded.

By way of illustration, an example of *A*1 first difference over time is shown in Fig. 4 for one token with [a] in each context for a given speaker with the utterances aligned at the beginning of the vowel. The maximum first difference in [an] of an-C_s is about 18 dB

TABLE I. Maximum first difference of the first four formant amplitude time contours averaged over speakers and repetitions for the vowel–nasal consonant contexts. The sum of the four first differences is in the last column. The standard deviations (SD) are also given

VN context	$A1$ (SD) (dB)	$A2$ (SD) (dB)	$A3$ (SD) (dB)	$A4$ (SD) (dB)	Sum (dB)
a-n	15.2 (3.6)*	12.3 (3.4)*	8.6 (4.3)*	9.9 (3.7)*	46.0 (10.4)
an-C_s	13.3 (3.2)*	11.0 (3.3)*	9.3 (3.6)*	10.3 (3.1)*	43.9 (9.3)
aŋ-C_s	10.2 (4.2)*	8.4 (3.8)*	3.9 (3.1)	6.7 (3.1)*	29.2 (9.2)
i-n	1.9 (1.5)*	10.6 (3.4)*	13.9 (3.6)*	13.0 (2.9)*	39.4 (6.4)
in-C_s	1.6 (1.5)*	9.1 (4.5)*	13.6 (4.7)*	11.8 (3.8)*	36.1 (8.5)
iŋ-C_s	0.9 (1.9)	8.4 (2.6)*	6.7 (3.7)*	6.5 (3.2)*	22.5 (5.6)
ə-n	13.8 (3.6)*	14.1 (4.7)*	8.5 (4.2)*	10.8 (2.8)*	47.2 (9.0)
ən-C_s	9.8 (3.4)*	9.5 (3.6)*	8.0 (4.5)*	9.5 (3.4)*	36.8 (9.3)
əŋ-C_s	5.4 (3.0)*	6.3 (2.4)*	5.0 (3.2)*	8.3 (2.7)*	25.0 (7.2)

*$p < 0.001$ according to a one-tailed test between the values of VN context and non-nasal context.

while it is about 3 dB for [a] of a-C_s for these single-token examples. Assuming the 15 dB difference along with those of other repetitions and of other speakers is significant, the maximum first difference within an-C_s for the given speaker, averaged across five repetitions (13 dB), can be used as a threshold for $A1$. For the repetition of an-a by the same speaker in Fig. 4, the maximum first difference for [an] is only 1 dB, which is lower than the 13 dB threshold. To determine whether the maximum first difference shows statistically significant variation between an-C_s and a-Cs, the same comparison is done for the other three formant amplitudes ($A2$–$A4$). If the relevant maximum first differences are all lower for an-a than their corresponding thresholds, then no V : N boundary exists and it is assumed that there is no oral closure.

The location of the V : N boundary can be estimated as well by the following proposed algorithm. If the maximum first differences of all four formants are determined from the same time slots, the half-way point of the two time slots is defined as the boundary (33% of the tokens); if all four maximum first differences are within adjacent time slots (56%) or the majority of the maximum first differences are in adjacent time slots (11%), the middle time slot is labeled as the boundary. In Fig. 4, the maximum first difference of $A1$ for an-C_s is obtained from $A1$ values at 95 and 105 ms. If the maximum first differences of the other three formants are also from those time frames, then the V : N boundary would be taken to be 100 ms.

3.2. *Results and Discussion*

The average and standard deviation (SD) of the $A1$–$A4$ maximum first differences across speakers and repetitions for words with V : N boundary are listed in Table I. (The variation across gender was examined for the maximum first differences of the formant amplitudes. However, there is no clear separation between the two groups. The only trend is that male speakers generally showed a greater range across speakers than the female speakers.) Several observations emerge from the data in Table I. The change in first-formant amplitude for [i] is clearly much smaller than that for the other two vowels. The small value of $A1$ change for [i] is presumably because F_1 for [i] is low and close to

TABLE II. The number of tokens out of 50 for each word spoken by the 10 speakers at a normal rate and at a fast rate produced without an oral closure for the nasal coda, as indicated by the maximum first difference of the formant amplitudes

Word	Normal rate	Fast rate
shan-ao	50	50
chang-ao	30	13
xin-yi	29	16
ding-yi	7	29
fen-e	29	13
deng-e	14	16

the lowest resonance for the nasal consonants. Thus, the movement from the vowel [i] to the nasal consonant produces little change in the frequency of the lowest resonance and hence little change in spectrum amplitude. Although the first formant of [a] is close in frequency to the second nasal resonance of [n] or [ŋ], in the nasal murmur the amplitude of the second formant may be about 15 dB lower than that of the first formant (Stevens, 1998) so that there is still a large maximum first difference for $A1$ of [a] followed by a nasal consonant. Across contexts, the change in $A2$ at the vowel–nasal boundary for [n] is in the range 9–14 dB, similar to the change reported for this acoustic discontinuity in English by Stevens (1998). These abrupt changes are presumably due to the rapid movement of a zero in the transfer function of the combined vocal and nasal tracts, toward a frequency that is in the range 1600–1900 Hz. This upward movement of the zero would also cause an abrupt downward change in the amplitudes $A3$ and $A4$. The change in $A4$ may be further enhanced by the fact that the tongue-blade closure creates a cavity anterior to the closure, and excitation of the natural frequency of this cavity is suddenly eliminated. The reduced maximum first differences in $A2$–A4 for the [ŋ] context relative to those for the [n] context can be attributed to the lower rate of decrease of the cross-sectional area of the oral cavity constriction for velars (Stevens, 1998). The general trend of reduced maximum first difference in the context of VN-C_s relative to those in the context of V-n may be attributed to the greater effect of nasal assimilation of word-final nasal consonants on the vowel than the effect of word-initial nasal consonants (Vaissière, 1988; Krakow, 1993).

With the technique described above, thresholds for the V : N boundary detection were obtained. The maximum first differences for each of the four formants, three vowel types, and three nasal contexts were compared with the corresponding vowel in non-nasal context for all of the speakers. Table I indicates the formants that showed statistical significance by a one-tailed test. All showed significant differences except $A3$ of aŋ-C_s and $A1$ of iŋ-C_s, which were not incorporated into the thresholds. Since it is the vowels in nasal coda contexts that generally showed nasal murmur deletion, the averages of the amplitude changes across repetitions in the VN-C_s context instead of those in the V-n context were used as thresholds for V : N boundary detection.

Once the threshold was determined, it was applied to the words with VN-V to detect those with nasal coda that were formed without oral closure. Table II shows the number out of 50 tokens for each word with Vn-V or Vŋ-V that did not make the oral closure at the normal and fast speaking rates. At the normal speaking rate, words with [a], independent of nasal coda, have about twice as much oral closure deletion as the other vowels. A possible explanation is that a lower jaw movement is required to make the low vowel so that there is a lack of time to raise the jaw to form the closure, particularly for [n]. Nasal coda [n], independent of vowel type, causes 1.7–4.1 times greater occurrence of oral closure deletion than [ŋ].

These results are in agreement with Chen (1975) who suggested that in the major Chinese dialects, word-final alveolar nasals are more likely to undergo absorption than word-final velar nasals. However, this result disagrees with the labial–coronal–velar hierarchy of nasal absorption going from low to high susceptibility, as proposed by Trigo Ferré (1988). Hajek (1997) gave examples from several languages to show the randomness of the effect of nasal place on nasal absorption. He suggested possible factors for such randomness, such as vowel length distribution and word-final consonant loss. A likely cause for the different effects of nasal coda type on oral closure may be that the formation of a closure is enhanced by raising the tongue body for [ŋ] in addition to the velum lowering.

Since not all nasal codas in VN-V produced at the normal rate were made without oral closure, an additional number of tokens produced at a fast rate was included in order to achieve more similar numbers of non-occluded tokens for each word. All of the tokens identified in Table II at both normal and fast rates were used for further acoustic analysis except for “shan-ao”, for which only the tokens at normal rate were used. Since all of its tokens produced at the normal rate already have oral closure deletion, none of the tokens spoken at the fast rate were used. In general, there are fewer items without oral closure at the fast rate than the normal rate, suggesting that the faster speaking rate may not necessarily enhance oral closure deletion for the nasal coda. This unexpected result requires a closer examination of the possible causes of oral closure deletion.

If there is a boundary, its position in time can also be located by the method described above. The location of the V : N boundary determined by this method was compared with that found manually according to the spectrogram and waveform. The Pearson product moment correlation coefficients are above 0.97 with an average deviation less than 7 ms in either direction. Therefore, by using the maximum first difference of the formant amplitudes, the V : N boundary may be accurately determined automatically.

4. Acoustic correlates of vowel nasalization

4.1. *Analysis*

The acoustic correlates of vowel nasalization in this study are based on the amplitudes in dB of the first formant ($A1$), the nasal peak above the first formant ($P1$), and the nasal peak at low frequencies ($P0$). These quantities were measured every 10 ms in the vowel, beginning with the first glottal pulse of the vowel and ending before more than half of the window moves past the end of the vowel. The end of the vowel is determined by the boundary-detection technique or by the last glottal pulse in the vowel if there is no nasal murmur. Generally, the maximum harmonic amplitudes in the vicinity of the first formant and nasal peaks were used for $A1$, $P1$, and $P0$. Fig. 1 shows three spectra obtained from the vowel with $A1$, $P1$, and $P0$ labeled. Theoretically, as the vowel is more nasalized, $A1$ is lowered while $P1$ and $P0$ are raised. The kinematics of vowel nasalization were characterized by the time course of the contours of $A1 - P1$ and $A1 - P0$. Chen

(1997) found that $A1 - P1$ and $A1 - P0$ are significantly smaller for nasal than for non-nasal vowels in English and French. Adjustment of $A1 - P1$ and $A1 - P0$ for formant frequency changes over time due to vowel type, nasal coda coarticulation, and speaker variation was made by removing the effects of the first two formants on the amplitudes $P0$ and $P1$ to obtain $A1 - P0n$ and $A1 - P1n$. The calculations of the amplitude boost were based on the first two formant frequencies, F_1 and F_2, and on the nasal peak frequencies, as suggested by Chen (1997).

The appropriate acoustic parameter was analyzed for each of the different vowel types. For [i], $A1 - P1n$ is a better measure than $A1 - P0n$ since F_1 at a low frequency affects and is affected by the $P0$ peak. This nasal peak may even occur at the same frequency as F_1 and therefore boost $A1$. On the other hand, if the nasal peak is at a frequency different from and with an amplitude higher than the first formant, it may be mistaken as the first formant. Since for a vowel adjacent to nasal consonant, the frequency of the first formant F_1 is expected to move up (Fujimura, 1960) while $A1$ is expected to decrease, and also (from data) the first-harmonic amplitude remains fairly constant throughout a nasalized vowel, the second-harmonic amplitude was used as $A1$ for nasalized [i]. For non-nasalized [i], the first or second harmonic with the maximum amplitude was used to measure $A1$. For [a], $A1 - P0n$ is a better measure than $A1 - P1n$ since the $P1$ peak sometimes is not distinctly observable due to the high F_1 and low F_2. For [ə], both $A1 - \mathrm{P}1n$ and $A1 - P0n$ contours were used to quantify nasalization.

4.2. *Results and discussion*

The effect of an underlying nasal coda on the quality of the preceding vowel, whether or not the oral closure is formed, can be detected by examining nasalization in the vowel. The results from analyzing the acoustic correlates $A1 - P0n$ and $A1 - P1n$ are presented according to vowel type.

4.2.1 *Vowel [a]*

Average values across speakers and repetitions were obtained from $A1 - P0n$ time contours of the vowel [a] in the initial syllable with different contexts. Because $A1 - P0n$ is inversely related to nasalization, the portion of the vowel having the greatest velopharyngeal (v–p) opening would be indicated by the lowest level (minimum) of this measure while the portion with the least nasal coupling would be indicated by the highest level (maximum). Fig. 5(a) shows that the range of $A1 - P0n$, and therefore the change of the v–p opening, is smaller for nasalized vowels without realization of the murmur ("shan-ao" and "chang-ao") and for vowels in non-nasal context ("da-ta"). Fig. 5(b) shows the slope of $A1 - P0n$, which is obtained by first plotting $A1 - P0n$ over time at 10 ms intervals over the entire vowel and then finding the slope of the line that best fits the plot. The positive and negative values reflect a general decreasing and increasing, respectively, of v–p opening. Fig. 5(c) plots the duration of vowel nasalization in nasal contexts, starting with the time when $A1 - P0n$ is below the minimum of non-nasal [a] in "da-ta" averaged across repetitions for the corresponding speaker and terminating with the end of the last glottal pulse of the vowel. The average period of nasalization is 50–100 ms for [a] in various nasal contexts with an-V having the longest duration.

The upper section of Table III indicates the statistically significant differences in word pairs (subtracting the mean value of the second word from that of the first word) of the

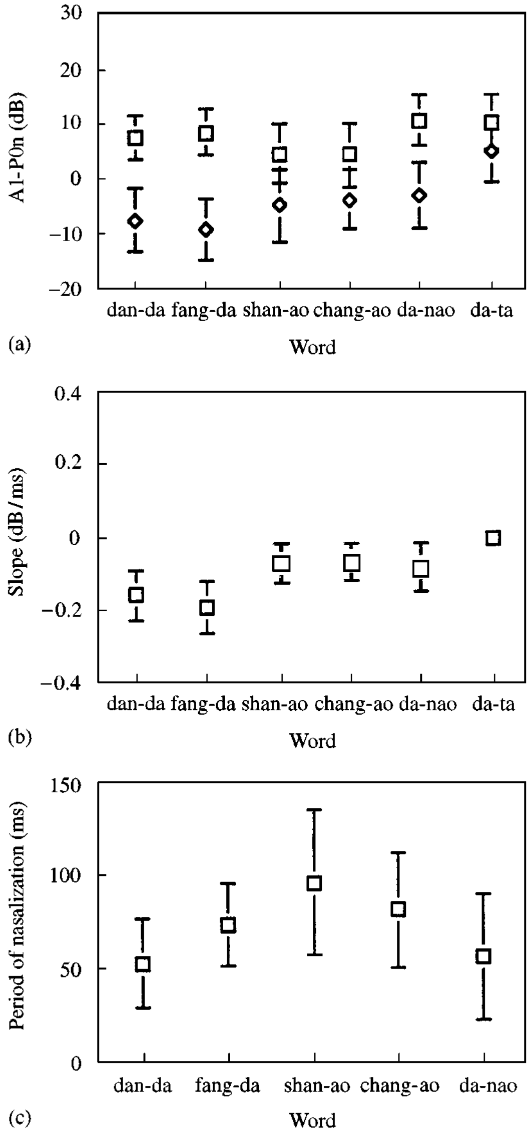

Figure 5. (a) Minimum (□) and maximum (◇) of $A1 - P0n$, (b) slope of $A1 - P0n$ (□) time contour and (c) duration of vowel nasalization obtained by using the minimum $A1 - P0n$ (□) of non-nasal vowel averaged across speakers and repetitions and shown for [a] in different contexts.

maximum, the minimum, and the slope of $A1 - P0n$ timing contour, and the duration of nasalization for [a]. Comparison of the pairs in non-nasal and nasal coda contexts shows that the $A1 - P0n$ maxima and minima of the latter are lower and vowels with nasal coda have a more negative slope. The significant maximal and minimal differences suggest a greater degree of nasalization and the slope differences indicate a greater rate of nasalization for vowels in nasal contexts. Furthermore, the pairs comparing a-n and aN-C_s, where N is [n] or [ŋ], show that when nasalization of the vowel arises from

TABLE III. The differences within word pairs for parameters obtained from $A1 - P0n$ or $A1 - P1n$ (in italics) timing contours for [a], [i], and [ə] in non-nasal and syllable-initial nasal contexts compared to nasal coda contexts

Word pair	Maximum (dB)	Minimum (dB)	Slope (dB/ms)	Period$_{\text{nasal}}$ (ms)
a-C$_s$; an-C$_s$	2.7*	12.1*	0.15*	—
a-C$_s$; aŋ-C$_s$	1.6*	13.9*	0.19*	—
a-C$_s$; an-V	5.8*	9.7*	0.07*	—
a-C$_s$; aŋ-V	6.0*	8.5*	0.07*	—
a-n; an-C$_s$	3.0*	4.0*	0.07*	3.4
a-n; aŋ-C$_s$	1.9*	5.8*	0.11*	– 16.8*
a-n; an-V	6.1*	1.6	– 0.01	– 39.8*
a-n; aŋ-V	6.3*	0.4	– 0.02	– 24.8*
i-C$_s$; in-C$_s$	_4.5*_	_2.9_	_0.17*_	—
i-C$_s$; iŋ-C$_s$	_9.1*_	_10.1*_	_0.21*_	—
i-C$_s$; in-V	_2.8*_	_6.1*_	_0.12*_	—
i-C$_s$; iŋ-V	_5.4*_	_7.5*_	_0.18*_	—
i-n; in-C$_s$	_4.0*_	_2.0_	_0.14*_	_1.3_
i-n; iŋ-Cs	_8.6*_	_9.2*_	_0.18*_	_– 11.7_
i-n; in-V	_2.3_	_5.2*_	_0.09*_	_– 23.2_
i-n; iŋ-V	_4.8*_	_6.6*_	_0.14*_	_– 15.1_
ə-C$_s$; ən-C$_s$	_5.5*_	10.5*	0.34*	—
ə-C$_s$; əŋ-C$_s$	_5.7*_	10.7*	0.33*	—
ə-C$_s$; ən-V	_4.1*_	10.5*	0.25*	—
ə-C$_s$; əŋ-V	_6.3*_	13.1*	0.22*	—
ə-n; ən-C$_s$	– 0.6	3.0*	0.20*	_15.9*_
ə-n; əŋ-C$_s$	– 0.6	3.2*	0.20*	_6.2_
ə-n; ən-V	– 3.1*	3.0	0.11*	_– 25.1*_
ə-n; əŋ-V	2.5	6.4*	0.07†	– 46.7*

*$p < 0.001$,
†$p < 0.005$.

a nasal consonant across a syllable boundary, the degree and rate of nasalization is smaller than when nasalization is due to a nasal coda with oral closure. This result corresponds to the finding by Krakow (1993) that a vowel preceding a syllable-final [m] has lower velic height than when the vowel precedes a syllable-initial [m]. The amount of anticipatory nasal coarticulation, indicated by the period of nasalization, Period$_{\text{nasal}}$, is significantly greater for [a] followed by nasal coda [ŋ] with murmur than for [a] followed by [n] across a syllable boundary but not for an-C$_s$ vs. a-n. A much longer period of nasalization is also found for [a] with nasal coda without oral closure, an-V and aŋ-V, than in a-n, but the degree and rate of nasalization do not show significant differences except for the maxima.

4.2.2. *Vowel [i]*

Measures obtained from $A1 - P1n$ time contours during the vowel [i] in various contexts are shown in Figs 6(a)–(c). According to Fig. 6(a), the vowels in non-nasal ("bi-ti") and syllable-initial nasal context ("bi-ni") have a smaller range than those with nasal coda, except for "xin-pi", which has short vowel duration yielding measurements

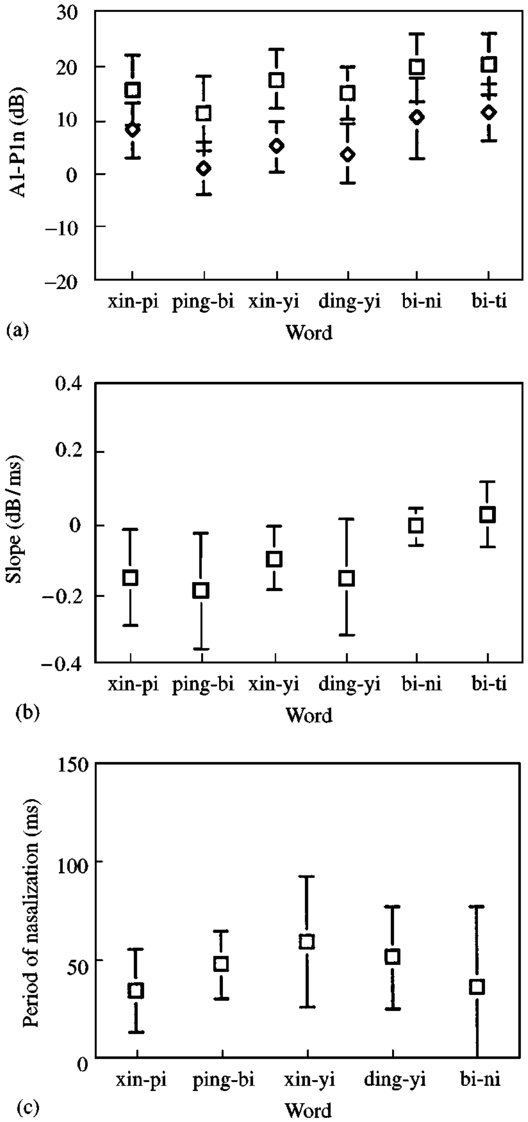

Figure 6. (a) Minimum (◇) and maximum (□) of $A1 - P1n$, (b) slope of $A1 - P1n$ (□) time contour and (c) duration of vowel nasalization obtained by using the minimum $A1 - P1n$ (□) of non-nasal vowel averaged across speakers and repetitions and shown for [i] in different contexts.

from 1 to 3 time frames. The negative slopes for vowels in nasal contexts, as shown in Fig. 6(b), indicate increasing nasalization into the the vowel due to anticipatory coarticulation. The more negataive slopes for [i] with nasal coda, even the ones without nasal murmur, show a faster rate of nasalization. The nearly zero slope of [i] followed by a syllable-initial [n] and the positive slope of [i] in non-nasal context reflect the lack of nasalization. Fig. 6(c) shows the duration of nasalization in [i] with the starting time

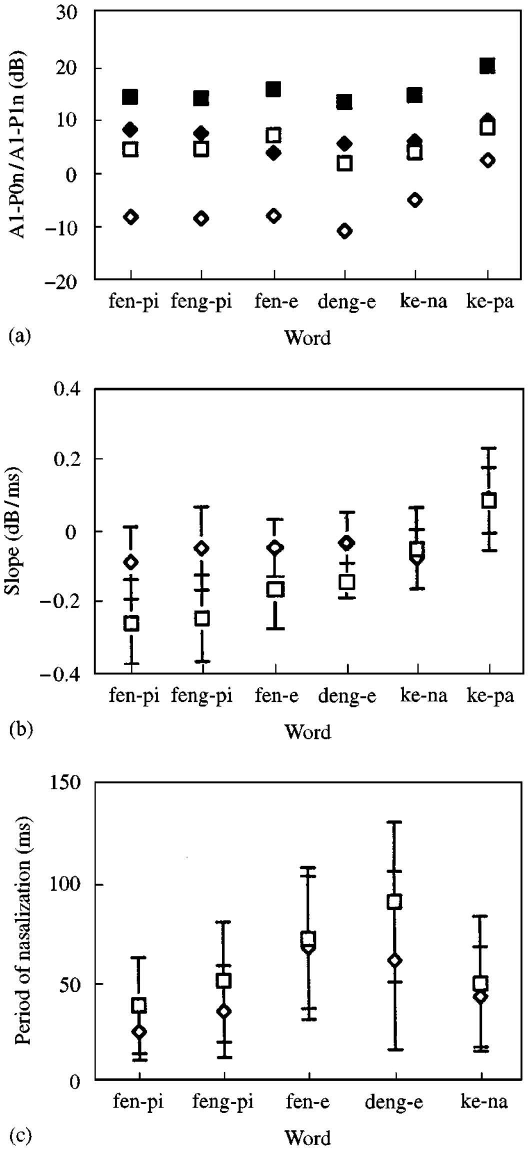

Figure 7. (a) Minimum and maximum of $A1 - P0n$ and $A1 - P1n$, $A1 - P0n$ maximum (□); $A1 - P0n$ minimum (◇); $A1 - P1n$. maximum (■); $A1 - P1n$. minimum (◆), (b) slope of $A1 - P0n$ and $A1 - P1n$ time contour and (c) duration of vowel nasalization obtained by using the minimum $A1 - P0n$ and $A1 - P1n$ of non-nasal vowel averaged across speakers and repetitions and shown for [ə] in different contexts.

based on the average $A1 - P1n$ minium of [i] in the non-nasal context for the corresponding speaker. The period of nasalization is 30–60 ms for [i] in different nasal contexts, with "xin-yi" having the longest duration, although the SD is large.

The middle section of Table III lists the significant differences of maximum, minimum, and slope form $A1 - P1n$ timing contour and period of nasalization within word pairs with [i]. Comparison of the pairs in non-nasal context and nasal coda contexts shows that the latter has smaller $A1 - P1n$ maximum and minimum, suggesting greater v–p opening for [i] followed by nasal coda than [i] followed by stop consonant. The slopes are also more negative for [i] with nasal coda, indicating faster increase of the v–p opening. For all pairs comparing i-n and iN, the maximum and/or minimum is smaller for the latter, suggesting a greater degree of nasalization. The nasal coda, with or without murmur, also causes a faster rate of nasalization in the vowel, which is indicated by the more negative slopes. However, there is no significant difference in the duration of nasalization.

4.2.3. *Vowel [ə]*

The values measured from $A1 - P0n$ to $A1 - P1n$ time contours for [ə] in different contexts are shown in Figs 7(a)–(c). The standard deviation is not shown in Fig. 7(a) to avoid confusion. There is no consistent trend in the range for $A1 - P0n$ and $A1 - P1n$; however, the trend of the maxima is the same for the two parameters. Fig. 7(b) shows that the slopes of $A1 - P0n$ and $A1 - P1n$ are negative for [ə] in nasal contexts but positive in the non-nasal context, indicating increasing nasalization into the vowel for the former. The pattern of the slopes for $A1 - P0n$ is similar to that for $A1 - P1n$ except for "ke-na". Fig. 7(c) shows that the trend for the duration of vowel nasalization as indicated by $A1 - P0n$ and $A1 - P1n$ is similar except for "deng-e". The duration of nasalization is between 25 and 90 ms.

The statistically significant differences of the maximum, minimum, and slope of $A1 - P0n$ or $A1 - P1n$ timing contour and the duration of nasalization for word pairs with [ə] are indicated in the lower section of Table III. The value obtained from $A1 - P0n$ or from $A1 - P1n$ is listed, depending on which parameter showed a greater difference between the words in the word pair. The values from $A1 - P1n$ are listed in italics. In the pairs comparing [ə] in non-nasal context and in nasal coda contexts, the latter have a lower maximum, especially from $A1 - P1n$, and a lower minimum, especially from $A1 - P0n$, suggesting a greater degree of nasalization for [ə] with nasal coda. The vowels with nasal coda also have more negative slope, suggesting a faster rate of nasalization into the vowel. For the pairs comparing [ə] followed by a nasal across a syllable boundary and by a nasal coda, $A1 - P0n$ shows the vowel with nasal coda to have a lower minimum, reflecting a greater degree of nasalization, except for ən-V. The $A1 - P0n$ slopes are more negative, indicating a faster rate of nasalization, for vowels with nasal coda. The period of nasalization is longer for [ə] followed by a nasal coda without oral closure than followed by a nasal initial.

5. Vowel formant frequencies

5.1. *Analysis*

The first three formant frequencies (F_1–F_3) measured throughout the vowel in the first syllable were used to determine the place of articulation of the nasal coda. The frequencies were determined from spectra taken every 10 ms within the vowel. The frequency of the harmonic with the maximum amplitude in the formant vicinity was used as a measure of the frequency of a formant. The end-frequencies and time-averaged frequencies over the vowel duration were used to test for differences distinguishing place

TABLE IV. The differences within word pairs (with the values of the second word subtracted from the values of the first word) of the first three formant frequencies at the end of the vowel [a], [i], and [ə] in the first syllable. The time-averaged frequency differences are shown below the end frequencies in italics. Statistical significance according to a one-tailed test is indicated

Word pair	End and *average* frequencies F_1 (Hz)	F_2 (Hz)	F_3 (Hz)
an-C_s	58	534*	262*
aŋ-C_s	*7*	*520**	*241**
an-V	– 76†	246*	4
aŋ-V	*– 93**	*335**	*33*
in-C_s	33	– 51	343*
iŋ-C_s	*46**	*– 101*	*265**
in-V	– 11	– 7	436*
iŋ-V	*– 8*	*20*	*418**
ən-C_s	– 21*	615*	– 34
əŋ-C_s	*– 16**	*571**	*26*
ən-V	– 107*	89*	– 94
əŋ-V	*– 26*	*98**	*– 2*

*$p < 0.001$
†$p < 0.005$.

of articulation for the nasal coda. Measurements were made for all of the repetitions for each word with Vn-C_s and Vŋ-C_s. For words with Vn-V and Vŋ-V, only those without nasal murmur were examined, so that the information regarding the intended nasal coda would have to be solely in the vowel. Those with nasal murmur are assumed to have the same properties as the vowel in VN-C_s.

5.2. *Results and discussion*

Once the presence of a nasal coda is detected, whether by the boundary detection technique or by the acoustic correlates of nasalization, the next step is to determine the type of nasal coda. The end and time-average formant frequencies of the vowel can be used to distinguish different nasal coda contexts, whether or not a nasal murmur is present. Table IV compares the word pairs with statistical significance according to a one-tailed test indicated. The average differences within word pairs, with the values of the second word subtracted from the values of the first word, across speakers and repetitions are shown. The time-average frequency differences are listed below the end-frequency differences in italics. The time-average frequency is examined since the crucial place information may not be present solely at the end of the vowel, especially when there is no oral closure for the nasal.

5.2.1 *Vowel [a]*

The differences in the formant frequencies were observed in the spectrograms as well as measured from the spectra throughout the vowel. The spectrograms in Figs. 2(a) and (b)

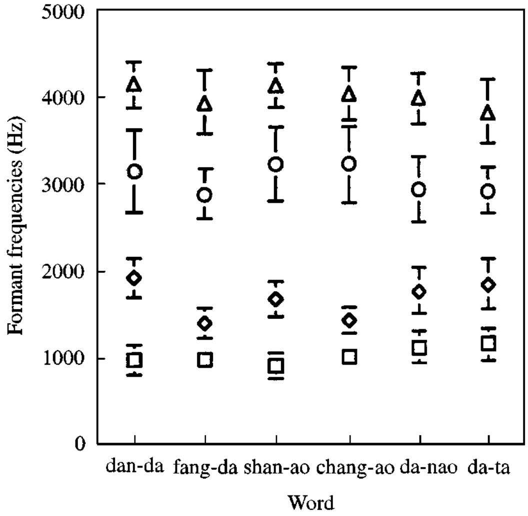

Figure 8. The first formant end or time-average frequencies, whichever shows greater differences among the words, for [a] across speakers and repetitions. F1.avg (□); F2.end (◇); F3.end (○); F4.end (△).

illustrate the differences in formant frequencies between "dan-da" and "fang-da" spoken by a male speaker, particularly for F_2 in this example. Fig. 8 shows the average formant frequencies for the end or time-averaged frequencies, whichever shows greater differences among the words. The average was across repetitions and speakers. According to the upper section of Table IV, [a] followed by [n] has F_2 and F_3 end and average frequencies greater than [a] with [ŋ] when murmur is present. Furthermore, the comparison of the second pair, an-V and aŋ-V with no oral closure for the nasal coda, indicates that F_1 is lower and F_2 is higher when an alveolar nasal is intended. The difference in F_2 is not as great as for the pair an-C_s and aŋ-C_s.

The formant frequencies in the vowel are related to anticipation of the nasal coda. The data for F_1 suggest that there is no significant difference in tongue height for the vowels with different nasal codas in aN-C_s, but in aN-V the tongue is higher when N is [n] than [ŋ]. From the F_2 values, which indicate the front–back tongue position of the vowel, it can be concluded that the tongue is further back in [a] of aŋ-C_s than for an-C_s. The tongue body is further forward for the alveolar consonant, as expected. The F_2 values are similar for vowel preceding the two alveolar consonants [n] and [t^h] in Fig. 8. Even for nasalized vowels without oral closure of the nasal coda, the tongue is further back for aŋ-V than for an-V, with F_2 frequencies higher for the latter. The differences for F_3 show trends similar to those for F_2, with the vowel in an-C_s having higher frequencies than those in aŋ-C_s. The formant frequency variability may also be influenced by the nasalization of the vowel and by the operation of phonological rules.

5.2.2 *Vowel [i]*

Fig. 9 shows the end or time-average formant frequencies for [i], whichever shows greater differences among the words. Averages are given across speakers and repetitions. The middle section of Table IV lists the end and time-averaged formant frequency

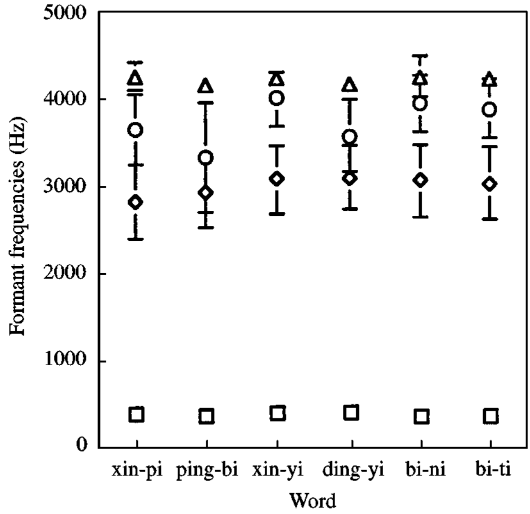

Figure 9. The first formant end or time-average frequencies, whichever shows greater differences among the words, for [i] across speakers and repetitions. F1.avg (□); F2.end (◇); F3.end (○); F4.end (△).

differences within word pairs with the vowel [i]. According to the figure and table, F_3 is higher in the vowel with [n] (and with [t^h]) than [ŋ], even without oral closure for the nasal consonant. The average frequency of F_1 can also be used for distinguishing in-C_s and iŋ-C_s.

The difference in formant frequencies within word pairs may indicate the position variation of the oral articulators. The slightly higher average F_1 in the vowel of in-C_s relative to the vowel of iŋ-C_s indicates that the tongue body is slightly higher in anticipation of [ŋ]. This result is opposite to that for an-V *vs.* aŋ-V, which shows a lower F_1 for the former, indicating vowel-type dependence of tongue height in nasal context anticipation. The F_2 differences for [i] show less anticipatory front–back position of the tongue body than for [a] since the higher and more fronted tongue position [i] makes anticipation more difficult. From Fig. 9, the F_2 and F_3 frequencies are closer to each other for [i] with nasal coda [ŋ] than [n], even without murmur. One probable explanation is that in anticipating [ŋ], the constriction is positioned so that the lowest resonance of the front cavity and the half-wavelength resonance of the back cavity are close in frequency. As a result, F_3 frequencies are lower for [i] preceding [ŋ] than before [n].

5.2.3 *Vowel [ə]*

The bottom section of Table IV shows the end and time-averaged frequency differences within word pairs with the vowel [ə]. Fig. 10 shows the end or time-average formant frequencies, averaged across speakers and repetitions. Whether or not the oral closure is formed for the nasal coda, the end and time-averaged F_2 frequencies are higher for [ə] followed by [n] than [ŋ]. The difference is greater by about 500 Hz when the nasal consonant is followed by a stop consonant. For these words, there is always a nasal murmur following the vowel [ə]. On the other hand, the first formant frequency, especially the end frequency, is lower for [ə] with nasal coda [n] than [ŋ]. There is no significant difference for the F_3 frequencies.

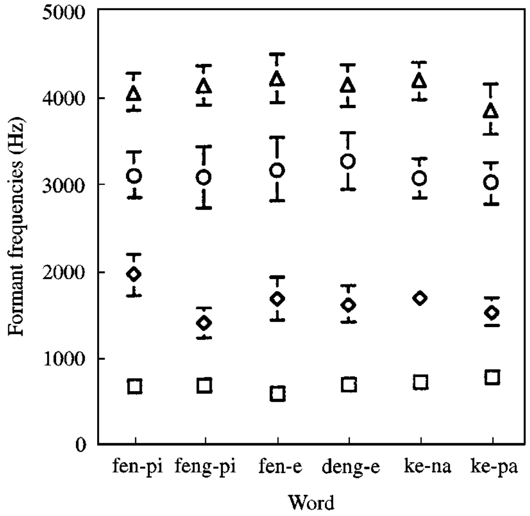

Figure 10. The first four formant end or time-average frequencies, whichever shows greater differences among the words, for [ə] across speakers and repetitions. F1.end (□); F2.avg (◇); F3.avg (○); F4.end (△).

The formant frequencies reflect the tongue-body position. As shown in Fig. 10, the end F_1 frequency of [ə] with nasal coda [n] is lower than with nasal coda [ŋ] when there is no oral closure for the nasal coda, indicating a higher tongue position toward the end of the vowel with an intended [n]. The trend found in the F_1 end frequency differences in Table IV is only partly reflected in the F_1 average frequencies, suggesting that anticipatory coarticulation of the tongue height does not involve the entire vowel duration, especially when the nasal is [ŋ]. On the other hand, the end and average F_2 frequency differences show a similar trend, indicating that anticipation of the front–back tongue body position involves the entire vowel duration and not just the end of the vowel. When [ə] is followed by [n], the F_2 frequency is higher than it is with following [ŋ], even without oral closure, due to the tongue body moving to the front in anticipation for [n] but with a more backward movement for [ŋ].

In summary, one or more of the first three formant frequencies taken at the end of the vowel or time-averaged throughout the vowel showed statistically significant difference for the two types of nasal codas. This result holds whether or not the nasal coda is manifested acoustically as a nasal murmur. Therefore, it may be possible to determine the place of the nasal coda solely from the preceding vowel. An attempt is also made to link the acoustic results to articulation in order to refine speech modeling of vowels in nasal contexts.

6. Conclusion

The presence of nasal codas in Standard Chinese can be detected based on the vowel–nasal consonant boundary if oral closure is formed for the nasal consonant. The boundary between the vowel and the nasal is determined objectively by using the maximum first-difference amplitude obtained from the first four formant time contours in the vowel and the nasal consonant. In fact, the boundary measured according to the quantitative method suggested in this study is highly correlated with the boundary measured subjectively, with an average difference of 7 ms. As shown in Table I, significant one-tailed test results indicate that the maximum first difference of any one of the four formant amplitudes can be used to distinguish a vowel followed by a nasal consonant from a non-nasal vowel, except for $A3$ in aŋ-Cs and $A1$ in iŋ-C_s. On average, the nasal coda can also be distinguished from a syllable-initial [n] across a syllable boundary by the V : N boundary detection technique. The sum of the maximum $A1$–$A4$ first differences may be used, independently of vowel type. As shown in the last column of Table I, the sum is 2–10 dB greater when [n] is syllable-initial than when it is a nasal coda [n], and the sum is 17–22 dB greater than when it is a nasal coda [ŋ], indicating a smoother transition from vowel to nasal within the syllable. Note that the values mentioned above are obtained from instances when a boundary was known *a priori*. Although the result shows a possible method of detecting a V : N boundary, to determine absolute cutoff values, training data and test data of different vowel types and contexts need to be measured,.

It was found that, in some cases of VN-V, there is no nasal murmur due to the lack of closure for the nasal coda, especially when the utterance has a low vowel and nasal coda [n]. For those cases, more than the V : N boundary detection technique is needed to indicate the presence of the underlying nasal coda. The detection of the nasal coda when the oral closure is not formed can be done by using time contours of normalized $A1 - P0$ and $A1 - P1$ (amplitude differences between the first formant and nasal peaks), $A1 - P0n$ and $A1 - P1n$. Similarities in these contours among the vowel types were observed from the maximum and minimum values, average rate of change (indicated by the slope), and time duration. As shown by the maximum, minimum, and slope of $A1$–$P0n$ and $A1 - P1n$ in Table III, all three vowels, [a], [i], and [ə] in nasal coda contexts were found to have greater degree and faster rate of nasalization than the corresponding vowels in non-nasal contexts. Comparisons of vowels followed by a nasal coda, with or without murmur, and vowels followed by syllable-initial [n] showed that anticipation of the nasal is weaker in the case of the latter, as indicated by a smaller degree of nasal coupling, slower rate of nasalization, and/or shorter duration of vowel nasalization. The significant differences suggest that the acoustic correlates of nasalization discussed here are promising parameters for detecting a nasal coda.

There are also differences in the $A1 - P0n$ and $A1 - P1n$ time contours according to vowel types. From the $A1 - P0n$ slope within the vowel in a nasal context in Figs 5(b) and 7(b), [ə] has more negative slope than [a], reflecting a slower rate of nasalization for the low vowel. On the other hand, from the slope of $A1 - P1n$ in Figs 6(b) and 7(b), [i] has more negative slopes than [ə], reflecting a faster rate of nasalization for the high vowel. These results correspond to findings from Velotrace measurements (Krakow, 1993) that the rate of velum change is faster for a high vowel than for a low vowel. The patterns of vowel nasalization duration from $A1 - P0n$ of [a], $A1 - P1n$ of [i], and $A1 - P1n$ of [ə] are the same for the different vowel contexts, as shown in Figs 5(c), 6(c), and 7(c), respectively; however, the range of nasalization duration for [i] (30–60 ms) is less than for [a] (50–100 ms) with the range for [ə] (25–90 ms) being mostly in between. One can conclude that a nasalized low vowel has a longer period of nasal coupling than a high vowel. This result corresponds to the finding by Clumeck (1976) who used a nasograph, a photoelectric device, to measure the velopharyngeal opening. He found that low vowels show lower velar height and longer nasalization than high vowels.

One cue for the place of articulation of the nasal coda is based on the vowel–nasal consonant boundary detection technique of determining the maximum first differences over the vowel and nasal consonant for the first four formant amplitudes. A simple measure is the sum of the maximum $A1-A4$ first differences, independent of vowel type. As shown in the last column of Table I, for a given vowel the change at the boundary is smaller when the nasal coda is [ŋ] than [n] with the average sums differing by at least 11 dB. However, this technique can only be used to distinguish nasal types when the oral closure is formed.

When the murmur is not produced, only the vowel contains information about the place of the intended nasal ending. The parameters measured from the first three formant frequency (F_1–F_3) time contours of the vowel can be used as cues for nasal place of articulation. In comparing vowels followed by [n] as opposed to [ŋ], with or without oral closure, F_2 frequency is higher for [a]; F_3 frequency is higher for [i]: and F_1 frequency is lower and F_2 frequency is higher for [ə]. This result is consistent with other studies of SC monosyllables (Ren, 1988; Lin & Yan, 1991). Vowels with nasal coda realized with nasal murmur show greater statistical variation for F_2 across different places of the nasal coda than do those without nasal murmur. The F_1 and F_2 frequencies of the formants may be explained by the tongue moving to a higher and a more fronted position for [a] and [ə] in anticipation of [n] relative to [ŋ]. The F_3 frequency is influenced by the closeness of front and back cavity resonances for [i]. One concern is that the C_s in "xin-pi", "ping-bi", "fen-pi", and "feng-pi" is a labial, which may influence the nasal coda, since labials tend to spread their place features onto the preceding nasal consonant (Stanley, 1967; Wright, 1975). However, results of this study indicate that the effect of the labial consonant on the nasalized vowel is minimal in Standard Chinese.

In summary, this study of simple vowels in Standard Chinese disyllables reveals the distinctiveness of vowels with a nasal coda and suggests a possible direction for developing an algorithm for identifying the nasal coda and its place of articulation. The presence of a nasal coda may be detected by the boundary detection technique if the oral closure is formed for the nasal. It can be further verified by using the acoustic correlates $A1 - P0n$ and $A1 - P1n$ to quantify nasalization in the vowel. The acoustic correlates can also detect the presence of a nasal coda when there is no nasal murmur. The place of articulation of the nasal coda may be determined by the first three formant frequencies of the vowel. If the oral closure is formed for the nasal, the maximum first difference of the formant amplitude can also verify the place of articulation.

This research was supported by MIT International Science and Technology Initiative. Special thanks to my mentors in the Laboratory of Phonetics at the Chinese Academy of Social Sciences. My deepest gratitude to Kenneth N. Stevens for his thorough editing of the paper. I am also grateful for the helpful comments by the editor and the reviewers of this paper.

References

Bell-Berti, F. (1993) Understanding velic motor control: studies of segmental context. In *Nasals, nasalization, and the velum: phonetics and phonology V* (M. K. Huffman & R. A. Krakow, editors), pp. 63–85. New York: Academic.

Chen, M. (1973) Cross-dialectal comparison: a case study and some theoretical considerations, *Journal of Chinese Linguistics*, **1**(1), 39–63.

Chen, M. (1975) An areal study of nasalization in Chinese. In *Nasálfest* (C. A. Ferguson, L. M. Hyman & J. J. Ohala, editors), pp. 81–125. Stanford: Language Universals Project, Dept. of Linguistics, Stanford University.

Chen, M. & Wang, W. S. (1975) Sound change: actuation and implementation, *Language*, **51,** 255–281.

Chen, M. Y. (1997) Acoustic correlates of English and French nasalized vowels, *Journal of the Acoustical Society of America*, **102**(4), 2360–2370.

Clumeck, H. (1976) Patterns of soft palate movements, *Journal of Phonetics*, **4,** 337–351.

Delattre, P. (1968) From acoustic cues to distinctive features, *Phonetica*, **18,** 198–230.

Fujimura, O. (1960) *Spectra of nasalized vowels.* Massachusetts Institute of Technology Quarterly Progress Report, Research Laboratory of Electronics, Vol. 58, pp. 214–218.

Fujimura, O. & Lovins, J. (1978) Syllables as concatenative phonetic units. In *Syllables and segments* (A. Bell & J. Hooper, editors), pp. 107–120. Amsterdam: North-Holland.

Hajek, J. (1997) *Universals of sound change in nasalization*. Oxford: Blackwell.

Hanyu fangyin zihui [*Syllabarics of the Chinese dialects*] (1989). 2nd edition. Beijing: Wenzi Gaige Chubanzhe.

Harrington, J. (1994) The contribution of the murmur and vowel to the place of articulation distinction in nasal consonants, *Journal of the Acoustical Society of America*, **96,** 19–32.

House, A. S. (1957) Analog studies of nasal consonants, *Journal of Speech and Hearing Disorders*, **22,** 190–204.

Krakow, R. A. (1993) Nonsegmental influences on velum movement patterns: syllables, sentences, stress, and speaking rate. In *Nasals, nasalization, and the velum: phonetics and phonology V* (M. K. Huffman & R. A. Krakow, editors), pp. 87–116. New York: Academic Press.

Kurowski, K. & Blumstein, S. (1984) Perceptual integration of the murmur and formant transitions for place of articulation in nasal consonants, *Journal of the Acoustical Society of America*, **76,** 383–390.

Kurowski, K. & Blumstein, S. (1987) Acoustic properties for place of articulation in nasal consonants, *Journal of the Acoustical Society of America*, **81,** 1917–1927.

Larkey, L. S., Wald, J. & Strange, W. (1978) Perception of synthetic nasal consonants in initial and final syllable position, *Perceptual Psychophysiology*, **23,** 299–312.

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.

Maeda, S. (1993) Acoustics of vowel nasalization and articulatory shifts in French nasal vowels. In *Nasals, nasalization, and the velum: phonetics and phonology V* (M. K. Huffman & R. A. Krakow, editors), pp. 147–167. New York: Academic Press.

Malécot, A. (1956) Acoustic cues for nasal consonants: an experimental study involving a tape-splicing technique, *Language*, **32,** 274–284.

Mermelstein, P. (1977) On detecting nasals in continuous speech, *Journal of the Acoustical Society of America*, **61,** 581–587.

Miller, J. L. (1977) Nonindependence of feature processing in initial consonants, *Journal of Speech and Hearing Research*, **20,** 519–528.

Moll, K. L. & Daniloff, R. G. (1970) Investigation of the timing of velar movements during speech, *Journal of the Acoustical Society of America*, **50,** 678–684.

Ohala, J. J. (1975) Phonetic explanations for nasal sound patterns. In *Nasálfest* (C. A. Ferguson, L. M. Hyman & J. J. Ohala, editors), pp. 289–316. Stanford: Language Universals Project, Dept. of Linguistics, Stanford University.

Ohde, R. N. (1994) The development of the perception of cues to the [m]-[n] distinction in CV syllables, *Journal of the Acoustical Society of America*, **96,** 675–686.

Pickett, J. M. (1965) Some acoustic cues for synthesis of the n-d distinction, *Journal of the Acoustical Society of America*, **38,** 474–477.

Raphael, L. J., Dorman, M. F., Freeman, F. & Tobin, C. (1975) Vowel and nasal duration as cues to voicing in word-final stop consonants: spectrographic and perceptual studies, *Journal of Speech and Hearing Research*, **18,** 389–400.

Recasens, D. (1983) Place cues for nasal consonants with special reference to Catalan, *Journal of the Acoustical Society of America*, **73,** 1346–1353.

Ren, H. M. (1988) *Temporal structure of the syllable with nasal ending*. Report of Phonetic Research, Institute of Linguistics, (CASS), Beijing, China, pp. 53–79.

Repp, B. H. (1986) Perception of the [m]-[n] distinction in CV syllables. *Journal of the Acoustical Society of America*, **79,** 1987–1999.

Repp, B. H. & Svastikula, K. (1988) Perception of the [m]-[n] distinction in VC syllables. *Journal of the Acoustical Society of America*, **83,** 237–247.

Seitz, P. F., McCormick, M. M., Watson, I. M. C. & Bladon, R. A. (1990) Relational spectral features for place of articulation in nasal consonants, *Journal of the Acoustical Society of America*, **87,** 351–358.

Stanley, R. (1967) Redundancy rules in phonology, *Language*, **43,** 393–436.

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

Trigo Ferré, R. L. (1988) *The phonological derivation and behavior of nasal glides*. Doctoral dissertation, Massachusetts Institute of Technology.

Vaissière, J. (1988) Prediction of velum movement from phonological specifications, *Phonetica*, **45,** 122–139.

Wright, J. (1975) Nasal-stop assimilation: testing the psychological reality of an English MSC. In *Nasálfest* (C. A. Ferguson, L. M. Hyman & J. J. Ohala, editors), pp. 389–397. Stanford: Language Universals Project, Dept. of Linguistics, Stanford University.