Comparing Tongue Positions of Vowels in Oral and Nasal Contexts

Takayuki Arai
Department of Electrical and Electronics Engineering
Sophia University, Tokyo, Japan
arai@sophia.ac.jp

Abstract

We studied the formant frequency shifts of a nasalized vowel from the aspects of acoustics and production. For the acoustic aspect, formant frequencies were measured in the following three cases: 1) vowels produced by physical models of a vocal tract, 2) nasal and oral vowels intentionally uttered by a human speaker, 3) nasalized vowels uttered in a nasal context. As predicted by the acoustic theory and the perceptual findings, bidirectional formant shifts in F1 frequency were observed by acoustic analyses: increasing F1 for high vowels and decreasing F1 for low vowels. Then, we tried to answer to the next question, that is, whether or not speakers and/or listeners compensate for the formant shifts in production and/or in perception. The perceptual experiment by Arai [J. Acoust. Soc. Am., 115, p.2541 (2004)] showed that compensation occurs when an isolated vowel has nasalization and is accompanied by formant transitions. This result agrees with the findings of Krakow et al. [J. Acoust. Soc. Am. 83, 1146-1158 (1988)] which reported that the formant shifts due to nasalization do not necessarily lead to the misperception of a vowel in a word with nasal context. In the production experiment, the measurement of the positions of the articulators showed almost no compensation except for the lowest vowel /ɑ/.

Introduction

The nasal tract couples to the main vocal tract due to several reasons; one of them is due either to anatomical or functional problems (Stevens et al., 1986). Cleft palate patients, for example, often have velopharyngeal insufficiency that causes hypernasality. Inadvertent nasalization, a speech disorder with the velopharyngeal port opened excessively during vowel production, is also one of the most common problems of deaf speakers (Stevens, et al., 1976; Chen, 1995). Inability to decouple the nasal cavities from the pharyngeal and oral cavities will result in severely distorted speech, and the ability to control coupling of the nasal cavities to the vocal tract is crucial for the production of normal speech (Bell-Berti, 1980).

Some languages, such as French, Portuguese and Hindi, distinguish phonemically between nasal and nonnasal vowels, whereas other languages, including English, do not have a phonemic distinction between an oral vowel and its nasal counterpart. Even in the latter case, however, nasal coupling occurs during vowels adjacent to nasal consonants. Vowels next to a nasal consonant are often nasalized due to an overlap in gestures of the velum and the tongue or lips. If the vowel is preceded by an obstruent consonant and followed by a nasal consonant, the velopharyngeal port is closed at the time of release of the obstruent consonant; the port then opens during the vowel in preparation for the formation of the oral closure for the nasal consonant (Stevens, 1998). This results in overlapping gestures, and therefore, the preceding vowel becomes nasalized before a nasal consonant.

The degree of acoustic coupling between the vocal and nasal tracts is controlled by the velum, or soft palate, as well as the posterior and lateral pharyngeal walls (Chen, 1997). The opening to the nasal tract allows airflow through the nose as well as through the mouth, and the acoustic coupling causes the vowel to be nasalized. Therefore, a simple model for a nasalized vowel could be a main vocal tract with a side branch, and different degrees of opening of the velopharyngeal port control different degrees of nasalization. According to acoustical theory (Fant, 1960; Fujimura, 1960, 1961; Fujimura and Lindqvist, 1971; House and Stevens, 1956), the basic difference between the transfer function for the vocal tract with a side branch and that for a nonnasal vowel is that additional poles and zeros are introduced to the vocal-tract transfer function as a consequence of acoustic coupling to the nasal tract.

The additional poles and zeros due to nasal coupling cause modifications in the spectrum, such as reduction in amplitude of the first formant (F1), broadening the bandwidth of F1, shifting F1 upwards in frequency, and a relative strengthening of the spectrum in the vicinity around 250 Hz (House and Stevens, 1956; Hattori et al., 1958; Fant, 1960; Fujimura, 1960; Fujimura and Lindqvist, 1971; Hawkins and Stevens, 1985; Maeda, 1993). The higher frequencies may also be modified by nasal coupling. The main effect of nasalization, however, is the perturbation of the low-frequency spectrum by replacing the first formant with a shifted F1 (F1'), a nasal formant (Fn), and a nasal zero (Fz) (Fant, 1960; Fujimura and Lindqvist, 1971; Stevens et al., 1986). As the cross-sectional area of the velopharyngeal opening is gradually increased, the spacing between the pole and zero introduced in the vicinity of the first formant increases, F1 frequency shifts, and F1 bandwidth increases.

Calculations of the acoustic consequences of nasal coupling predict distinctively different modifications depending on vowel identity (Fujimura, 1960; Fujimura & Lindqvist, 1971). For high vowels, the theory predicts F1 shifts upwards in frequency, and a nasal formant appears in the spectral valley between F1 and F2. For low vowels, F1 also shifts upwards in frequency. At the same time, F1, however, comes close to the zero, because the first pole-zero pair is lower in frequency than F1 in the corresponding oral vowel, and as a result, F1 is weakened and seemingly split into two peaks (Stevens et al., 1986). At low degrees of coupling, the nasal pole (Fn) is almost canceled by the nasal zero (Fz), and in this case, the prominence of Fn is small. At higher degrees of coupling, on the other hand, Fn increases in prominence (Fujimura and Lindqvist, 1971), and Fn could be identified as a formant (Maeda, 1993).

Thus, nasal coupling can shift F1 frequency and this may affect perceived vowel height. Due to the upwards shift in F1 frequency, nasalization might be expected to lower perceived vowel height (Ohala, 1986). In addition, due to the prominence of Fn, nasalization might be expected to raise perceived vowel height for low vowels. This seems to be a possible explanation for the bidirectional shifts in perceived nasal vowel height (Krakow et al., 1988; Stevens, 1998) that were observed in the perceptual experiments of previous studies (Wright, 1975, 1986).

This F1 frequency shift might lead to listener misperceptions of vowel height. Krakow et al. (1988) examined the hypothesis that perceived nasal vowel height is not entirely determined by the spectral shape of the nasal vowel, but rather the context in which the nasal vowel occurs can affect the way in which the nasalization of that vowel is perceived. This hypothesis was tested by comparing listeners' perception of nasal vowels in the presence and absence of an adjacent nasal consonant, and Krakow et al. (1988) showed that nasal coupling does not necessarily lead to listener misperceptions of vowel quality when the vowel's nasality is coarticulatory in nature.

In this study, there are two major goals. The first goal of this study is to confirm through acoustic analyses the manner in which formant frequencies shift due to nasalization, especially the bidirectional shifts in F1 frequency. Even though both the acoustic theory and the perceptual findings show the bidirectional movements of the F1 frequency, there are not many references in the literature which show clear acoustic evidence for the bidirectionality. We, therefore, confirm that the F1 frequency tends to shift in a more central direction when nasalized based on several measurements of formant frequencies for various vowels.

In English and many other languages, vowels should be perceived as the same phoneme regardless of nasalization. In other words, a speaker and/or a listener might tend to compensate for such formant shifts. The next question is, therefore, whether or not compensation exists in production and/or in perception.

For the perception side, we predicted in Arai (2004) that it might be possible to see the compensation as long as enough information is available for a listener to determine that the vowel is next to a nasal consonant even if the vowel is isolated. This prediction was based on the study by Krakow et al. (1988) for compensation in nasal contexts. In their study, oral and nasal vowels were embedded in oral and nasal contexts and the compensation due to the context was tested by using an articulatory synthesizer. In fact, the vowel itself usually contains contextual information in natural speech, such as formant transitions, and we do not know if such contextual information within a vowel is enough to yield this compensation. Arai (2004) therefore conducted a perceptual experiment and confirmed that this compensation effect still exists in isolated vowels as long as they contain contextual information such as formant transitions.

As the second goal of this study, we investigate the compensation effect for the production side. We measure the positions of the articulators, especially tongue height, and compared them in oral and nasal contexts using the Electromagnetic midsagittal articulometer (EMMA) system (Perkell et al., 1992).

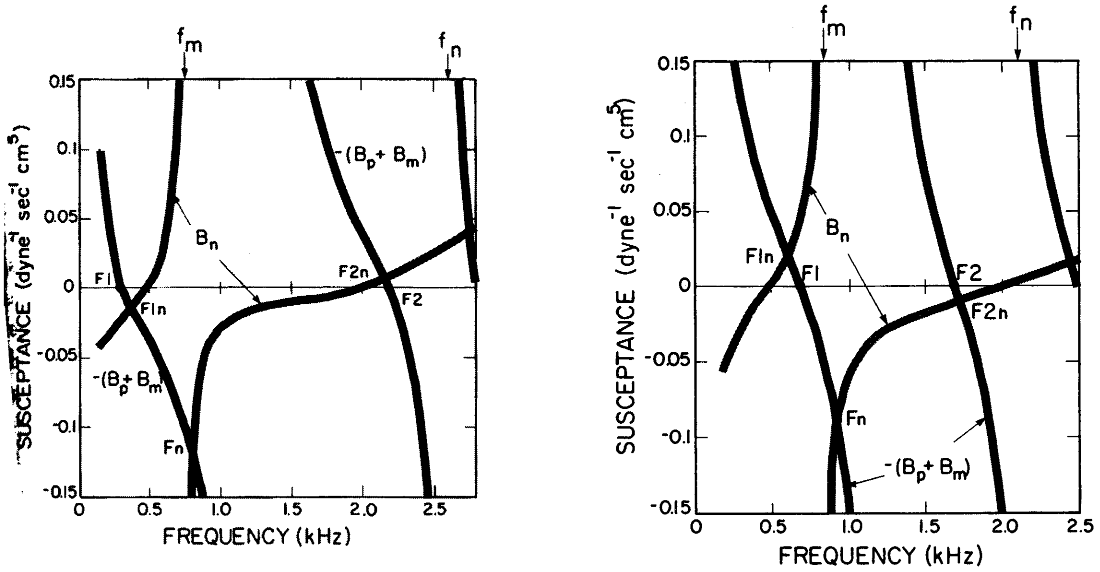

Fig. 1. Estimated curves of susceptance vs. frequency of nasalized vowels: the vowel [i] with a velopharyngeal area of 30 mm^2 (top panel), and the vowel [æ] with a velopharyngeal area of 50 mm^2 (bottom panel) (adapted from Stevens, 1998).

Theoretical considerations

The shift in formant frequencies due to nasal coupling during uttering a vowel can be predicted by the acoustic theory of nasalized vowels. The velopharyngeal opening when the main vocal tract is in a vowel-like configuration introduces poles and zeros in the transfer function of the vocal tract. These additional poles and zeros are approximated by applying an electric-circuit analog based on a simple model (Fujimura, 1960; Fujimura and Lindqvist, 1971; Stevens, 1998). In this model, a side branch is attached to the main vocal tract, and the susceptances looking in different directions from the velopharyngeal port are examined. This model has one input volume velocity (Us) and two outputs: the volume velocity Um at the mouth and the volume velocity Un at the nose (define U as volume velocity). Because the sound pressure at a distance is a result of the combined output Um+Un, the transfer function becomes (Um+Un)/Us = Um/Us + Un/Us. The transfer functions of Um/Us and Un/Us have the same poles but different zeros.

Let Bn, Bm and Bp be the susceptances looking into the nose, the mouth, and the pharynx, respectively. The poles of the transfer functions are located at frequencies where Bn+Bm+Bp = 0. Therefore, the natural frequencies of the entire system can be determined by the graphical analysis, that is, finding the intersections of the curves of Bn vs. frequency and –(Bm+Bp) vs. frequency as shown in Fig. 1. The Bn curve is determined by the velopharyngeal port and the nasal cavity. As the velopharyngeal port is opening, the Bn curve is shifting, whereas the –(Bm+Bp) curve does not change as long as the main tract shape is kept constant. The zeros of Bn occur at the natural frequencies of the nasal cavity with a closed velopharyngeal port. Based on the measured transfer function of the nasal tract from above the closed velopharyngeal port to the nostril output, as measured using a sweep-tone source (Lindqvist and Sundberg, 1972), the susceptance Bn is estimated to have zeros at about 500 and 2000 Hz (Chen, 1997; Stevens, 1998). Therefore, those frequencies are set to 500 and 2000 Hz in Fig. 1. On the other hand, the –(Bm+Bp) curve highly depends on the main vocal-tract shape. The zeros of Um/Us are located at frequencies for which Bn = ∞, while the zeros of Un/Us are located at frequencies for which Bm = ∞.

Fig. 2. Physical models of the human vocal tract with the nasal cavity for [i] (left) and [ɑ] (right).

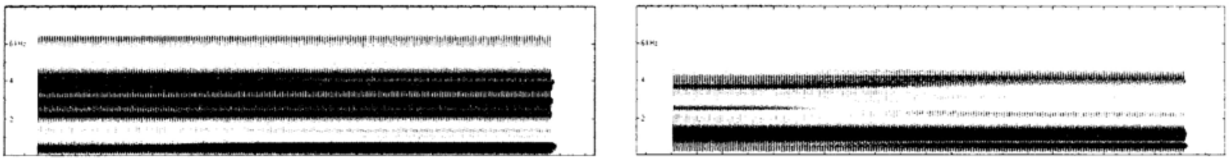

Fig. 3. Spectrograms for the physical models of [i] (top) and [ɑ] (bottom). In each case, the velopharyngeal port was first closed, and then opened in the middle of the utterance.

This model predicts how F1 is replaced by F1’ where Fn and Fz depend on vowel height (Stevens, 1998). For all vowel types, as the area of the velopharyngeal port increases, an extra pole-zero pair (Fn and Fz) starts to appear near 500 Hz and shifts upwards, and the spacing of the pole-zero pair gets wider. An increase in coupling also corresponds to an upward shift of F1 frequency. For the vowel /i/, as the degree of nasal coupling continuously increases, F1’ gradually shifts upwards from F1 of the non-nasal vowel and reaches a frequency lower than 500 Hz; Fn gradually shifts upwards from 500 Hz and reaches a frequency lower than 1 kHz; and Fz shifts upwards from 500 Hz to a frequency higher than 1 kHz. As a result, an extra pole occurs in the spectral valley between F1 and F2 in the spectrum of a nasalized vowel /i/. For the vowel /ɑ/, as the degree of nasal coupling continuously increases, things are a little complicated (Stevens, 1998; Maeda, 1993). F1’ gradually shifts upwards from F1 of the non-nasal vowel, but at the same time, Fz rapidly shifts upwards from 500 Hz. On the other hand, Fn gradually shifts upward from 500 Hz and reaches a frequency lower than F1 of the non-nasal vowel. As a result, F1' is getting weakened by Fz, and Fn is gradually becoming dominant. Consequently, F1' first shifts upwards; then the F1 region has two peaks of F1' and Fn; and finally, Fn acts as F1, as Fn is dominant.

Acoustical analyses

As we discussed previously, the acoustic theory shows that F1 tends to shift in a more central direction when nasalized. To confirm the formant shifts, especially the bidirectional shifts in F1, due to nasalization, we conducted acoustic analyses on three sets of vowels: A) mechanically produced nasalized vowels, B) intentionally repeated oral-nasal continuum, and C) naturally produced vowels in nasal and non-nasal contexts.

Mechanically produced nasalized vowels

To clearly observe the poles and zeros, we first made physical models of the human vocal tract for [i] and [ɑ] with the nasal cavity as shown in Fig. 2. This model is supposed to have less acoustic loss than the human vocal tract, so we can obtain peaks with less bandwidth. Each of the models is made from five acrylic plates. The center (black) plate is 1-cm thick and has a schematic midsaggital cross-section for each vowel. On both sides of the center plate, there are two transparent 3 cm thick plates. These plates have holes to achieve the proper area functions for the vocal-tract configuration of the vowels [i] and [ɑ] with nasal cavities. The outer-layered plates are 1-cm thick, and are also transparent.

The velum is made of rubber and may be rotated around a pivot located roughly at the boundary of the soft and hard palates. This movable velum acts as the velopharyngeal port and allows us to simulate different degrees of nasal coupling. The velopharyngeal opening is controlled by the rotating valve.

Method

With the physical models of the human vocal tract, we mechanically produced oral and nasalized vowels. As a sound source, we used the KLGLOTT88 voicing source model (Klatt and Klatt, 1990) with the default settings. Its fundamental frequency was fixed at 100 Hz. The velopharyngeal port was first closed, and then opened in the middle of the utterance. The final areas of the velopharyngeal port were approximately 30 mm^2 for [i] and 50 mm^2 for [ɑ].

The sound source was played from a laptop computer via the digital-to-analog (D/A) converter of a digital audio amplifier (Onkyo, MA-500U) that was connected to the laptop computer with the USB interface. The sampling frequency was 16 kHz. The amplifier then drove a driver unit (TOA, TU-750) used for a horn speaker. To avoid unwanted coupling between the neck and the area behind the neck of the driver unit, we inserted a close-fitting hard rubber cylindrical filler inside the neck. We made a hole in the center of the rubber filling with an area of 30 mm^2. The neck part was attached to the glottal end of the vocal tract models.

The utterances were recorded with an Electro-Voice model 054 omnidirectional dynamic microphone and a pre-amplifier (Shure Professional Microphone Mixer). The microphone was placed approximately 20 cm in front of the model's lips in a partially sound-attenuated room. For this microphone position, the distances from the microphone to the mouth and to the nose were about equal. All signals were digitized with sampling frequency of 16 kHz.

Results and discussion

It has been reported that the velum is higher in high vowels and lower in low vowels (Bell-Berti, 1980). To achieve the same degree of perceived nasality, we need a greater velopharyngeal opening for low vowels, whereas high vowels only need a small opening (House and Stevens, 1956). Thus, it was reasonable that we used different final areas of the velopharyngeal opening for the vowels [i] and [ɑ] (they were approximately 30 mm^2 and 50 mm^2, respectively).

Figure 3 shows the spectrograms for [i] and [ɑ]. For the vowel [i], one can observe that the bandwidth of F1 becomes wider when the velopharyngeal port opens. Simultaneously, the upward shift of the F1 frequency was also observed. Furthermore, the extra pole Fn between F1' and F2 was observable. For the vowel [ɑ], one can observe a pole-zero pair associated with the velopharyngeal opening below the original F1 frequency. The frequencies of the pole and zero increased as the velopharyngeal opening became wider. At the same time, the distance between the pole and zero also increased as the velopharyngeal opening became wider. Eventually, the extra pole below the original F1 becomes dominant as predicted by the acoustic theory.

Intentionally repeated oral-nasal continuum

A similar analysis was done for natural vowels by a real speaker. In this case, oral and nasal vowels were intentionally produced. Because there might be a small variation for each utterance in natural vowels, several repetitions were made.

Method

A speaker was asked to produce oral and nasalized vowels continuously by first raising, then lowering the velum. The utterance was repeated 12 times for vowels [i] and [ɑ]. During each repetition, the speaker maintained the main vocal-tract shape for [i] or [ɑ] with a stationary tongue configuration.

The utterances were recorded in a partially sound-attenuated room with an Electro-Voice model 054 omnidirectional dynamic microphone and a pre-amplifier (Shure Professional Microphone Mixer). The microphone was again placed approximately 20 cm in front of the speaker's lips to get equal distances from the microphone to the mouth and to the nose. All signals were digitized with a sampling frequency of 16 kHz.

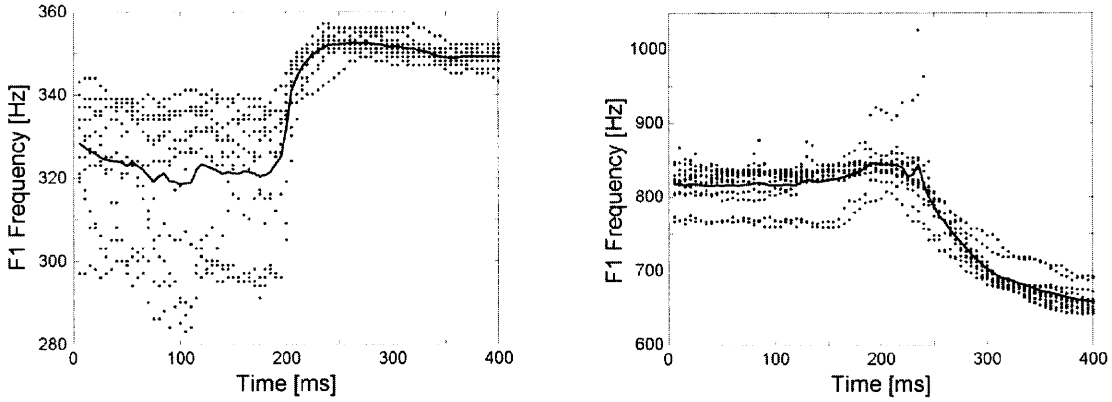

Fig. 4. Measured F1 frequency vs. time for utterances that transition from an oral vowel to a nasalized vowel: (a) [i], and (b) [ɑ]. The dots represent the measured F1 frequencies at each frame among 12 repetitions overlaid on top each other. The solid line indicates the average F1 trajectory over 12 repetitions.

Formant tracking of the vowels was done by using LSPECTO in XKL, which is a revision of the software package developed by Klatt (1984). The formant-tracking algorithm was based on 20th order linear predictive coding (LPC). For each frame, a 25-ms Hamming window was used, and each frame output was generated every 5 ms.

Results and discussion

Figure 4 shows the F1 trajectories as a function of time. The dots represent the measured F1 frequencies at each frame among 12 repetitions overlaid on top each other. The solid line indicates the average F1 trajectory over 12 repetitions. As shown in this figure, F1 frequency shifted upwards for the high vowel [i] as nasal coupling occurred; this was consistent with the acoustic theory. The F1 frequency moved roughly from 320 Hz to 350 Hz.

For the low vowel [ɑ], on the other hand, the F1 movement was complicated as also predicted by the acoustic theory. The average line shows the upward movement in the F1 frequency with a small velopharyngeal opening, whereas the line goes down with a larger velopharyngeal opening. Finally, the F1 frequency moved between the range of approximately –160 to +30 Hz.

Naturally produced vowels in oral and nasal contexts

To investigate the formant shifts of natural vowels in real speech, we analyzed a speaker's vowels in oral and nasal contexts.

Method

The speech samples were mono-syllabic nonsense words bVC uttered by a male native speaker of

American English. The vowel V was either /i/, /ɪ/, /ɛ/, /ʌ/, /æ/, or /ɑ/; and the consonant C was either /b/ or /m/. The target words were embedded in the carrier phrase "Say ________, again." All 12 combinations were repeated five times in a random order (60 utterances, in total). The utterances were recorded in a partially sound-attenuated room with an Electro-Voice model 054 omnidirectional dynamic microphone and a pre-amplifier (Shure Professional Microphone Mixer). The microphone was placed approximately 20 cm in front of the speaker's lips to get equal distances from the microphone to the mouth and to the nose. All signals were digitized with a sampling frequency of 16 kHz. Formant tracking of the vowels was done by using LSPECTO, which was based on the 20th order LPC (25-ms Hamming window and 5-ms frame shift).

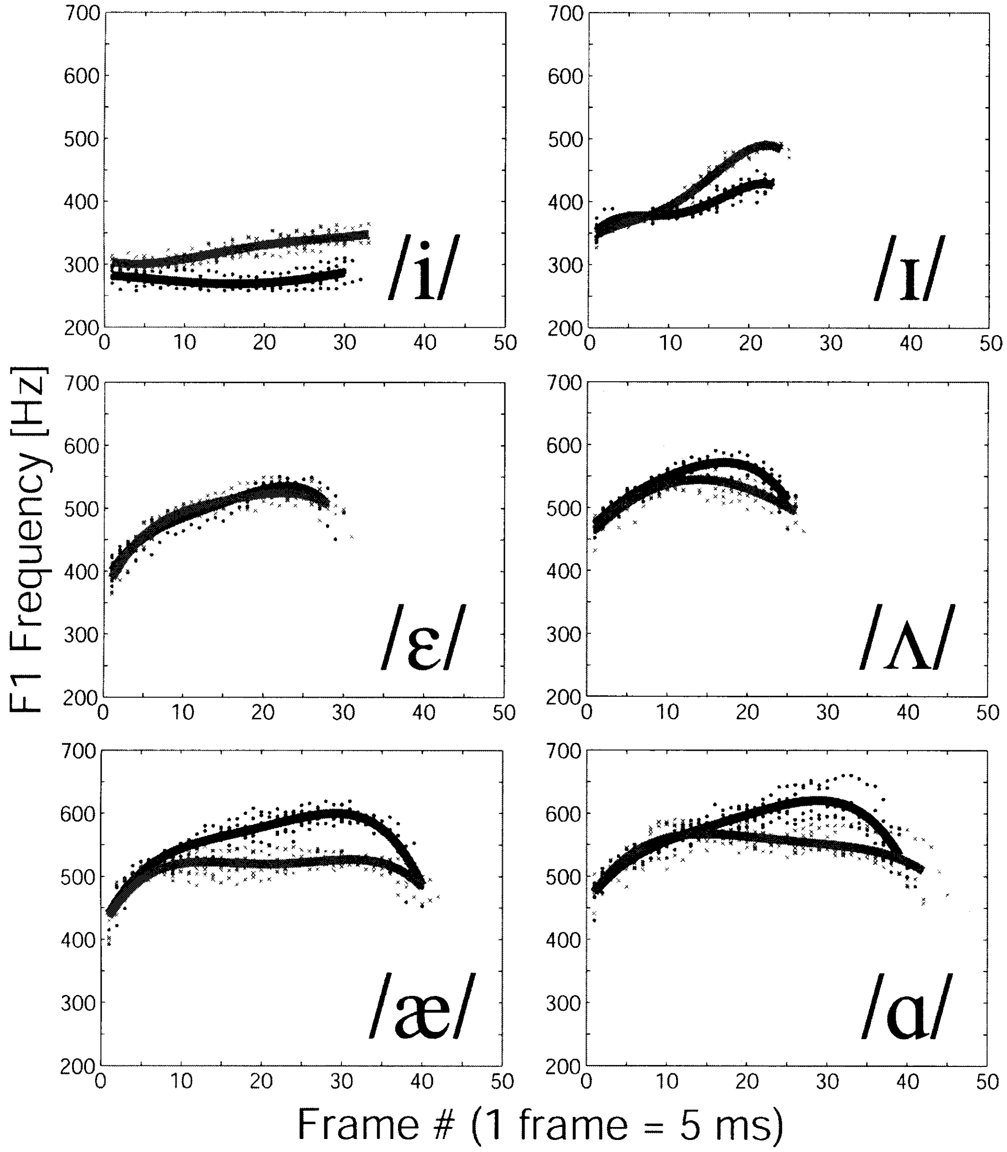

Fig. 5. Measured F1 frequency vs. time for vowels uttered in oral (/bVb/) and nasal (/bVm/) contexts. The dots '.' and 'x' represent the F1 frequency at a particular time in /bVb/ and /bVm/, respectively. The blue (dark) and red (light) lines are the average curves in /bVb/ and /bVm/ among five repetitions obtained by smoothing with a 3rd-order polynomial approximation.

Table 1.F1 frequencies of V in oral (/bVb/) and nasal (/bVm/) contexts at the point where the difference between them reaches maximum.

V	F1 of (b)V(b)	F1 of (b)V(m)	Difference
	[Hz]	[Hz]	[Hz]
i	273.6	335.2	61.6
ɪ	420.7	483.4	62.8
ɛ	535.0	524.5	-10.5
ʌ	565.6	530.1	-35.5
æ	599.3	524.7	-74.7
ɑ	619.8	550.9	-68.8

Results and discussion

Figure 5 shows the results of the formant tracking (formant frequencies versus time) for F1 of the six vowels. Each plot contains the results of the formant tracking of each /bVb/-/bVm/ pair (we refer to the first one of the pair as "oral context", and the second as "nasal context"). This pairing, in fact, allows us to see only the effect of nasalization, because it can be assumed that the only difference between the pair is whether or not there is velopharyngeal opening, and the rest of the articulatory movements are identical. In each plot, the dots '.' and 'x' represent the first formant frequency at a particular time in /bVb/ and /bVm/, respectively. The blue (dark) and red (light) lines are the average curves in /bVb/ and /bVm/ among five repetitions obtained by smoothing with a 3rd-order polynomial approximation.

From these plots in Fig. 5, we observed the bidirectional shifts in terms of F1 due to nasalization, that is, the F1 frequency shifts downward for the low vowels /ɑ/ and /æ/ and upward for the high vowels /i/ and /ɪ/.

To model the bidirectional shift in F1 frequency due to nasalization, the maximum difference between the two average F1 curves in oral and nasal contexts was examined (in Table 1). This maximum difference as function of F1 frequency (in oral context) can be modeled by a 2nd-order polynomial approximation:

$$\Delta F_1 = 67.8 \left[\frac{2}{1+e^{0.0435(F_1-529.5)}} - 1 \right] \text{[Hz]}.$$

From this model, we can thus see that the F1 frequency tends to shift bidirectionally toward the central region that is around 530 Hz in this case.

American English. The vowel V was either /i/, /ɪ/, /ɛ/, /ʌ/, /æ/, or /ɑ/; and the consonant C was either /b/ or /m/. The target words were embedded in the carrier phrase "Say ________, again." All 12 combinations were repeated five times in a random order (60 utterances, in total). The utterances were recorded in a partially sound-attenuated room with an Electro-Voice model 054 omnidirectional dynamic microphone and a pre-amplifier (Shure Professional Microphone Mixer). The microphone was placed approximately 20 cm in front of the speaker's lips to get equal distances from the microphone to the mouth and to the nose. All signals were digitized with a sampling frequency of 16 kHz. Formant tracking of the vowels was done by using LSPECTO, which was based on the 20th order LPC (25-ms Hamming window and 5-ms frame shift).

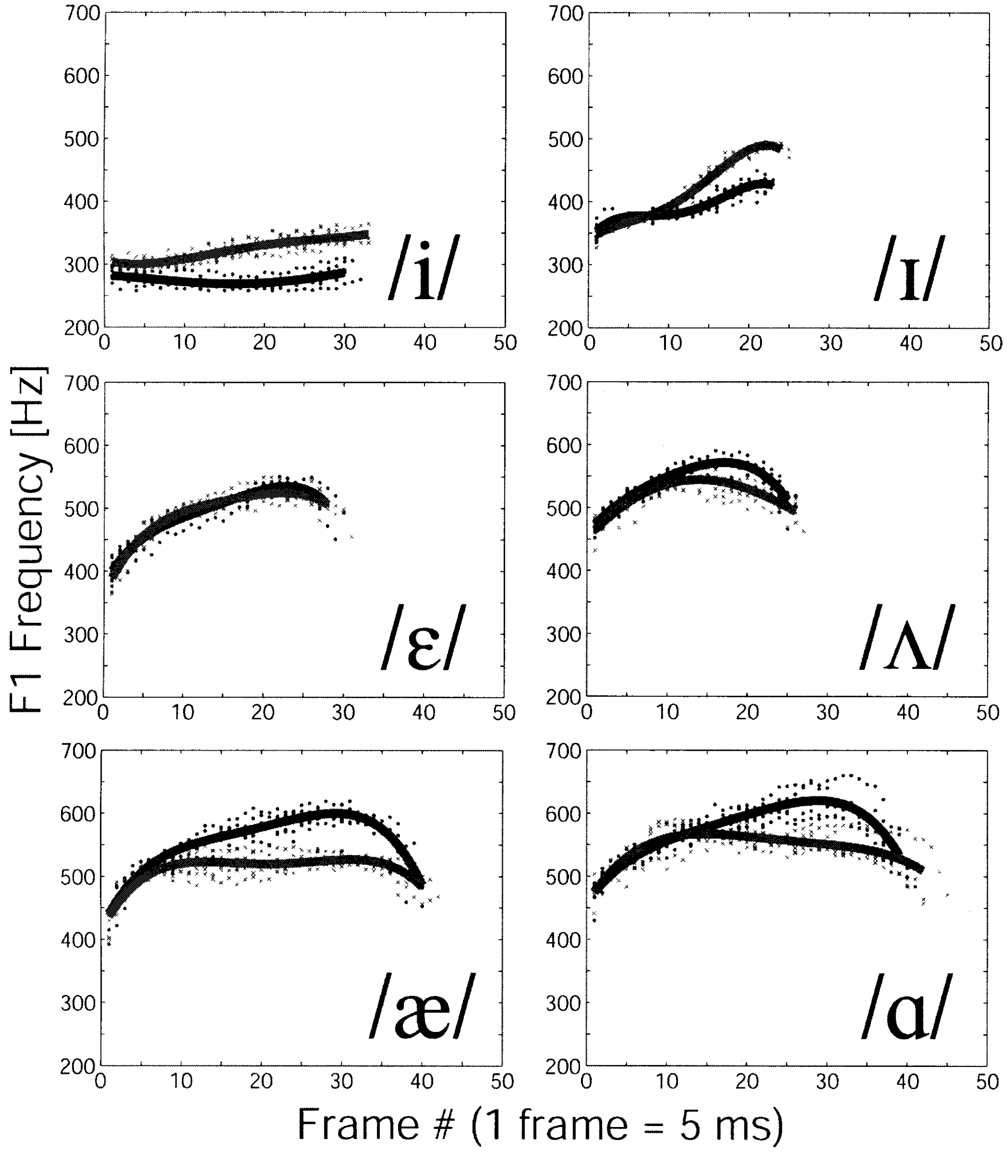

Fig. 5. Measured F1 frequency vs. time for vowels uttered in oral (/bVb/) and nasal (/bVm/) contexts. The dots '.' and 'x' represent the F1 frequency at a particular time in /bVb/ and /bVm/, respectively. The blue (dark) and red (light) lines are the average curves in /bVb/ and /bVm/ among five repetitions obtained by smoothing with a 3rd-order polynomial approximation.

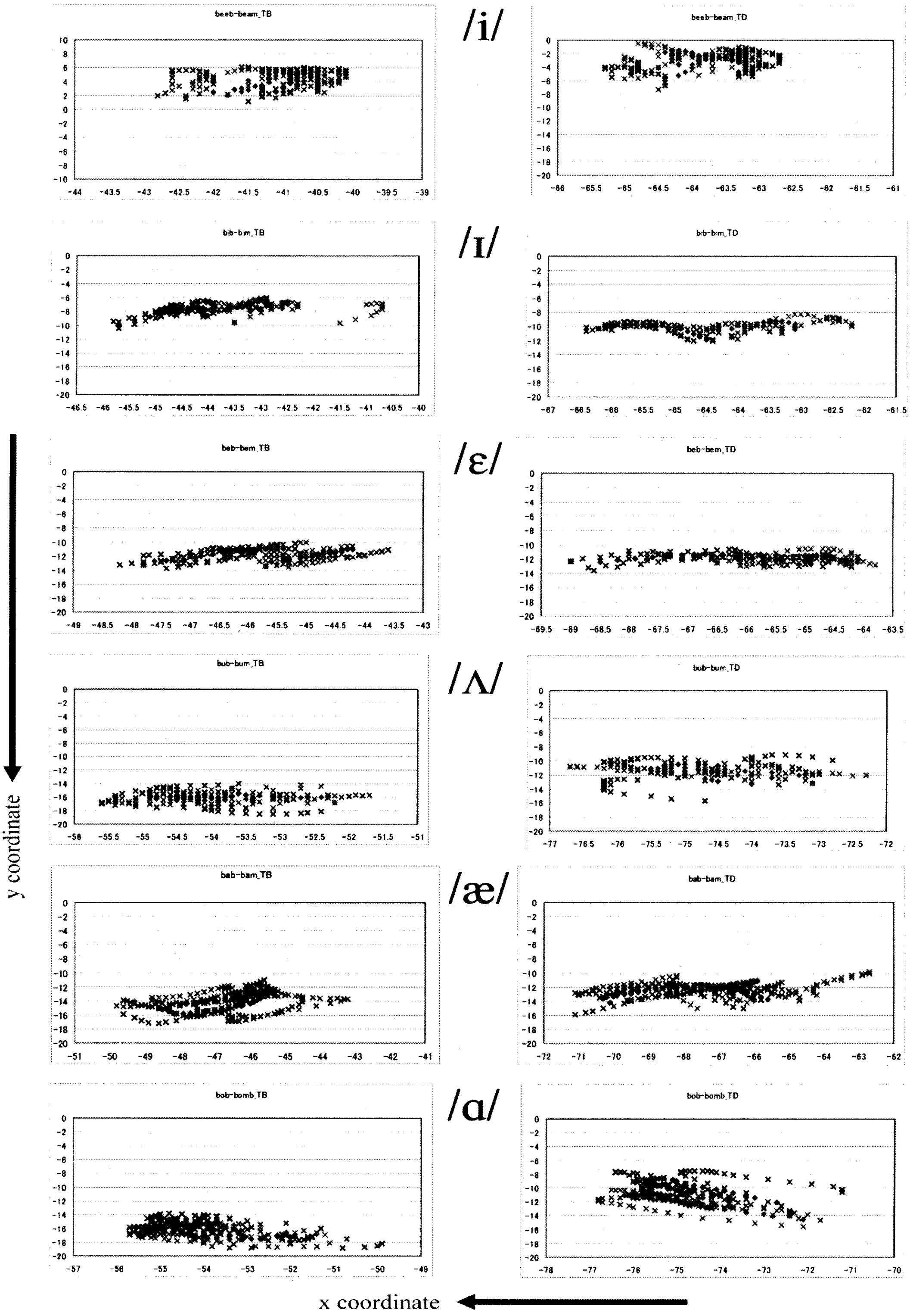

Fig. 6. The x and y coordinates of the TB and TD transducers between each minimal pair during the vowels in oral and nasal contexts (left column: TB; right column: TD). The blue (dark) and red (light) dots represent the tongue positions in oral and nasal contexts, respectively.

In this figure, the x-coordinate runs from back to front and the y-coordinate runs from bottom to top on the midsagittal plane, and the origin is located at the upper incisor.

For /i/, /æ/ and /ʌ/, the two distributions in oral and nasal contexts are fairly overlapped in both TB and TD cases. For /ɪ/ and /ɛ/, the distributions of the y-coordinate of the TB and TD transducers are overlapped, whereas the x-coordinates are somewhat separated (i.e., tongue positions are more advanced in nasal context). Because we are interested in tongue height, however, we can conclude that there is no major difference in y-coordinate between the two contexts. For /ɑ/, the distributions of the x-coordinate of the TB and TD transducers are fairly overlapped. There is, however, a difference between the two distributions in the y-coordinate. The difference of the x-coordinates of the TD transducer is approximately 2-3 mm.

As a result, the distributions between oral and nasal contexts were overlapped each other by a fair amount, and the difference was not significant except the lowest vowel /ɑ/. In case of the vowel /ɑ/, the result shows that the tongue dorsum is lower in nasal context (i.e., /bɑm/) than in oral context (i.e., /bɑb/). In other words, this speaker might have tried to make a more extreme /ɑ/ to compensate for the F1 shift due to nasalization.

General discussions

For naturally produced vowels in oral and nasal contexts in Section 3.3, we compared the minimal pairs /bVb/-/bVm/. It is a common observation that vowels adjacent to nasal consonants are nasalized. This is due to overlapping gestures of the velum and the tongue or lips. Such overlapping gestures occur in the both cases: a vowel followed by a nasal consonant, and a vowel preceded by a nasal consonant. However, the overlap is relatively longer in the former case than the latter case (Stevens, 1998).

If the vowel V is preceded by an obstruent consonant C in CVN context where N is a nasal consonant, the velopharyngeal port is closed at the time of release of the consonant C; the port then opens during the vowel V in preparation for the formation of the oral closure for the nasal consonant N (Stevens, 1998). This avoids build up of the intra-oral pressure due to the oral closure for the nasal consonant. Because it takes time to move velum down, the velopharyngeal port starts opening before the complete closure of the oral cavity.

Ohala (1971) has reported greater nasal coarticulation effects in vowels preceding nasals than in vowels following nasals, and states that velar lowering begins as soon as elevation is no longer required for obstruent articulation. Moll and Daniloff (1971) also reported that movement toward opening of the velar port began during articulator movement toward the first vowel in a CVN sequence.

For any vowel, F1' increases when the vowel is nasalized. At the same time, Fn was observed in frequencies lower than F1 for the low vowel [ɑ], and the Fn became dominant as the velophayngeal port area increases (especially, in Section 3.1). Thus, this observation from the acoustic measurements matches well to the prediction by the acoustic theory, and it supports the explanation that the

F1 of a low vowel shifts downward due to the dominant Fn. Because the same phenomenon is observed in Section 3.2, this “F1 transfer” from F1’ to Fn is considered to occur in real speech. It is a little bit harder to see such an F1 transfer for /ɑ/ from the results in Section 3.3 because of the initial transition from the preceding obstruent (i.e., /b/). However, the difference between the average F1 contours obtained from the vowels in nasal and oral contexts allows us to see an initial upwards movement, and this might be the evidence of this F1-transfer phenomenon. For high vowels, because there is no such F1 transfer, F1’ monotonically shifts upward in frequency. As a consequence, one can observe the bidirectional frequency shifts of F1.

In nasalization of vowels, we could see the “quantal nature” of speech (Stevens, 1972, 1989). Let us consider that the velum lowers from the raised position (the complete closure of the velopharyngeal port). Although the lowering speed is constant, the perceived nasality does not increase constantly. Especially for a low vowel, the nasality does not increase unless the velopharyngeal opening reaches a certain area, as described earlier. In fact, the velopharyngeal port does not close completely when we utter a low vowel with no perceived nasality. Interestingly, we might rather be able to hear less nasality with a slight opening of the velopharyngeal port (Maeda, 1993). Thus, for a low vowel, a plateau exists where a small perturbation in the velopharyngeal opening does not influence the perceived nasality. This type of nonlinear relations between articulation and acoustics or perception is a part of the nature of quantal aspects of speech (Stevens, 2003).

The quantal aspects of speech can also be observed in the formant frequency shifts of nasalized vowels. Especially for a low vowel, the complicated F1 movement due to the “transfer” phenomenon can be observed during the transition from complete closure to some degree of opening at the velopharyngeal port. Before the F1 transfer occurs, the perceived nasality is low and the F1 frequency is more stable. After the F1 transfer, the perceived nasality is higher and the F1 frequency is somewhat stable. This nonlinear relation between articulation and acoustics is again an aspect of the quantal nature of speech, and one might predict that a nasal vowel tends to be produced in such a stable region, especially in languages which have a phonemic distinction between nasal and oral vowels, such as French (Maeda, 1993).

Conclusions

In this study, the formant frequency shifts of a nasalized vowel from the aspects of both acoustics and perception were investigated. From the acoustic measurements, bidirectional formant shifts in terms of F1 frequency were observed as predicted by the acoustic theory (ranging from about –75 Hz for low vowels to about +65 Hz for high vowels in Section 3.3).

As a result of the articulatory measurements using the EMMA system in Section 4.1, the distributions of the maximum tongue height between each minimal pair during the vowels in oral and nasal contexts overlapped each other a fair amount, and the difference was not significant except the lowest vowel /ɑ/. This might be an indication that a speaker is not compensating his/her tongue height except the extreme one.

If a speaker knows that F1 frequency shifts when nasalized, there might be a chance that the speaker tries to compensate for the shift during the production of speech. Furthermore, we might be able to observe the compensatory correlation, or "motor equivalence," between velar and tongue heights. However, it could be only true in languages that distinguish nasal and oral vowels phonemically (such as French, Portuguese, and Hindi). American English, on the other hand, does not make such distinction, and there might be no compensation on the production side.

Acknowledgments

This study was done while I was a Visiting Scientist of the Speech Communication Group in the Research Laboratory of Electronics, Massachusetts Institute of Technology (Cambridge, MA, USA) during the time period from 2000 through 2004. I would like to thank all of the people who helped me in various ways, especially Kenneth N. Stevens, Joseph S. Perkell, Stefanie Shattuck-Hufnagel, Sharon Manuel, Janet Slifka, Helen Hanson, Majid Zandipour, Mark Tiede, other members of the Speech Communication Group at MIT, Bernard Gold of MIT Lincoln Laboratory, John J. Ohala of University of California, Berkeley, Tsutomu Sugawara and Setsuko Imatomi of Sophia University, Haruko Miyakoda of Tokyo University of Agriculture and Technology, Shin'ichi Tokuma of Chuo University, and Kyoko Takeuchi of University of Tokyo. This research was supported in part by Grants-in-Aid for Scientific Research (A-2, 16203041) from the Japan Society for the Promotion of Science.

References

Arai, T. 2004. Formant shift in nasalization of vowels. J. Acoust. Soc. Am., 115, 2541.

Bell-Berti, F. 1980. Velopharyngeal function: A spatial-temporal model. *Speech and Language: Advances in Basic Research and Practice* 4, In N. J. Lass (ed.), New York: Academic, 291-316.

Chen, M. Y. 1995. Acoustic parameters of nasalized vowels in hearing-impaired and normal-hearing speakers. *J. Acoust. Soc. Am.* 98, 2443-2453.

Chen, M. Y. 1997. Acoustic correlates of English and French nasalized vowels. J. Acoust. Soc. Am. 102, 2360-2370.

Fant, G. 1960. *Acoustic Theory of Speech Production,* The Hague: Mouton.

Fujimura, O. 1960. Spectra of nasalized vowels. *Res. Lab. Electron. Q. Prog. Rep.* No. 58, MIT, 214-218 (July 15, 1960).

Fujimura, O. 1961. Analysis of nasalized vowels. *Res. Lab. Electron. Q. Prog. Rep.* No. 62, MIT, 191-192 (July 15, 1961).

Fujimura, O. and Lindqvist, J. 1971. Sweep-tone measurements of vocal-tract characteristics. *J.*

Acoust. Soc. Am. 49, 541-558.

Hattori, S., Yamamoto, K., and Fujimura, O. 1958. Nasalization of vowels in relation to nasals. *J. Acoust. Soc. Am.* 30, 267-274.

Hawkins, S. W. and Stevens, K. N. 1985. Acoustic and perceptual correlates of the non-nasal--nasal distinction for vowels. *J. Acoust. Soc. Am.* 77, 1560-1575.

House, A.S. and Stevens, K. N. 1956. Analog studies of the nasalization of vowels. *J. Speech Hear. Disord.* 21, 218-232.

Klatt, D. H. 1984. The new MIT speech VAX computer facility. *Speech Communication Group Working Papers IV,* Research Laboratory of Electronics, MIT, Cambridge, 73-82.

Klatt, D. H. and Klatt, L. C. 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J. Acoust. Soc. Am.* 87, 820-857.

Krakow, R. A., Beddor, P. S., Goldstein, L. M., and Fowler, C. A. 1988. Coarticulatory influences on the perceived height of nasal vowels. *J. Acoust. Soc. Am.* 83, 1146-1158.

Lindqvist, J. and Sundberg, J., 1972. Acoustic properties of the nasal tract. *Speech Transmission Laboratory Quarterly Progress and Status Report* 1, 13-17.

Maeda, S. 1993. Acoustics of vowel nasalization and articulatory shifts in french nasal vowels. In Huffman, M. K. and Krakow, R.A. (eds.), *Phonetics and Phonology: Nasals, Nasalization, and the Velum* 5, SanDiego: Academic Press, 147-167.

Moll, K. L. and Daniloff, R. G. 1971. Investigation of the timing of velar movements during speech. *J. Acoust. Soc. Am.,* 50, 678-684.

Ohala, J. J. 1971. Monitoring soft palate movements in speech. *Project on Linguistic Analysis Reports* (Phonology Laboratory, Department of Linguistics, University of California, Berkeley), 13, J01-J015.

Ohala, J. J. 1986. Phonological evidence for top-down processing in speech perception. In Perkell, J. S. and Klatt, D. H. (eds.), *Invariance and Variability of Speech Processes,* Hillsdale: Erlbaum, 386-401.

Perkell, J., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I., and Jackson, M. 1992. Electromagnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements. *J. Acoust. Soc. Am.* 92, 3078-3096.

Stevens, K. N. 1972. The quantal nature of speech: Evidence from articulatory-acoustic data. In Denes, P. B. and David, E.E. Jr. (eds.), *Human communication: A unified view,* New York: McGraw Hill, 51-66.

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

Stevens, K. N. 1998. *Acoustic Phonetics,* Cambridge: MIT Press.

Stevens, K. N. 2003. Acoustic and perceptual evidence for universal phonological features. *Proc. of International Congress of Phonetic Sciences,* 33-38.

Stevens, K. N., Fant, G., and Hawkins, S. 1986. Some acoustical and perceptual correlates of nasal vowels. In Channon, R. and Shokey, L. (eds.), *Honor of Ilse Lehiste,* Dordrecht: Foris, 241-254.

Stevens, K. N., Nickerson, R. S., Boothroyd, A., and Rollins, A. M. 1976. Assessment of nasalization in the speech of deaf children. *J. Speech Hear. Res.* 19, 393-416.

Wright, J. T. 1975. Effects of vowel nasalization on the perception of vowel height. In Ferguson, C.A., Hyman, M. and Ohala, J.J. (eds.), Nasalfest: Papers from a Symposium on Nasals and Nasalization, (Language Universals Project, Stanford University, Stanford, CA), 373-388.

Wright, J. T. 1986. The behavior of nasalized vowels in the perceptual vowel space. In Ohala, J. J. and Jaeger, J. J. (eds.), *Experimental Phonology,* Orlando: Academic, 45-67.