Acoustic Correlates of Nasality in Speech

by

Marilyn Y. Chen

B.S. E.E., University of Illinois
(1989)

M.S. E.E., Massachusetts Institute of Technology
(1991)

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

Doctor of Philosophy in Medical Engineering

at the
Massachusetts Institute of Technology
February 1996

Author ....
Harvard-MIT Division of Health Sciences and Technology
January 12, 1996

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

Accepted by ....
Roger G. Mark
Chairman, Committee for Graduate Students

Acoustic Correlates of Nasality in Speech

by

Marilyn Y. Chen

Submitted to the Harvard-MIT Division of Health Sciences and
Technology on January 12, 1996 in partial fulfillment of the
requirements for the degree of Doctor of Philosophy

ABSTRACT

One of the prevalent abnormalities in speech is inadvertent vowel nasalization. Acoustic analysis of nasalized vowels in the frequency domain indicates the presence of extra spectral prominences between the first two formants and below the first formant which introduce nasal peaks with amplitudes P1 and P0 (dB), respectively. Furthermore, the amplitude of the first formant, A1 (dB), is reduced. These acoustic characteristics can be explained by speech production theory. The relative amplitudes A1-P1 and A1-P0 were measured for vowels adjacent to nasal consonants and those adjacent to stops for English speakers. French utterances with nasal-non-nasal vowel distinction were also examined. The differences between the non-nasal vowels and nasal vowels were statistically significant for A1-P1, especially in the non-low vowels and for A1-P0, especially in the non-high vowels. These acoustic correlates were systematically manipulated in synthetic words, which were presented to listeners in perceptual experiments. The results indicate that A1-P1 and A1-P0 have high correlation with perceived nasality. In addition, nasal consonants and vowels spoken by patients who have gone through endonasal sinus surgery were analyzed to determine how changes in the nasal cavity anatomy can affect the spectra in the vicinity of the first formant, particularly changes in A1-P1 and A1-P0 for nasal vowels. The surgery was found to have opposite effects both acoustically and perceptually on the low vowels and high vowels. Influences of formant location and breathiness on A1-P1 and A1-P0 were addressed, as well as the variability of these acoustic correlates over time.

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

Acknowledgements

It has been a great joy and privilege to have Ken Stevens as my advisor. His insight and love for research in speech communication have been a great inspiration to me. His encouragement has meant much to me, especially during those dark hours when I thought that I could not make it at MIT. I am also grateful for his diligence and patience in reading and correcting my writings.

I am thankful for the other members of my thesis committee for their time and support. Bob Hillman has helped me to see my research from a clinician's point of view and started me on the project examining the speech of patients who have gone through endonasal sinus surgery. Bill Peake has given me detailed and constructive comments on my thesis that taught me much about technical writing and data analysis. Joe Perkell has asked me stimulating questions that helped me to improve my research methods.

I want to thank HST and Dr. Ralph Metson and his patients for making possible clinical application of my research. I am also grateful to the French speakers who participated in my experiments. This research has truly been a "lab-grown" project with the participation of almost every member of the speech lab as speakers and/or listeners for my experiments. Thanks to Abeer Alwan, Corine Bickley, Karen Chenausky, Laura Dilley, Peter Guiod, Helen Hanson, David Horowitz, Mark Johnson, Joyce Manzella, Noel Massey, Kelly Poort, Sharon Manuel, Alice Turk, Lorin Wilde, and Jane Wozniak for their patience and kindness. I also want to thank Melanie Matthies for her crash course on statistics, Sharon Manuel and Yi Xu for their expertise on the Macs, and Seth Hall, Keith North, and Arlene Wint for their technical support. Furthermore, I have been blessed by the encouragement from Stefanie Shattuck-Hufnagel, Jeff Kuo, Sharlene Liu, and Walter Sun.

It has been a great blessing for me to live in the same dorm as Phoebe Cheng, Elaine Chew, Yanqing Du, Judy Wang, Alice Liu, Janet Pan, and Chiong Zhang. They have truly made Green Hall a home away from home for me.

I am grateful for my friends at Boston Chinese Evangelical Church and especially those from the MIT Chinese Christian Fellowship. Thanks to Tim Chow, Dorothy Hsieh, and Chi Wong for their faithful concerns and prayers. Thanks to Jerry Chen, Melissa Lee, Phil Lin, and Tom Wang for their empathy in my academic and personal trials. Thanks to Irene Lowe, Nancy Chin, Mei Ching Ng, Lorraine Ho, and Davina Ling for their advice, understanding and prayers.

I am blessed with the persevering and sacrificial love of my family. Thanks to Mom and Dad for immigrating our family to this country to better my education, for the countless care packages, and for the expensive visits. Thanks to Shary, Bob, Stephanie, and Jennifer for opening their home to me, for the free meals and haircuts, and for adding to our family bundles of joy and laughter. Thanks to Lynn and Mark for their zeal to live their new life, for their deep sharing and encouragement, and for their faithful prayers.

Ultimately, I thank God-- the Alpha and the Omega for being the originator and sustainer of speech. Without Him holding all things together, none of my experiments could be performed or repeated. I am also forever grateful to God-- Immanuel, for becoming man in order to communicate with us through human speech. He has used a way so tangible that even the analysis and synthesis methods presented in this thesis can be applied to it.

This research was financially supported in part by an NSF fellowship, an NIH grant, and the Lebel fellowship.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

The degree of acoustic coupling between the vocal tract and the nasal tract is controlled by the velum (soft palate) as well as the posterior and lateral pharyngeal walls. This coupling to the nasal cavity which is defined as nasalization, greatly contributes to the overall effect of perceived nasality. In English, the velopharyngeal (v-p) port is opened during the production of the nasal consonants /m/, /n/, and /ŋ/, but it is raised for the obstruent consonants to allow pressure build-up. During normal vowel production, the v-p port is generally closed to prevent airflow into the nasal cavity unless it is adjacent to a nasal consonant. If the v-p port is opened excessively during vowel production, it may be considered a speech disorder, labeled as “inadvertent nasalization". The effect of nasalization on intelligibility would be greater on consonant production and it is reflected in the adjacent nasal vowels reflecting coarticulation. The timing of opening the v-p port can also affect intelligibility. In languages, such as French, Gujarati, Hindi, and Bengali, in which some vowels are intentionally nasalized, inadvertent vowel nasalization would influence the intelligibility of the language.

Some populations tend to nasalize differently from normal speakers. For instance, the development of speech by persons who are born deaf or are deafened prior to the age of five years is greatly retarded by impaired ability to hear their own vocalizations and the speech of others. According to many speech teachers, inadvertent nasalization is one of the most common problems of deaf speakers (Brehm, 1922). It is also difficult for people with velopharyngeal insufficiency due to a cleft palate, paralyzed velum, or adenoidectomy to contrast nasal and non-nasal sounds. Furthermore, patients who have had endonasal sinus surgery for treating severe sinusitis have often noticed a change in their speech post-operatively. It is likely that the acoustic signal during the production of nasal sounds is being affected by anatomical changes from the surgery.

It is difficult to obtain consistent subjective judgments of nasality, especially when other speech disorders are present. Several abnormalities such as speech tempo, misarticulations, and pitch variations, in addition to inadvertent lowering of the velum, influence the perception of nasality (Spriestersbach, 1955). However, nasalization is a well defined articulatory gesture that greatly contributes to the overall effect of nasality. Since the size of the opening to the nasal cavity is not readily visible and most people are not conscious of velum movement (Shelton *et al.*, 1970), some objective measure is needed to estimate the velopharyngeal opening.

1.2 Literature Survey

1.2.1 Articulatory Studies

Various methods of imaging used to determine the vocal tract shape and dimensions can give some quantitative information on the velopharyngeal opening. Earlier studies consist of lateral radiographic images (Abramson and Cooper, 1963; Perkell, 1969) which only yield two-dimensional data, requiring the third dimension to be determined by applying transformations to the sagittal measurements. More data are needed to determine the transformations adequately. X-ray computer tomography (CT) has also been used (Kiritani *et al.*, 1977; Johansson *et al.*, 1983). In principle, tomography can supply the three-dimensional vocal tract information. However, one needs to be concerned with the danger of using even low x-ray dosages, especially for non-medical uses. Also, the CT system is limited in its maneuverability so that images of the upper vocal tract are difficult to obtain. X-ray microbeam (Fujimura *et al.*, 1973) and magnetometers (Schonle *et al.*, 1987) can supply dynamic information but only in the midsagittal plane. More recently, magnetic resonance imaging (MRI) has been used to examine the vocal tract (Baer *et al.*, 1991; Moor, 1992) and the nasal cavities (Dang *et al.*, 1994). It can provide three-dimensional images without using ionizing radiation or repositioning the subject as in CT scans. However, because the time required for the imaging process (several tens of seconds to several minutes), the speech sound needs to be sustained.

Past studies using the imaging techniques have yielded information on vowel nasalization. Moll (1962) quantified velopharyngeal closure in vowels within various consonant contexts for normal English speakers by using cinefluorography. He concluded that the high vowels, /i/ and /u/, exhibit significantly greater velar height and velopharyngeal contact than low vowels, /æ/ and /ɑ/. Also, he found that v-p opening during vowel production is dependent on the adjacent consonants. Vowels with nasal consonant context, especially /n/, exhibit v-p opening much more often than those with other consonant contexts and no-consonant context. Furthermore, vowels that precede a nasal consonant have a greater velum-pharynx distance than vowels following a nasal consonant. In fact, according to a similar study by Moll and 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. Stevens *et al.* (1976) also concluded that a vowel between two nasal consonants is usually produced with a partial v-p opening. According to McClumpha (1966), who examined cinefluorographic films of CV syllable production, there is significant variation in the v-p ports between deaf and normal subjects.

The anatomy of the nasal cavity was also clarified by imaging. From anatomical atlases, sample skulls, and lateral x-ray pictures, House and Stevens (1956) modeled the nasal tract as cylindrical sections and implemented the model with electrical analogs. They concluded that although part of the nasal tract consists of two channels, assuming symmetry allows one to model the tract using a single tube within the frequency of interest for speech. From MRI studies by Dang *et al.* (1994), the dimensions of nasal sinuses were measured and found to act as separate resonators that are coupled to the nasal passages.

From past studies based on imaging, coarticulation with nasal consonant or deafness often result in particular distinguishing characteristics of the acoustic signal within the vowel due to nasalization. Anatomically, by assuming symmetry one can model the nasal tract as a single tube coupled with nasal sinuses as resonators.

1.2.2 Acoustic Studies

Other studies have used acoustic and aerodynamic signals to obtain measures of the velopharyngeal opening objectively. These measures include nasal pressure (Weiss, 1954; Shelton *et al.*, 1967), nasal vibration (Stevens *et al.*, 1975; Stevens *et al.*, 1976), and nasal flow (Quigley *et al.*, 1964). Although Weiss found a high correlation coefficient, 0.94, between nasality perception judgments and the pressure parameter, Shelton et al. and others were not able to confirm his results. Stevens et al. reported a high correlation coefficient as well, 0.78, between the amplitude of nasal vibration and nasality judgments. By using a nasal accelerometer to provide a measure of velopharyngeal opening during voiced sounds of severely hearing-impaired and normal-hearing children, they concluded that there is improper nasalization of vowels for a large number of the hearing-impaired children due to inadequate control of the velopharyngeal articulators. Nasal pressure, nasal vibration, and nasal flow, however, are all dependent on the intensity of the speech. Horii (1980) suggested the Horii Oral-Nasal Coupling index (HONC), which is a ratio of nasal accelerometric amplitude to voice amplitude. However, according to Redenbaugh and Reich (1985), no correlation was found between the HONC accelerometric values and nasality judgments for sentences containing nasal semivowels and vowels. In 1986, Kay Elemetrics introduced the Nasometer, which is a microcomputer-based device that determines the amount of nasal acoustic energy relative to nasal-plus-oral energy during speech production. In a study by Dalston *et al.* (1991), the measurements made by the Nasometer were compared to perceived hyponasality. They found a correlation of -0.65, which indicates that the Nasometer may not be sensitive to the acoustic characteristics that contribute to nasality. The problem arised especially with audible nasal emission during the production of consonants that normally are produced with the velopharyngeal port closed, which may give a high score on the nasometer but a low nasality judgment. For all of these measures, a sensing device at the nose is needed in addition to the microphone. Also, the measurements do not reflect any details about the nasal-cavity anatomy, which may significantly influence the perception of nasality in addition to nasalization.

Analysis of the acoustic signal in the frequency domain, on the other hand, is more ideal in a clinical setting since it requires only the output from the microphone. Furthermore, such analysis is expected to reflect the condition of the nasal cavity such as a larger total surface area and coupling to the paranasal sinuses. The frequency domain analysis involves examining the deviation of the nasalized vowel spectrum from that of the non-nasalized vowel. Delattre (1954) believed that the reduction in amplitude of the first formant spectral peak is the primary cue of nasalization. Fant (1960) came to similar conclusion, and indicated that it is the most consistent characteristic of a nasal. An amplitude decrease of the first formant and a spectral peak developing around 1 kHz with an anti-resonance in the range of 700-1800 Hz were observed for a large velopharyngeal opening in the electrical analog circuit modeling by House and Stevens (1956).

Theoretically, a pole-zero pair in the vicinity just above F1 is expected to accompany vowel nasalization (Fant 1960; Fujimura, 1960; Fujimura and Lindqvist, 1971). According to Huffman's study of contextual nasalization (1990), the nasal pole occurred 40-50 ms after vowel onset in a word like "bean," in the region of 600-1000 Hz, and the zero was 200-300 Hz above the pole. From other studies, nasal coupling was found to introduce in the region of F1 a pole and a zero whose locations depended on the vowel quality (Hawkins and Stevens, 1985; Stevens *et al.*, 1987). Maeda (1982b) did a systematic study of spectral modification by nasal coupling through synthesis based on an acoustic simulation of a realistic vocal-tract model. An increase in coupling corresponded to an upward shift of nasal formant-antiformant frequencies and a larger magnitude of the extra peak. With larger nasal coupling, energy was more evenly distributed over the F1-F2 region due to the nasal formant, so that the spectrum appeared to be more flattened. However, in the study, a measure of flatness closely corresponding to the degree of nasal coupling was not found. The extra pole-zero pair affected F1 and F2 differently, depending on the formant locations for the various vowel qualities.

From Maeda's computer simulation results (1982a), he concluded that another pole-zero pair at low frequencies may be introduced by paranasal sinuses. From spectrogram examination done by Hattori, Yamamoto, and Fujimura (1958), nasalized vowels have certain distinctive acoustic features: an extra pole around 250 Hz, an extra zero around 500 Hz, and components added to the valleys between formants. In Fujimura and Lindqvist's sweep-tone measurement of the vocal tract during the production of nasalized vowels (1971), a pole-zero pair below the first formant peak was observed. Similarly, Lindqvist and Sundberg in their measurement of the nasal transfer function (1972) showed evidence of the extra peak in the low frequencies, which they attributed to the frontal and maxillary sinuses. Båvegård *et al.* (1993) suggested that a low frequency pole-zero pair around 300 Hz is associated with the maxillary sinus since it has the largest volume. Dang and Honda (1995) used sound pressure gradient in the nasal tract and sound pressure at the nostril to determine the volume velocity transfer function, and showed that the nasal zero can be as low as 310 Hz.

They suggested that this zero may be due to the sphenoid sinus as well as the maxillary sinus.

In the higher frequencies, nasalization may introduce shifts in formants, modification of formant amplitudes, and additional spectral peaks. However, these effects are not as consistent as those in the vicinity of the first formant (Hawkins and Stevens, 1985).

1.2.3 Perceptual Studies

As early as 1942, Hudgins and Numbers carried out an experiment to show that excessive nasality could be perceived by a listener. According to Wright (1980), nasal vowels could be distinguished from oral vowels even by speakers of languages that do not make a distinction between the two. Furthermore, Hawkins and Stevens (1985) discovered that American listeners, who do not distinguish nasal and oral vowels in their language, were better than Gujarati, Hindi, and Bengali listeners (who do distinguish nasal and oral vowels) at discriminating nasality even when the nasalization was minor.

Past studies have shown that vowel context and vowel quality influence nasality judgments. According to McIntosh (1937), vowels were judged to be more nasal in some consonant contexts than others: those preceding a nasal consonant had a greater tendency to be judged nasal than those following a nasal consonant. This finding may be due to the different effects that vowel context has on nasalization. Krakow and Beddor (1991) found that American English speakers perceived nasality better with nasal vowels either in isolation or in oral context than with nasal vowels in nasal context. This result indicates that one may first attribute nasality of the vowel to a nasal context but if there is no nasal context, nasality is attributed to the nasalization of the vowel. Furthermore, vowel quality influences nasality. From a perceptual test in which relative judgments of nasality were obtained, House and Stevens (1956) observed that for the same degree of nasality judgment, high vowels required less nasal coupling than low ones. Also, the responses suggested that there is an area of nasal opening for which additional enlargement would not contribute further to nasality perception. In a study of the French vowels, /i/, /ɑ/, and /u/ by computer simulation, Maeda (1982b) also found that perceptual effects of nasal coupling are vowel dependent. For /i/, the degree of perceived nasality increased rapidly relative to the amount of nasal coupling simulated. This trend was also apparent for the high vowel, /u/, although, for a given opening to the nasal cavity, the degree of nasality was judged to be weaker compared to /i/. For the low vowel, /ɑ/, much greater nasal coupling was necessary for the same effect on nasality perception.

Perceptual experiments with synthetic stimuli have been performed to examine individual spectral effect of nasalization, such as a wider first formant bandwidth or a nasal prominence. Huffman (1990) evaluated the contribution of the F1 amplitude to the perception of nasality by using analysis and synthesis. The amplitude difference between the first formant and the first harmonic was used to measure the change in the bandwidth of the first formant. It showed that F1 prominence contributed to the perception of nasality for /i/ and /I/ while for /æ/, a spectral change over time was needed for nasality perception. She concluded that the spectral change may be influenced by diphthongization, laryngeal adjustments, or changes in velopharyngeal opening. However, for all of the stimuli, manipulation of the nasality parameter only resulted in less than 50% of the times in which a given stimulus was presented that it was perceived to be nasal. House and Stevens (1956), in a perceptual study of stimuli from an analog synthesizer, found that the amplitude of F1 needed to be reduced by 8 dB for the nasality response to reach the 50% level. Stevens *et al.* (1988) did perceptual tests with Portuguese, French, and English listeners using stimuli generated with the Klatt synthesizer (Klatt, 1980). They discovered that about 6 dB lowering of the F1 amplitude was necessary to achieve a significant level of nasality perception.

Other studies have shown that the addition of a pole-zero pair to the spectrum is also effective in enhancing nasality perception when it is introduced in the vicinity of the first formant. In experiments on the perception of nasal vowels Hawkins and Stevens (1985) replaced the first formant of oral vowels by a pole-zero-pole to simulate an effect due to nasal coupling. The synthesis of the oral vowels was done by manipulating the Klatt synthesizer parameters to match that of the natural speech. The extra pole remained at the same frequency while the zero was manipulated to vary the extra peak prominence. Wider spacing of the pole-zero pair, usually above the first formant, was found to be necessary for the perception of a nasal vowel, especially for /e/, /ɑ/, and /u/. By computer simulations, Maeda (1982a) observed that modification of the valley between the first two formants contributes to the perception of nasality, especially for /i/. He also calculated auditory spectra in relative loudness density (sone/Bark) for /i/, /ɑ/, and /u/ with six different nasal coupling magnitudes (1993). He used the difference in frequency of spectral peaks (N2-N1) in the low-frequency region to measure nasality. N1 and N2 may include the first formant peak, the nasal formant peak, or the second formant peak, depending on which two of the three peaks are stronger. He found a good match between the perceived nasality and N2-N1 for /i/ and /ɑ/ but not for /u/. According to Malme (1959), the zero, on the other hand, contributed little to the nasality discrimination of the listener. Rather, it adjusted the adjacent formant amplitude to maintain the vowel type.

Other studies have indicated that another pole-zero pair at low frequencies also influences nasality perception. Hattori *et al.* (1958) carried out perceptual studies by feeding natural oral vowel signals into a special electric filter system. They found that it was necessary especially for non-high vowels to introduce an extra pole and zero at 250 Hz and 500 Hz, respectively, for a noticeable nasal effect. By using computer simulations, Maeda (1982a) connected to the nasal cavity a side cavity representing the maxillary sinuses. This side cavity introduced a pole-zero pair below the first formant. An informal listening of low vowels indicated that they require a greater v-p opening than high or mid vowels for the same degree of nasality perception. He suggested that the first formant region is primarily responsible for nasality perception. For low vowels which have F1 at higher frequency, greater coupling to the nasal cavity is required to shift the pole-zero pair to higher frequencies for spectral modification of the first formant region. The effect of this pole-zero pair on nasality perception was also examined by Klatt and Klatt (1990) through synthesis by raising the amplitude of the first harmonic by 6 dB or 10 dB through manipulating the open quotient of the glottal waveform. They found that the change enhanced nasality perception. The extra peak in the vicinity of the first few harmonics was also found to be significant in giving synthetic vowel a natural quality (Delattre, 1965).

The broadened first formant may affect the peripheral auditory system differently than the sharp prominence of an oral vowel (Stevens *et al.*, 1987). When the peak is narrower than a critical band, the outputs of the auditory filters in the vicinity of F1 show synchrony to F1; for a broadened first formant, synchronization is reduced (Delgutte,1980). On the other hand, if an extra peak is close to F1 so that they are less than a critical distance apart, they may be detected as a single prominence by the corresponding auditory filters. In either case, broadening the first formant causes fewer of the peripheral auditory neurons to respond synchronously to F1, and this response pattern could lead to perception of nasality.

1.3 Thesis Outline

The goal of this thesis is to quantify nasality due to nasalization objectively through measurements on the acoustic signal. It is motivated by the difficulty of obtaining consistent subjective judgments of nasality, especially when other speech disorders are present, the need of a better understanding of nasalization in order to synthesize more natural speech with nasal sounds, and the desire to develop a more accurate speech production model of the nasal cavity. Analysis and synthesis of the acoustic signal in the frequency domain are used to realize the goal.

In Chapter 2, a theoretical model of nasalized vowels, in which the main vocal tract is coupled to the nasal cavity, is introduced. The dimensions of the structural model are constrained by anatomical and physiological measurements. Spectral characteristics of vowel nasalization based on the suggested model are discussed.

In Chapter 3, the relevant acoustic correlates in nasalized vowels due to context or due to language (as in French) are quantified and compared to measures of the corresponding non-nasalized vowels for normal speakers. In particular, widening of the first formant and the introduction of extra prominences in the vicinity of the first formant are examined and quantified. The observations are consistent with speech-production theory for nasalized vowels.

Chapter 4 further tests the relevance of the acoustic parameters to nasality by using perceptual experiments with synthesized words in which the parameters were varied systematically. Specifically, the parameters varied are the first formant bandwidth, the nasal pole and zero frequencies, and the open quotient, OQ, which is the ratio of the glottis open time to the fundamental period. The average nasality judgments are correlated with the acoustic parameters.

Once a baseline has been established for the normal speakers, Chapter 5 presents results from the analysis of the speech of patients with endonasal surgery, noting the effect of anatomical alterations on the relevant acoustic correlates.

Chapter 6 further considers several possible problems when using acoustic correlates based on spectral measurements in the vicinity of the first formant. Factors other than nasalization that may influence the first formant and the extra peaks are addressed. These include the frequency of the formants and the acoustic attributes due to breathiness of the vowel. Variations of the measurements over time as well as over different speakers are also quantified. Examination of these factors will help to establish robust acoustic correlates of nasality.

Chapter 7 summarizes the results and gives suggestions for future research.

Chapter 2

An Acoustic Model of Nasalization

2.1 Velar Physiology and Nasal Structure

The velum, or soft palate, plays a major role in nasalization. In an oral vowel, it is raised so that its posterior end is in contact with the rear wall of the pharynx to prevent airflow into the nasal cavity. When it is lowered with the relaxation of the pharyngeal walls, however, the opening to the nasal cavity allows airflow through the nose as well as through the mouth, and the acoustic coupling causes the vowel to be nasalized. The velum is a flap of tissue about 4 cm long, 2 cm wide, and 0.5 cm thick (Broadbent *et al.*, 1975) connected to the posterior end of the hard palate. The velum is raised by the palatal levator muscles which originate above and behind the soft palate near the opening of the Eustachian tube and insert at the midline of the soft palate. Velum elevation is also assisted by the palatal tensor muscles. The velum's posterior end moves up and back toward the adenoids of the posterior pharyngeal wall. At the same time, the lateral pharyngeal walls move medially against the sides of the velum to prevent airflow. The velum is lowered when the palatal levator muscles relax while the palatoglossus and the palatopharyngeus muscles contract.

Beginning about 3 cm from the posterior end of the nasal cavity, the nose is divided into two passages by a septum which consists of a bony part and a mobile cartilaginous part with the anterior wall formed by the external part of the nose. Figure 2.1 shows the lateral view of the nasal cavity. Anteriorly, the passages communicate with the outside environment through the nostrils; posteriorly, they open into the nasopharynx. The roof consists of the ethmoid bone and the body of the sphenoid which contains the sphenoid sinuses. The floor is formed by the hard palate. The

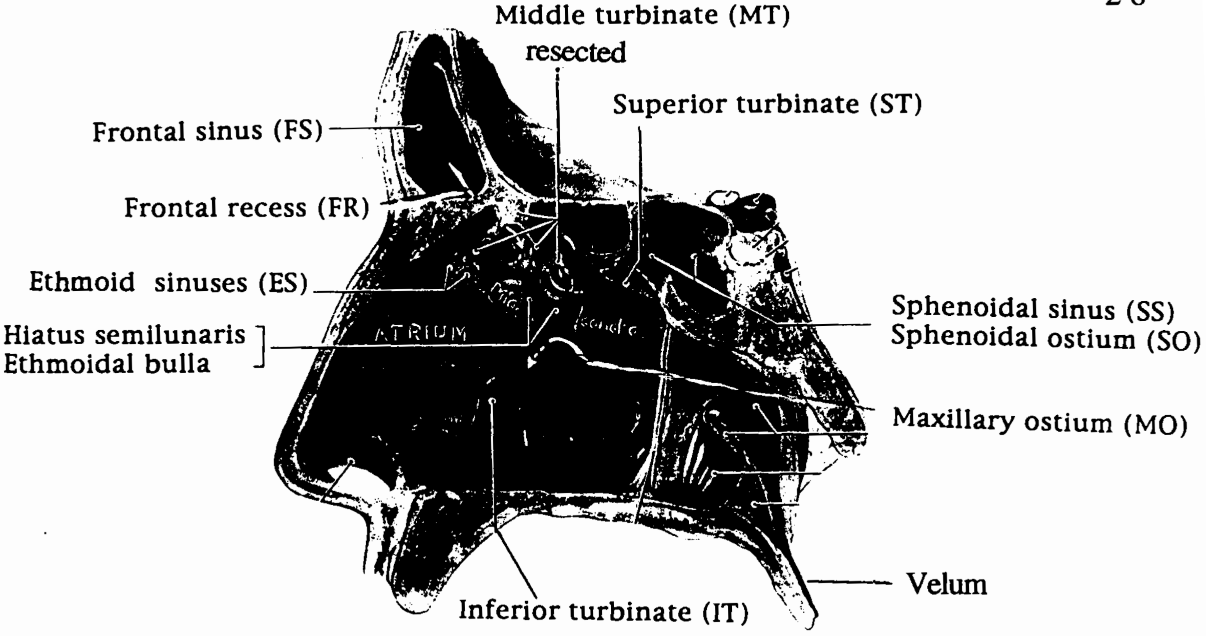


Figure 2.1: Lateral view of the nasal cavity with paranasal sinuses and ostia, frontal recess, ethmoid bulla, hiatus semilunaris, and turbinates labeled. (revised figure from Agur, 1991)

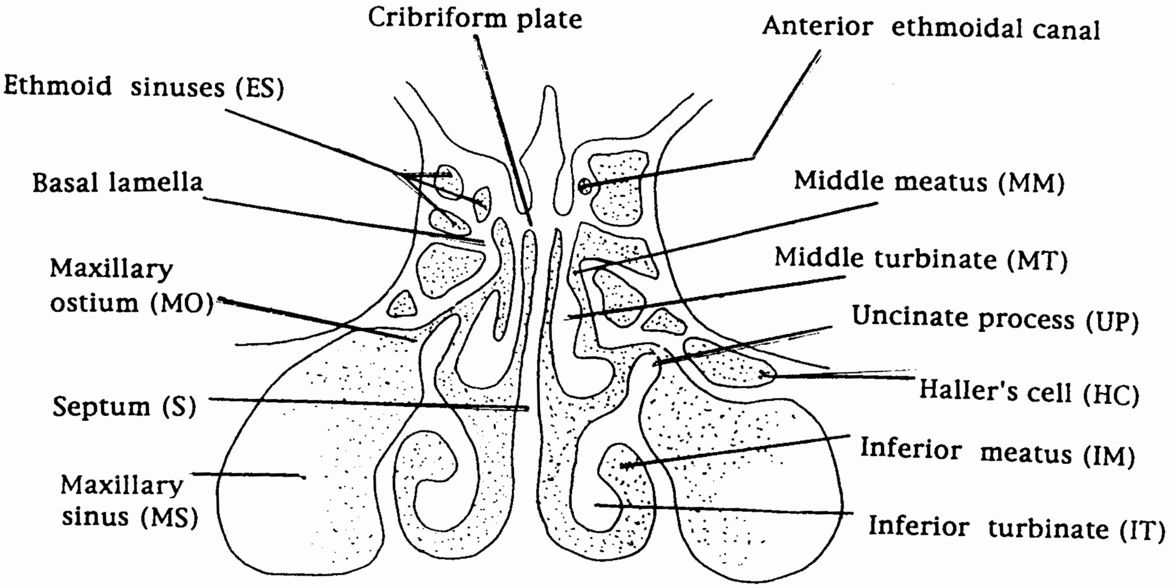


Figure 2.2: This coronal view of the nasal cavity was taken about the middle of the nasal cavity so that the frontal sinuses, the sphenoid sinuses, and the superior turbinates are not shown. The middle and inferior turbinates and the paranasal sinuses increase the total surface area of the nasal cavity. (revised figure from Rice, 1993).

lateral walls of the nasal passages consist of inferior, middle, and superior turbinates.

Figure 2.2 shows a coronal view of the nasal cavity. The inferior and middle turbinates curve inferiorly and medially from the lateral wall to cover the corresponding meatus. The superior turbinate, shown in Fig. 2.1, is smaller and is anterior to the sphenoid sinus. The turbinates augment the total surface area of the nasal cavity. Bjuggren and Fant (1964) made measurements of the effective cross-sectional area of the nose of an adult male cadaver. The total surface area of the nasal cavities was about 3.5 times as large as that of a cylindrical tube with the same total cross-sectional area. By magnetic resonance imaging (MRI) of four subjects, Dang *et al.* (1994) found asymmetry between the nasal passages. The area function of the nasal tract varied among subjects but all had greater area in the region posterior to the septum. On the other hand, the middle region had a much larger average circumference of 20.2 cm, compared with 5.8 cm and 8.6 cm for the posterior and anterior portions, respectively. The average length of the nasal tract was 11.6 cm with a standard deviation of 0.13 cm and the average volume was 25.5 cm^3 with a standard deviation of 9.1 cm^3.

Paranasal sinuses are connected to the lateral wall of the main nasal passages through narrow bony pathways called ostia. They are the maxillary, ethmoid, frontal, and sphenoid sinuses. The maxillary sinus with its base as part of the lateral nasal cavity and the apex in the zygomatic process is inferior to the orbital cavity. The maxillary ostium, elliptically shaped with the long axis oriented horizontally, is found at the junction of the floor of the orbit and the medial maxillary wall. It is half way between the anterior and posterior maxillary walls. In one study (May *et al.*, 1990), 10% of the patients had an accessory ostium in the posterior fontanelle due to the breaking down of the membrane; however, mucociliary flow is not directed toward this ostium.

The ethmoid sinus has large anatomical variations since it can consist from 4 to 17 air cells (Van Alyea, 1939). The posterior ethmoidal cells open into the superior meatus and the anterior ethmoidal cells open into the middle meatus. The roof of the ethmoid sinus is the orbital plate of the frontal bone.

The frontal sinus is superior to the orbital cavity and is a cavity in the frontal bone connected to the frontal recess. The frontal recess is at the highest part of the middle meatus behind the anterior attachment of the middle turbinate. It is enclosed anteriorly by the agger nasi ethmoid cells, medially by the middle turbinate, and posteriorly by the ethmoidal bulla. Generally, the recess receives secretions from the agger nasi, supraorbital ethmoid cells, and the frontal sinus. Then it passes the secretion into the superior ethmoidal infundibulum or directly into the nasal cavity. The agger nasi cells generally originate from the anterior superior portion of the infundibular groove. It may form part of the anterior wall of the frontal recess and drainage pathway of the frontal sinus.

The infundibulum is the region where secretions from the frontal, maxillary, and anterior ethmoid sinuses converge. It is bounded anteromedially and anteroinferiorly by the uncinate process, laterally by the medial wall of the maxillary sinus and the lamina papyracea, and posteriorly by the bulla ethmoidalis. Medially, it communicates with the middle meatus through the hiatus semilunaris. Anteriorly and superiorly, the infundibulum forms the frontoethmoidal recess in the majority of the people, or it may freely communicate with the nasofrontal duct (Messerklinger, 1987).

The sphenoidal sinus is in the body of the sphenoid bone. It opens into the sphenoethmoidal recess through an anterior opening. The sinus extends laterally inferior to the optic nerve, inferiorly to the pterygoid process, and posteriorly inferior to the hypophysis cerebri. The septum separating the right and left sphenoidal sinuses often is deflected to the side.

Dang *et al.* (1994) determined the average volume for the sinuses: right and left frontal sinuses averaged 2.6 cm^3 and 3.6 cm^3, respectively; right and left sphenoid sinuses averaged 9.0 cm^3 and 8.9 cm^3, respectively; right and left maxillary sinuses averaged 17.3 cm^3 and 15.7 cm^3, respectively. The volume of the ethmoid sinuses was difficult to measure since they have many air cells. For a single subject, the radii of the sphenoid and maxillary sinus ostia ranged between 0.10 cm and 0.15 cm; the lengths of the ostia ranged between 0.29 cm and 0.45 cm.

2.2 Acoustics of Nasal Coupling

Speech production theory can predict how changes of the vocal tract configuration affect acoustic spectra.

2.2.1 Pole and Zero Due to the Main Nasal Passages

During vowel production, the glottal source can be modeled by a volume-velocity source with high internal impedance since for most glottal vibrations the impedance of the glottis is much larger than that of the vocal tract. To a first approximation, the source at the glottis is dependent only on the glottal area waveform and not the shape of the vocal tract. The vocal tract can be modeled as an acoustic tube with a cross-sectional area that varies along its length. Since the cross dimensions are small compared to the wavelengths under consideration, it is valid to assume that only uniform plane waves are propagating in the vocal tract. For a schwa vowel, the vocal tract can be modeled roughly by a uniform tube opened at one end with the volume-velocity source, U_s, at the other end. When no loss is assumed, the transfer function from the glottal source U_s to the volume velocity U_o at the open end is

$$T(\omega) = \frac{U_o}{U_s} = \frac{1}{\cos k\ell} \qquad 2.1$$

where $k = \omega/c$, $\omega = 2\Pi*\text{frequency}$, c = speed of sound, and ℓ = length of the tube. The transfer function is characterized by evenly-spaced poles. As the vocal-tract shape changes to form different vowels, the formants shift accordingly.

For a nasalized vowel, there is coupling between the nasal cavity and the rest of the vocal tract. Asymmetry of the nasal passages may introduce pole-zero pairs (Fujimura and Lindqvist, 1964; Lindqvist-Gauffin and Sundberg, 1976; Dang *et al.*, 1994). From their dual-tube model, Dang et al. concluded that the transfer function would have a pole-zero pair in the 2 kHz to 2.5 kHz range. Apparently, then, the dual passages do not affect the transfer function in the vicinity of the first formant. Assuming symmetry of the two channels in the nasal cavity, one can model the velopharyngeal opening and the cavity as a constriction followed by a side branch, as schematized in Fig. 2.3. As a result of the side branch, extra poles and zeros are added to the all-pole system of the oral vowel (Fant, 1960) with the transfer function $T(\omega)$ in Eq. 2.1.

The location of the pole-zero pairs can be determined by graphical analysis of susceptances (Fujimura, 1960). However, only the pair at the lowest frequencies is examined since it occurs more consistently than the other pairs. In Fig. 2.3, B_n, B_m, and B_p are the driving-point susceptances at the junction looking toward the nose, the mouth, and the pharynx, respectively. The natural frequencies, the formants and the extra poles, of the entire system for the transfer function from the glottal source to the total output from the

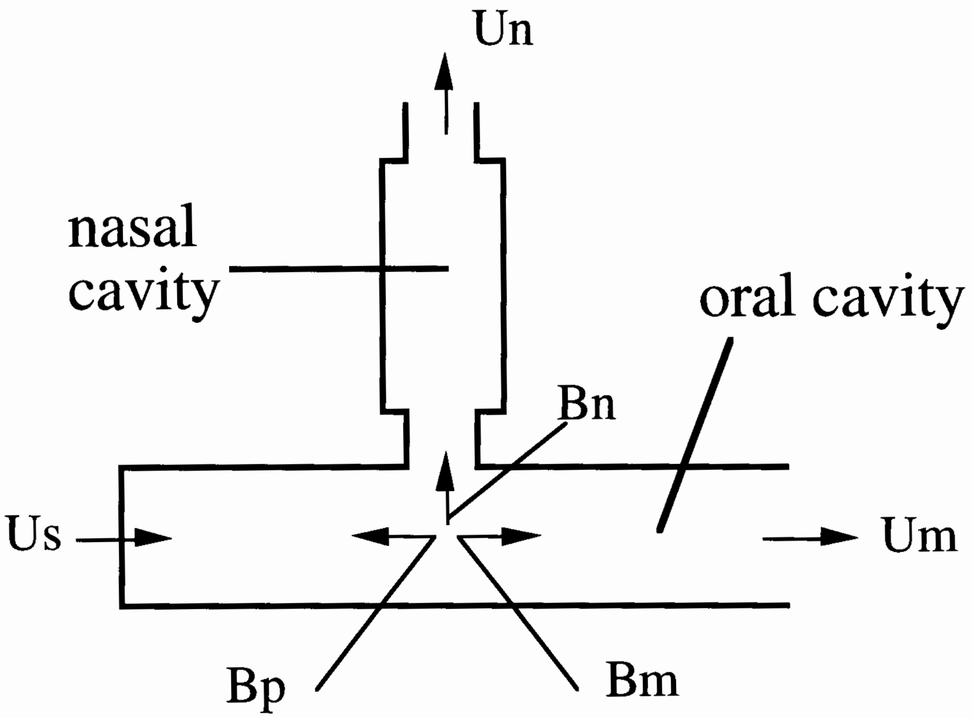


Figure 2.3: Simplified model of the acoustic system assumes symmetry during a nasalized vowel and uses susceptance analysis. B_p, B_m, and B_n are the susceptances looking toward the pharynx, the mouth, and the nose, respectively.

mouth and the nose occur when $B_n+B_m+B_p$ is equal to zero. The susceptance B_n is the (negative) reciprocal of the reactance X_n, which can be approximated as a series connection of an acoustic mass, representing the velopharyngeal opening, and the reactance of the nasal cavity proper. Based on the experimentally measured transfer function of the nasal cavity from a sweep-tone source above the closed v-p port to the nostril output (Lindqvist and Sundberg, 1972), the susceptance B_n is estimated to have zeros at about 500 and 2000 Hz, which gives information on constructing the B_n curves. Figure 2.4 models the vocal tract as a transmission line with the nasal tract represented by a black box. If low frequencies and small losses are assumed, a lumped element circuit can be used to represent the side branch. The acoustic mass and acoustic capacitance, M_s and C_s, respectively, are associated with the properties of the side branch; the coupling element M_c is equal to $\rho L_{vp}/A_{vp}$, where ρ is the density of air, L_{vp} and A_{vp} are respectively the length and cross-sectional area of the velopharyngeal opening. Plots of the estimated B_n versus frequency are shown in Fig. 2.5 for two values of A_{vp}, 0.3 cm^2 and 0.8 cm^2, with L_{vp} as 2 cm and cross-

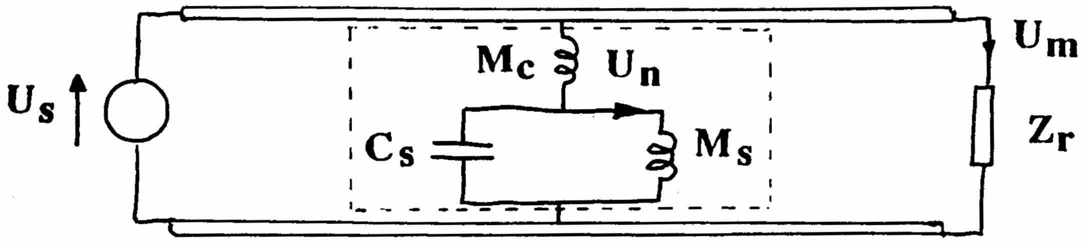


Figure 2.4: A lumped element circuit is used to represent the nasal branch at low frequencies and small losses. M_s and C_s are associated with the side branch, M_c is a function of the size of the opening to the side branch. This circuit is connected across a transmission line representing the vocal tract. Z_r is the radiation impedance.

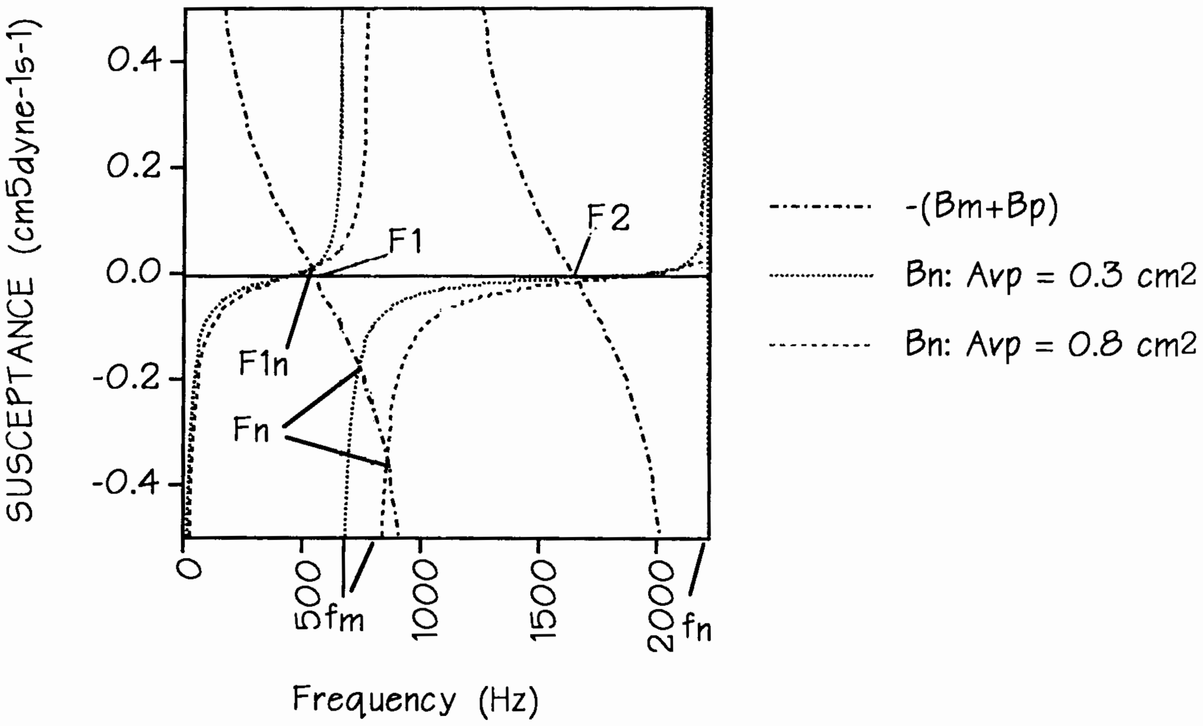


Figure 2.5: Susceptance curves B_n and $-(B_p+B_m)$, used to determine shifted F1 (F1n), nasal pole (Fn), and nasal zero (from fm and fn) are shown. The vowel configuration for this example has uniform cross-sectional area. Susceptance curves for velopharyngeal areas of 0.3 cm^2 and 0.8 cm^2 are shown.

sectional area of the nasal cavity as 2 cm^2. The frequency at which B_n goes to infinity is dependent on the size of the velopharyngeal opening: when the opening is larger, M_c is smaller, and B_n goes to infinity at a higher frequency, f_m.

The sum B_m+B_p is determined by solving the one-dimensional wave equation for the particular vocal-tract shape. The mouth at the end of the oral cavity is assumed to be open while the glottis at the end of the pharynx is assumed to be closed. In addition to the cross-sectional areas, B_m and B_p are dependent on L_m and L_p, which are the lengths of the oral cavity and the pharynx, respectively, and each is taken here to be 7.5 cm. The points where the $-(B_m+B_p)$ curve crosses the abscissa are the formant frequencies F1 and F2 of the oral vowel. The lowest frequency at which the sum goes to infinity is due to B_p and the next frequency at which the sum goes to infinity is due to B_m. In Fig. 2.5, the natural frequencies of the entire system are at the intersections of B_n and $-(B_m+B_p)$. The first formant frequency, F1, is shifted to $F1_n$ and the nasal pole is F_n.

The zero of U_m/U_s (where U_m = volume velocity at the mouth) occurs at frequency f_m for which the side branch acts as a short when B_n goes to infinity. The zero of U_n/U_s (where U_n = volume velocity at the nose) occurs at a frequency f_n where B_m goes to infinity. The frequencies of the two zeroes are needed to calculate the zero for the transfer function from the glottal source to the total output from the mouth and the nose. We assume the outputs from the mouth and nose are physically close to each other so that the distance between them is much smaller than a wavelength. The output is then U_m+U_n with the assumption that the two individual outputs do not influence each other and change the values of the susceptances B_n and B_m. At low frequencies, the side branch introduces a pole and a zero to the transfer function from the glottal source to the total output in addition to the first formant (Fant, 1960). The transfer function at the mouth output at low frequencies can be represented by the following expression:

$$\frac{U_m}{U_s} = \frac{a\left(\frac{s^2}{\omega_m^2}+1\right)}{\left(\frac{s^2}{\omega_a^2}+1\right)\left(\frac{s^2}{\omega_b^2}+1\right)} \qquad 2.\ 2$$

where $a = M_n/(M_m+M_n)$. The impedance of the nasal cavity, M_n is the acoustic mass of the constriction M_c plus that of the side branch

M_s; M_m is the acoustic mass of the main tube from the velum to the mouth. The frequencies $f_a = \omega_a/2\Pi$ and $f_b = \omega_b/2\Pi$ are the lowest natural frequencies of the circuit and $f_m = \omega_m/2\Pi$ is the frequency at which the impedance looking into the side branch is zero. The transfer function at low frequencies at the nose output is:

$$\frac{U_n}{U_s} = \frac{(1-a)\left(\frac{s^2}{\omega_n^2}+1\right)}{\left(\frac{s^2}{\omega_a^2}+1\right)\left(\frac{s^2}{\omega_b^2}+1\right)} \qquad 2.3$$

with a zero at $f_n = \omega_n/2\Pi$, the frequency at which the impedance looking into the main tube toward the mouth is zero.

By adding Eqs. 2.2 and 2.3 one obtains the overall transfer function:

$$\frac{U_m+U_n}{U_s} = \frac{\left(\frac{s^2}{\omega_z^2}+1\right)}{\left(\frac{s^2}{\omega_a^2}+1\right)\left(\frac{s^2}{\omega_b^2}+1\right)} \qquad 2.4$$

where the lowest zero of the total output from both the mouth and the nose occurs at $f_z = \omega_z/2\Pi$. Note that the frequencies of the poles remain the same. From Eqs. 2.2, 2.3, and 2.4, f_z can be derived as:

$$f_z = \frac{1}{\sqrt{\frac{a}{f_m^2} + (1-a)\frac{1}{f_n^2}}}. \qquad 2.5$$

By manipulating Eq. 2.5,

$$f_z = f_m \frac{\sqrt{1 + \frac{M_m}{M_n}}}{\sqrt{1 + \left(\frac{f_m}{f_n}\right)^2 \frac{M_m}{M_n}}} \qquad 2.6$$

Table 2.1: Calculated frequency domain parameters for nasalized vowels according to the size of velopharyngeal port by Stevens (in preparation). The lowest zero of the total system is f_z, and the zero frequencies at the mouth and at the nose are f_n and f_m, respectively. F_n is the frequency of the pole introduced by the nasal coupling. $F1_n$ is the shifted first formant. The last column shows the height of the extra peak at F_n above that of the non-nasal reference.

Vowel/ opening (cm^2)	$F1_n$ (Hz)	f_n (Hz)	f_m (Hz)	f_z (Hz)	F_n (Hz)	$\frac{M_m}{M_n}$	Extra peak (dB)
i / 0.3	400	2600	760	1260	810	2.0	13.5
ɛ / 0.3	510	2170	760	970	830	0.8	7.5
ɔ/ 0.3	530	1670	760	910	770	0.6	7.3
ɑ / 0.3	600	1745	760	870	810	0.4	2.9
ɑ / 0.8	610	1745	920	1220	910	1.5	12.1

indicating that the antiformant is higher than f_m since $f_m < f_n$.

The values of F_n and f_z (with its corresponding f_n and f_m) as well as the shifted formant $F1_n$ have been calculated by Stevens (in preparation) using susceptance plots for a velopharyngeal opening of 0.3 cm^2 for several vowels, and also for a larger opening of 0.8 cm^2 for the vowel /ɑ/, as shown in Table 2.1. The anatomical dimensions on which the calculations at low frequencies are based are as follows. The velopharyngeal opening has a length of 2 cm. The nasal cavity proper has a length of 6 cm and an area of 2 cm^2. The nostrils have a length of 1 cm and area 0.5 cm^2. For a velopharyngeal opening of 0.3 cm^2, the nasal pole (in the frequency range 770-830 Hz) is below the nasal zero (in the range 870-1260 Hz). Across vowels, the location of the extra pole is fairly similar. With a large opening for the vowel /ɑ/, the nasal pole frequency is 910 Hz. The last column shows the height of the extra peak above that of the non-nasal reference introduced by adding the nasal pole and zero with bandwidths around 200 Hz. A larger opening causes the zero to increase in frequency much more than the pole and therefore introduces a more prominent extra peak, as shown by the two opening sizes for the vowel /ɑ/. A large velopharyngeal opening for this vowel can introduce a prominence as large as 12.1 dB. Even with a small opening of 0.3 cm^2 for the vowel /i/, a prominence of 13.5 dB can be introduced. For a given velopharyngeal area, the perturbation differs depending on the vowel height.

Theoretically, coupling to the main nasal passages during vowel nasalization introduces to the transfer function from the glottal source to the total output of the mouth and the nose additional poles and zeros. By using susceptance analysis based on the nasal cavity transfer function from sweep-tone experiment and lumped circuit elements at low frequencies, the pole and zero frequencies of the transfer function can be approximated. The prominence of the nasal peak also can be calculated for different velopharyngeal openings. The nasal pole occurs up to 910 Hz with a prominence of about 13 dB.

2.2.2 Pole and Zero Due to Sinuses

From acoustic measurements of the nasal cavity transfer function, Lindqvist-Gauffin and Sundberg (1976) showed that a nasal pole-zero pair often occurs at frequencies below the first formant, especially for low vowels. Maeda (1982a) showed through computer simulations of the nasal tract that modeling a sinus as a side cavity to the nasal tract introduces a pole-zero pair below the first formant around 450 Hz. He found that when the sinuses are closed, this extra peak is weaker than the extra peak due to coupling to the main nasal passage. Other investigators have suggested that the sinuses of the nose may influence the nasal spectra greatly (Fujimura and Lindqvist, 1964; Fant, 1980; Dang *et al.*, 1994, Båvegård *et al.*, 1993).

A sinus can be modeled as a side cavity of the nasal tract system, as shown in Fig. 2.6. It is connected to the main nasal passage by an ostium, which has a radius of about 0.13 cm (Dang *et al.*, 1994). The narrow coupling allows one to model the sinus and its ostium as a Helmholtz resonator with resonance frequency of:

$$f_{zs} = \frac{c}{2\Pi}\sqrt{\frac{S_o}{VL_o}} \qquad 2.7$$

where S_o is the cross-sectional area of the ostium in cm^2 , L_o is its length in cm, and V is the volume of the paranasal cavity in cm^3. A sinus would introduce to the total system an extra zero and an extra pole at low frequencies. The zero frequency may be approximated by f_{zs} in Eq. 2.7 since the acoustic mass of the sinus ostium is much larger than that of the nasal tract and the vocal tract. The corresponding pole which determines the location of the extra peak is at a frequency lower than the zero due to the acoustic mass from the nasal cavity anterior to the sinus ostium in parallel with the acoustic masses of the nasal cavity posterior to the ostium, the velopharyngeal opening, and the oral cavity.

By using the anatomical dimensions of the sinuses, Dang *et al.* (1994) estimated using MRI and Eq. 2.7 for one subject the resonance frequency, f_{zs}, of the sphenoid and maxillary sinuses to be between 534 and 989 Hz. (Because of their small cavities, the frontal and ethmoid sinuses would affect the acoustic properties at much higher frequencies.) Calculations also show that the frequency of the pole is lower than that of the zero by 33 Hz to 127 Hz. These calculated frequencies of the nasal pole and zero due to the sinuses are higher than those observed in the spectral analysis. This discrepancy may be because of the wide variation of sinuses among subjects of the different studies and the inaccurate measurements of the ostium dimensions due to the limited MRI spatial resolution with sagittal views of the sinus cavity taken at 0.3 cm intervals. This limited spatial resolution may also obscure the boundary between the sphenoid and the ethmoid sinuses adding error to the volume measurement. Although dimensions of the sinuses have also been derived from cadaver specimens (Takeuchi *et al.*, 1977; Masuda, 1992), absence or dehydration of mucous membrane would

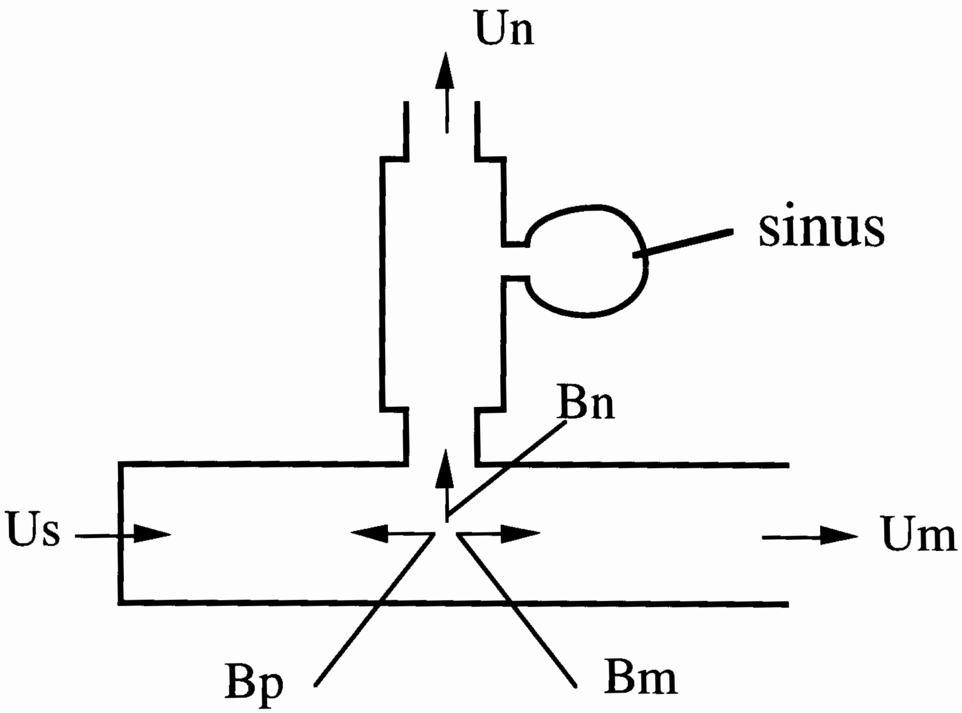


Figure 2.6: The model of the nasalized vowel is modified when the paranasal sinuses are included. The sinus with its large cavity and small ostium acts as a Helmholtz resonator.

cause deviation of the measurements from those of the living. If one assumes that the pole is at 250 Hz, the zero at 500 Hz (Hattori *et al.*, 1958), and the bandwidths are 160 Hz, (Lindqvist and Sundberg, 1972), the pole-zero pair would introduce a peak of 2.4 dB at 250 Hz. Boosting the amplitude of harmonics at frequencies below the first formant would tend to flatten the apparent spectral peak of F1.

2.2.3 Bandwidth Widening

The losses due to radiation, cavity wall vibration, viscous friction, and heat conduction contribute to the bandwidth of the formants. The radiation impedance of the mouth and the nose openings can be approximated by that of a piston in a sphere. From the analytical expression derived for the radiation impedance (Morse, 1948; Fant, 1960) and a perturbation calculation (Staelin *et al.*, 1994) for a uniform tube of area A and length ℓ, the formant bandwidth due to radiation loss may be approximated by:

$$B_r = \frac{F^2 A}{\ell c} K(\omega) \tag{2.8}$$

where F is the formant frequency, c is the speed of sound, and K, a function of $\omega = 2\Pi$ *frequency, accounts for the baffling effect of the head. By approximating the head as a simple source, K can be considered to be unity. The relatively small opening of the nostrils (0.5 cm^2) compared to that of the mouth (3 cm^2) causes B_r for a nasalized vowel to be small, especially at low frequencies.

The non-rigid, fleshy vocal tract walls introduces energy losses due to the resistive component of the impedance. According to House and Stevens (1958), for a uniform tube of cross-sectional area A and cross-sectional perimeter S, the addition to the bandwidth of a formant is:

$$B_w = \frac{G_{sw} S \rho c^2}{2\Pi A} \tag{2.9}$$

where G_{sw} is the specific acoustic conductance of the walls and ρ is the density of air. From Eq. 2.9, for a given formant frequency,

$$B_w \; \alpha \; \frac{S}{A}. \tag{2.10}$$

As mentioned above, viscous friction at the wall also contributes to the loss of energy. The equivalent resistance per unit length of the tube is:

$$R_v = \frac{S}{A^2}\sqrt{\frac{\omega\rho\mu}{2}} \qquad 2.11$$

where μ is the viscosity coefficient (Flanagan, 1972). The contribution of viscosity to the bandwidth of a uniform tube closed at one end and opened at the other is:

$$B_v = \frac{R_v\ A}{2\Pi\rho}\ . \qquad 2.12$$

From Eqs. 2.11 and 2.12,

$$B_v\ \alpha\ \frac{S}{A}\ . \qquad 2.13$$

Heat conduction at the walls of the tube also causes loss in the vocal tract. The equivalent conductance per unit length is:

$$G_h = S\frac{0.4}{\rho c^2}\sqrt{\frac{\lambda\omega}{2c_p\rho}} \qquad 2.14$$

where λ is the coefficient of heat conduction and c_p is the specific heat of air at constant pressure (Flanagan, 1972). The bandwidth due to heat conduction of a uniform tube closed at one end and opened at the other is:

$$B_h = \frac{G_h\rho c^2}{2\Pi A} \qquad 2.15$$

using Eqs. 2.14 and 2.15,

$$B_h\ \alpha\ \frac{S}{A} \qquad 2.16$$

The total surface area of the nasal cavity is augmented due to the turbinates in the lateral walls of the nasal passages. Dang *et al.* (1994) used MRI to measure the 3-dimensional geometry of the nasal tract. They found a shape factor (Fant, 1960)

$$S_f = \frac{S}{\sqrt{4\Pi A}} \qquad 2.17$$

of 1, 4, and 2 for the posterior, middle, and anterior portion of the nasal cavity, respectively. The average cross sectional area A for the posterior, middle, and anterior portion of the nasal cavity obtained by volume/length was 2 cm^2, 2.4 cm^2, and 1.4 cm^2, respectively. Assuming that the oral tract has a cross-sectional area of 4 cm^2 and a shape factor of 1, the posterior, middle, and anterior portion of the nasal cavity would cause the formant bandwidth to increase and therefore the peak amplitude to decrease by 3 dB, 14 dB, and 11 dB, respectively, relative to that of the cylindrical shaped oral cavity. On the average the bandwidth would more than triple and the amplitude would decrease by 11 dB for the whole nasal passage. However, the assumption that all of the acoustic energy goes to the nasal cavity was made. Since during nasal vowel production some of the energy goes to the oral cavity as well, 11 dB overestimates the amplitude change. More careful study of the energy distribution for coupling to the nasal cavity is necessary to better quantify the change in the formant bandwidth and amplitude. Nevertheless, for a nasalized vowel, the total formant bandwidth is expected to increase relative to that of the oral vowel due to a larger S_f and a smaller A.

On the other hand, there are other losses that were not taken into account by the S_f and A. The sinuses augment the total surface area of the nose. Constrictions and nostril hairs would also introduce nasal resistance, which would increase the bandwidth of the formants. Therefore, the amount by which the first formant amplitude decreases is underestimated if only the surface factor of the main nasal passage is considered. As an approximation, we assume an 11 dB amplitude decrease with losses.

The increase of the first formant bandwidth due to coupling to the nasal cavity can also be approximated by calculations based on a model in which elements representing the velopharyngeal opening and the nasal tract are introduced across the transmission line that represents the vocal tract at low frequencies. To simplify the procedure, calculations were made with the resistive element for the nasal tract either in parallel or in series, as shown in the Appendix. Losses due to heat conduction at the walls and yielding walls are represented by R_l in parallel with the inductance of the nasal cavity, as shown in Fig. A.1. Loss due to viscous friction at the walls is represented by R_o in series with the inductance of the nasal cavity, as shown in Fig. A.2. The values of these resistances were estimated from the bandwidth of the lowest resonance of the nasal cavity, measured from the reponse curves reported by Lindqvist and Sundberg (1972). The calculation of the lowering in amplitude of the first formant, $\Delta A1$, for the resistance in parallel and in series are shown in Fig. 2.7 and Fig. 2.8, respectively. The value of A_{vp} is as large as 2.5 cm^2 since that is the maximum size measured during nasalization according to X-ray measurements (Björk *et al.*, 1961, Moll and Daniloff, 1970). A resistance in series affects the amplitude more than a resistance in parallel for vowels with low F1 and for vowels with high F1 and small A_{vp}. As the velopharyngeal opening increases in size, the parallel resistance dominates in vowels with higher F1. Even though high vowels have been found to require a smaller A_{vp} to reach the same nasality level as for low vowels (Maeda, 1982b), the exact size of A_{vp} for various vowel types has not been measured. However, a minimum A_{vp} of 0.2 cm^2 has been determined for a non-nasal sound to be perceived as abnormal (Warren, 1979). Also, Maeda (1982b) has approximated a small coupling of 0.4 cm^2 for high vowels and 1.6 cm^2 for low vowels. One may assume that mid vowels may have A_{vp} about 1 cm^2. Most likely, the actual $\Delta A1$ is between that calculated for the model with the resistance in parallel and that with the resistance in series. From Figs. 2.7 and 2.8, for high vowels $\Delta A1$ due to nasalization would be between 3.3 dB and 6.4 dB ; for mid vowels it would be between 6.5 dB to 6.6 dB; for low vowels it would be between 5.7 dB and 8.9 dB. The average for the various vowel types is around 6.2 dB. If one assumes that A_{vp} can be any value less than 2.5 cm^2, for high vowels the maximum $\Delta A1$ due to nasalization would be between 5.9 dB and 10.0 dB ; for mid vowels it would be between 6.3 dB and 6.6 dB; for low vowels it would be 17.4 dB since the 23 dB from the series resistance is unlikely. 23.0 dB. The average for the various vowel types calculated is around 10.5 dB, which is similar to the result estimated above, using the shape factor and cross-sectional area of the nasal cavity.

2.3 Summary

A main effect of nasalization is the introduction of a pole-zero pair above the first formant so that the sharp spectral peak of F1 is replaced by a broader prominence consisting of two peaks. Theoretical analysis shows that the pole is usually below the zero. Across vowels, the location of the extra pole is fairly constant. A larger nasal opening than that of a non-nasal vowel corresponds to a

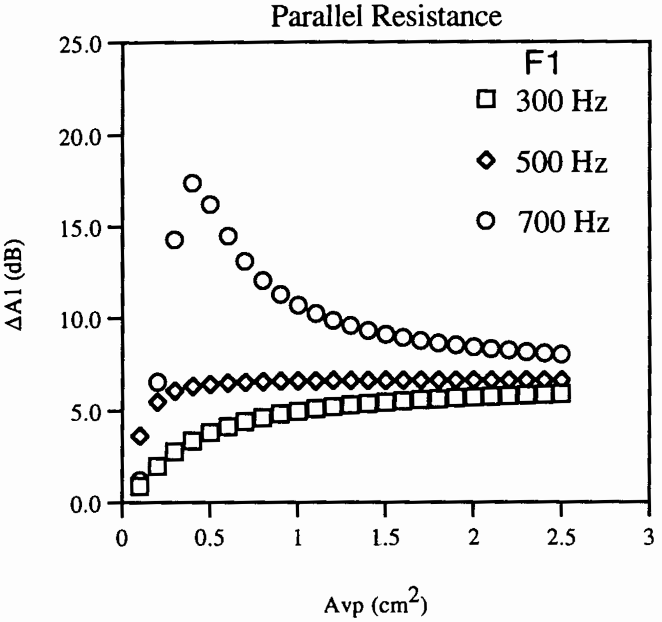


Figure 2.7: The lowering of amplitude for the first formant, ΔA1, when a parallel resistance representing losses in the nasal cavity due to heat conduction and yielding walls is introduced to the circuit in Fig. 2.4. The calculation was done for F1 at 300 Hz (squares), 500 Hz (diamonds), and 700 Hz (circles).

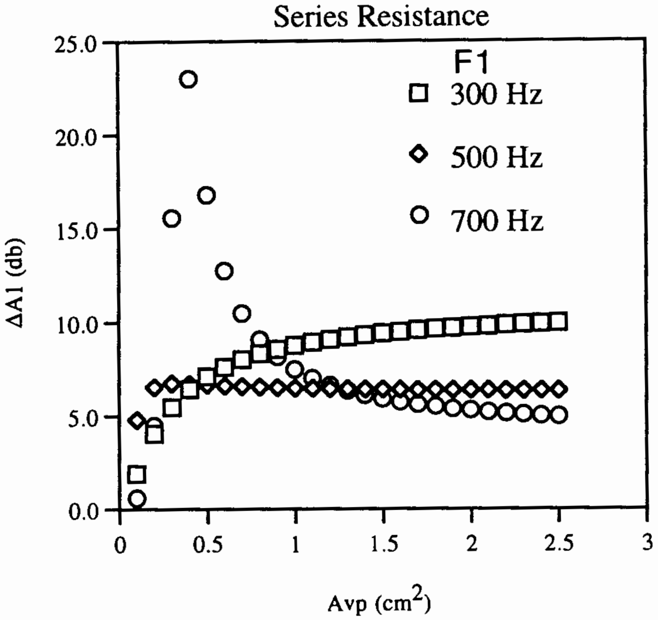


Figure 2.8: The lowering of amplitude for the first formant, ΔA1, when a series resistance representing loss in the nasal cavity due to viscous friction at the walls is introduced to the circuit in Fig. 2.4. The calculation was done for F1 at 300 Hz (squares), 500 Hz (diamonds), and 700 Hz (circles).

shift of the pole-zero pair to higher frequencies, with a greater effect on the zero frequency. Therefore, a larger velopharyngeal opening causes a greater separation between the pole and zero, and thus a greater peak amplitude at the frequency of the pole. The modification to the spectrum at the extra peak may be as large as 12 to 13 dB, depending on the vowel type.

According to the Helmholtz resonator model, the sinuses, especially the maxillary and sphenoid sinuses, are expected to introduce additional pole-zero pairs in the low frequencies, below that of the first formant, except when F1 is low. The pole is expected to be below the zero. Based on the frequencies and bandwidths of the pole-zero pair from acoustic analysis and nasal cavity response curves, the pair may introduce an extra peak at low frequencies that has an amplitude of about 2.4 dB.

A bandwidth of the F1 prominence is partly due to losses at the large surface area of the nasal cavity caused by wall impedance, viscosity, and heat conduction. The large shape factor due to the septum and turbinates of the nasal passages contributes to the losses. In addition, the sinuses augment the total surface area of the nose. Constrictions and nostril hairs would also introduce nasal resistance, which would increase the bandwidth of the formants. It is estimated somewhat arbitrarily that because of these increased losses, coupling to the nasal cavity may lower the F1 amplitude by about 11 dB. Calculations based on circuit theory by introducing resistance in parallel and in series with the inductance of the nasal cavity further support this amount by which A1 is lowered.

This study is only concerned with modifications of the spectra around the first formant. Even though in the higher frequencies, nasalization may introduce shifts in formants, modification of formant amplitudes, and additional spectral peaks, the effects are not as obvious as those in the vicinity of the first formant for various vowels and speakers. Also, the asymmetry of the main nasal passages can be ignored in the low frequencies since the extra pole-zero pair due to this asymmetry is expected to be well above the frequency of the first formant.

Chapter 3

Acoustic Analysis of Nasalized and Non-nasalized Vowels

Spectral characteristics motivated by speech production theory (Chap. 2) were quantified for utterances containing nasalized vowels. Measurements of the recorded speech of adult speakers were made by examining closely in the vicinity of the first formant for extra peak prominence and also changes of the formant bandwidth.

3.1 Data Acquisition

The recordings were made with a ceiling-hung Electro-Voice model 054 omnidirectional dynamic microphone placed approximately six inches in front of the speaker's lips in a partially sound attenuated room. For this microphone position, the distance from the microphone to the mouth and the nose was about equal, and the relative effect of reverberation was minimized. The microphone output was connected to a Shure microphone mixer followed by a Nakamichi LX-5 cassette recorder without Dolby and dbx so that the onsets were not distorted. All signals were lowpass filtered at 4.8 kHz using a TTE passive seven-pole elliptical low-pass filter, digitized at 10 kHz with 12-bit samples, and stored on a VAX-750 computer disk. The vowels were extracted from the digitized speech by editing the waveform files and were analyzed by using computer software. The frequency-domain analysis of the vowels was done by using KLSPEC93, which is a revision of the software package developed by Klatt (1984). This research utilized the discrete Fourier transform (DFT) magnitude and spectrogram-like-spectra from the software package. Spectra throughout the vowel were generated by using a 30-ms Hamming window and computing a 512-pt. DFT.

3.2 Acoustic Correlates

3.2.1 A1-P1 (Difference between A1 and P1 in dB)

As observed in Chapter 2, the large shape factor S_f due to the septum and turbinates of the nasal passages contribute to broadening the first formant and to lowering the formant amplitude A1 (dB). The amplitude was estimated as the amplitude of the peak harmonic closest to the expected F1. Theoretically, it is expected that coupling to the nasal cavity may lower A1 by about 11 dB. Also, a large velopharyngeal opening shifts the nasal zero to a higher frequency so that it moves further away from the nasal pole, causing a more prominent extra peak above the first formant. The spectral effect of the extra peak was characterized by its amplitude, P1 (dB), which was estimated by using the amplitude of the highest peak harmonic around 1 kHz. In the non-nasalized vowels, if no extra peak was observed, P1 was measured by using the highest harmonic close in frequency to that in the nasalized vowel. From the calculations in Table 2.1, for the vowel /ɑ/ when the velopharyngeal opening is 0.8 cm^2, P1 can be about 12.1 dB greater for the nasal vowel than for the non-nasal vowel, and P1 can be 13.5 dB greater even when the opening is only 0.3 cm^2 for the nasalized vowel /i/ compared to its non-nasalized version. The effect of nasal zero was assessed indirectly through its influence on P1. Since both parameters A1 and P1 are changed by nasal coupling and they are inversely related to each other, they were lumped together by the difference A1-P1 (dB). A difference was used so that the measurements would be independent of speech intensity. A given vowel that is nasalized should have a smaller A1-P1 and it should be perceived to be more nasal. Theoretically, A1-P1 can decrease by as much as 23 dB when a vowel is nasalized with A1 lowering and P1 raising.

Spectra of vowels adjacent to nasal consonants uttered by normal hearing adults, vowels uttered by normal hearing children, and vowels produced by hearing-impaired children were analyzed by Chen (1995). The average frequency of the extra peak due to coupling to the nasal passages was found to be 950 Hz and 938 Hz for the adults and children, respectively. These measurements are close to the theoretical average value of 910 Hz for a large velopharyngeal opening for the vowel /ɑ/ (Stevens, in preparation). Chen (1995) also reported perceptual judgments of nasality for the children's utterances by two trained phoneticians. The utterances consisted mainly of low vowels. The correlation coefficient between the average perceptual judgment and A1-P1 was -0.82. This relatively high correlation coefficient showed that A1-P1 is a promising measure of nasality. The current work examines A1-P1 in more depth by making measurements in English and French vowels spoken by adult speakers.

3.2.2 A1-P0 (Difference between A1 and P0 in dB)

Theoretically, the paranasal sinuses, especially the maxillary and sphenoid sinuses, can introduce extra pole-zero pairs in the transfer function of the total system in a nasal vowel at frequencies below that of the first formant. A larger velopharyngeal opening would cause greater coupling to the nasal sinuses and introduce a more prominent peak at the lower harmonics in the spectrum of a nasal vowel. The amplitude of this low-frequency peak is designated as P0. Theoretically, this amplitude can increase by 2.4 dB (see Sec. 2.2.2). It is recognized, however, that increased open quotient of the glottal waveform can also introduce increased low-frequency amplitude in the form of a more prominent first harmonic. As the open quotient changes from 50% to 90%, H1 increased about 5 dB (Klatt and Klatt, 1990), assuming other aspects of the glottal waveform remain fixed. The location of the extra peak P0 (dB) for nasalized and non-nasalized vowels was chosen based on which harmonic has the greatest amplitude at the low frequencies. It could be the first, second, or the third harmonic, depending on the fundamental frequency of the speaker. Both coupling to the paranasal sinuses and a large open quotient would introduce greater losses and therefore increase the bandwidth of the first formant as well as lowering its amplitude, A1 (dB), by around 11 dB (see Sec. 2.2.3). Therefore, another acoustic correlate used for measuring the degree of nasal coupling is A1-P0 (dB): a nasal vowel would have a smaller A1-P0 by as much as 13 dB due to the lowering of A1 and rising of P0. If the OQ is increased and the first harmonic is used to measure P0, A1-P0 may be decreased further by 5 dB. A1-P0 would be especially useful for vowels that have their first and/or second formant close to the frequency of P1, since peaks close in frequency influence each other and therefore obscuring the A1-P1 measurement. On the other hand, A1-P1 would be more useful for vowels with F1 at a low frequency that obscures P0.

Past studies have reported an extra peak in the low frequency region for nasal vowels, particularly for low vowels (Hattori *et al.*, 1958). It was expected that this peak would be more observable in a language like French since the non-nasal-nasal vowel distinction in

French would cause the nasal vowels to be more nasalized than English vowels nasalized due to coarticulation (Delattre, 1965; Mrayati, 1975). In a preliminary study of French vowels, speech of three male native French speakers was examined. Acoustic analysis was carried out on vowels in words embedded in the carrier phrase "dites _______ pour moi." Two nasalized vowels, /ɑ̃/ and /ɛ̃/, and five non-nasalized vowels /ɑ/, /i/, /e/, /u/, and /æ/ with various numbers of repetitions were analyzed. A prominent extra peak below F1 was usually observed in the nasal vowels of all three speakers. For one speaker, the first harmonic was used to define the amplitude of the extra peak while for the other two speakers, the second harmonic was used. A1-P0 was determined every 20 ms throughout the vowel and the minimum, (A1-P0)$_{min}$, was obtained. Figure 3.1 shows (A1-P0)$_{min}$ compared across vowels and speakers by counting the number of times a given value of (A1-P0)$_{min}$ occurred. A sharp distinction was found between the nasal and non-nasal vowels, although some overlap did occur. In general the nasal vowels had smaller (A1-P0)$_{min}$ than the non-nasal vowels; almost all of the nasal vowels had negative (A1-P0)$_{min}$, indicating that

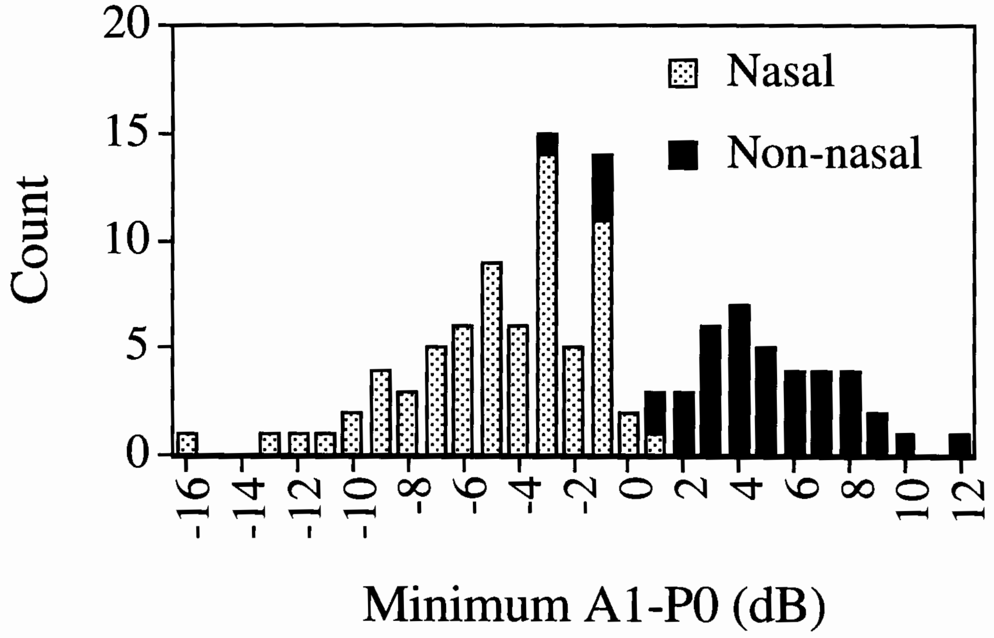


Figure 3.1: Comparison of the minimum values of A1-P0, (A1-P0)$_{min}$, for nasal and non-nasal French vowels of three male speakers by counting the number of times a given value of (A1-P0)$_{min}$ occurred.

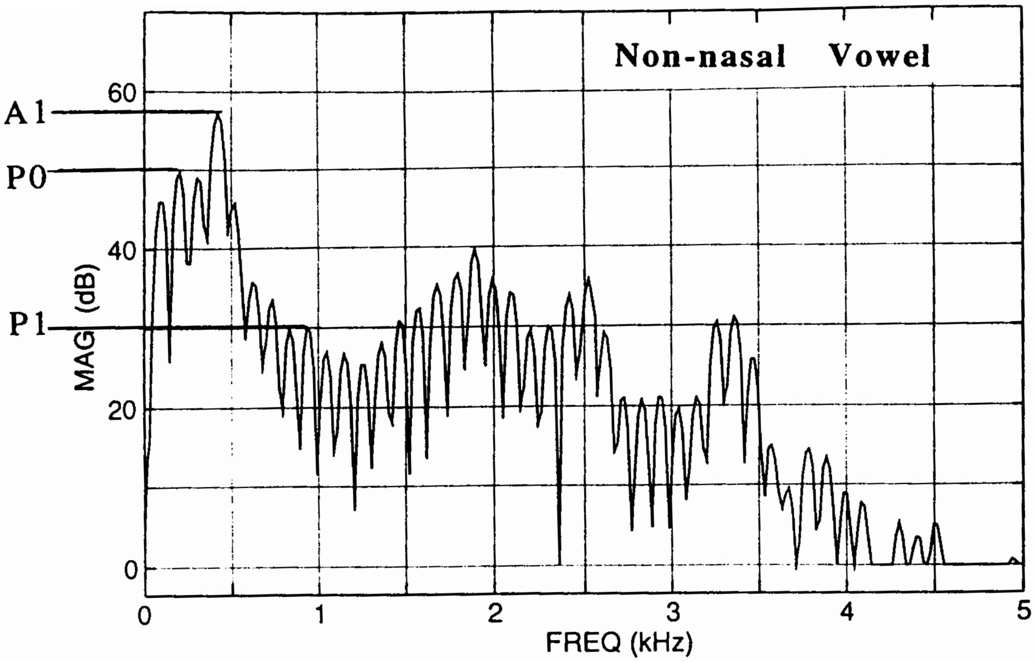
the peak at the low frequencies actually is larger than the first formant. A similar result is expected for nasalized English vowels, although A1-P0 may not be so distinctive between nasal and non-nasal vowels.

3.3 English Speakers: Minimal Pairs of Nasalized and Non-nasalized Vowels

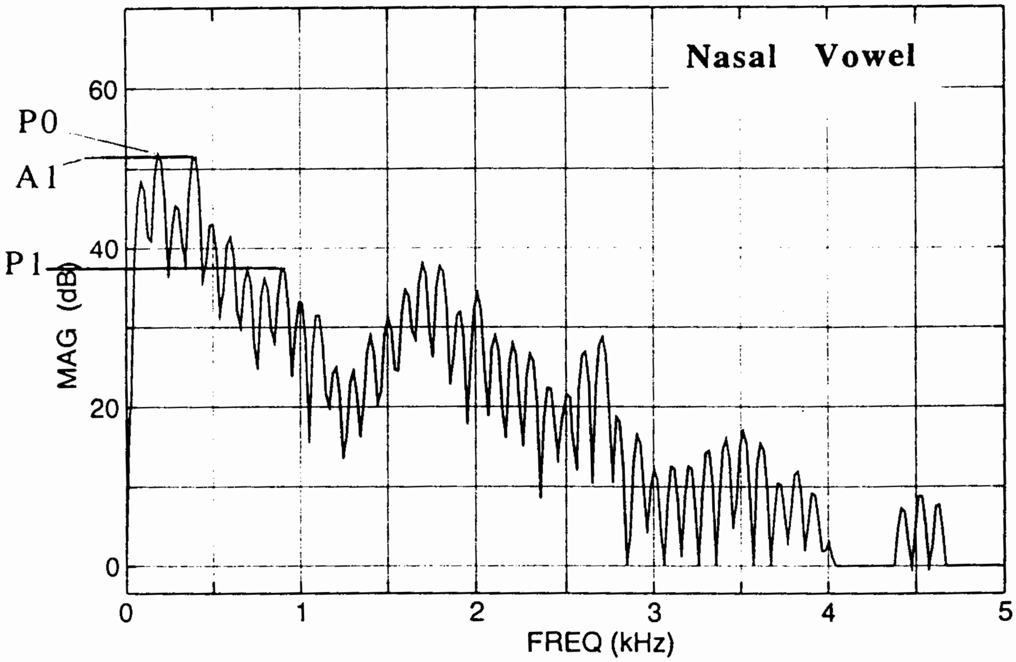
3.3.1 Method

To test the accuracy of A1-P1 and A1-P0 as measures of nasalization, these correlates were measured and compared for minimal pairs of nasalized vowels and their corresponding non-nasal vowels. Eight normal native-English speakers, four males (CM, MJ, NM, PG) and four females (JM, KT, LW, SM) participated in this study. They said monosyllable words with the vowels /i/, /ɛ/, /æ/, /u/, /ɑ/, and /ʌ/ in the carrier phrase "think ___________ clearly." The vowel was either adjacent to nasal consonants, e.g., mean, men, man, moon, mom, and mum, or adjacent to stops, e.g., beeb, bed, bad, booed, bob, and bud, so that a clear distinction between nasal and non-nasal vowels was expected. Three repetitions of each token were recorded. The frequency of the first and second formants, F1 and F2, respectively, and the frequency of P1 and P0, F_{P1} and F_{P0}, respectively, were measured at the beginning, middle, and end of the vowels. The correlates A1-P1 and A1-P0 were also measured at the beginning, the middle, and the end of the nasalized and non-nasalized vowels.

Figure 3.2a shows the spectrum obtained from a non-nasal vowel and Fig. 3.2b shows the spectrum from a nasal vowel with A1, P1, and P0 labeled. P1 in the non-nasal vowel was measured by using the harmonic with the maximum amplitude in the region of the extra peak for the nasal vowel; P0 in the non-nasal vowel was measured using the harmonic with the maximum amplitude in the low frequency region. In the example, the values for the nasal vowel, A1-P1 = 15 dB and A1-P0 = -1 dB, are much less than those in the non-nasal vowel, A1-P1 = 27 dB and A1-P0 = 7 dB.



(a)



(b)

Figure 3.2: (a) The spectrum of a non-nasal vowel is compared with (b) the spectrum of a nasalized vowel. The spectral peaks that determine A1, P1, and P0 are labeled. The amplitudes are measured by using the harmonic amplitudes closest to the expected peak according to theory. A1-P1 and A1-P0 are greater for the non-nasal vowel than for the nasal vowel.

3.3.2 Results and Discussion

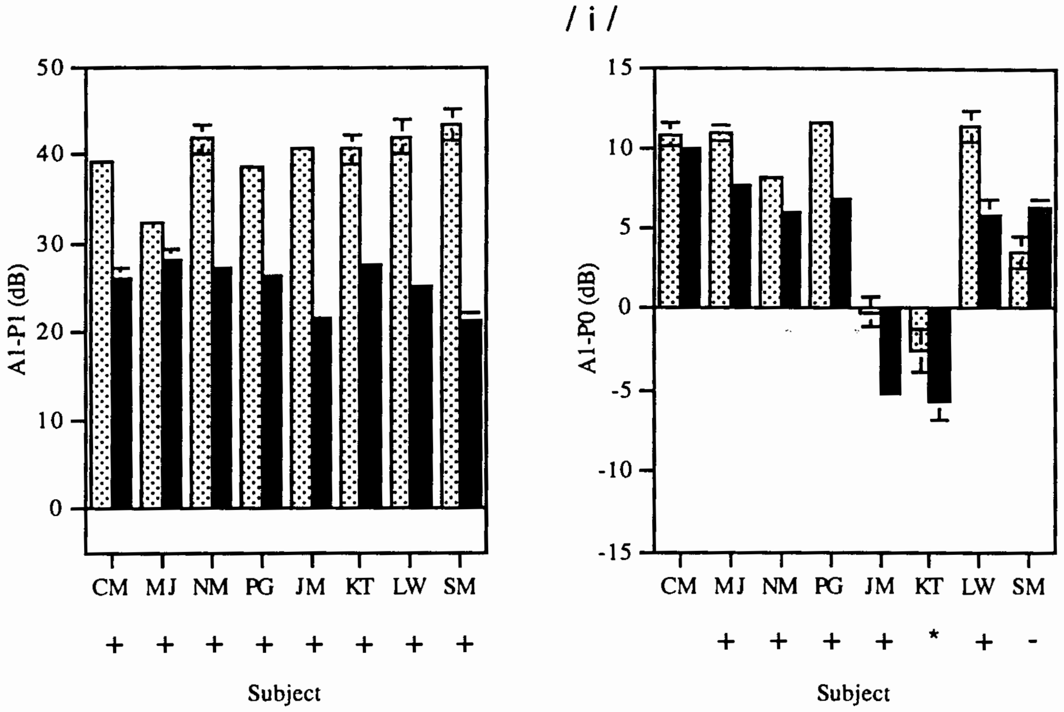
Table 3.1 shows the average frequencies measured from nasalized vowels. Since for /i/ F1 is at a low frequency close to P0, it was difficult to measure F_{P0}. For all of the vowels, except /i/, average F_{P0} ranged from 206 Hz to 223 Hz with an average across vowels of 216 Hz. For all of the vowels, average F_{P1} ranged from 924 Hz to 1032 Hz with an average across vowels of 966 Hz. Table 3.2 shows the average frequencies measured from non-nasalized vowels. The symbol '+' in Table 3.1 indicates $p < 0.01$ and '*' indicates

Table 3.1: Average frequency of the nasal peaks and the first two formants measured in English nasalized vowels of eight speakers and three repetitions of each. F_{P0} and F_{P1} are the frequencies of the nasal peaks with amplitudes P0 and P1, respectively. F1 and F2 are the first and second formants, respectively. The symbol '+' indicates $p < 0.01$ and '' indicates $p < 0.05$ between the nasal and non-nasal (average shown in Table 3.2) values.*

NASALIZED				
Vowel	F_{P0} (Hz)	F1 (Hz)	F_{P1}(Hz)	F2 (Hz)
/i/	-	273+	964+	2529+
/u/	206	397+	1032	1213+
/ɛ/	223	520*	982	1895
/ʌ/	223	627+	938	1194+
/æ/	212	562+	924	1903+
/ɑ/	217*	655	953	1178+

Table 3.2: Average frequency of the "nasal" peaks and the first two formants measured in English non-nasalized vowels of eight speakers and three repetitions of each. F_{P0} and F_{P1} are the frequencies of the corresponding "nasal" peaks with amplitudes P0 and P1, respectively. F1 and F2 are the first and second formants, respectively. The frequencies of the "nasal" peaks are as defined in the text.

NON- NASALIZED				
Vowel	F_{P0} (Hz)	F1 (Hz)	F_{P1}(Hz)	F2 (Hz)
/i/	-	300	992	2404
/u/	207	371	1046	1272
/ɛ/	222	496	986	1879
/ʌ/	224	558	935	1463
/æ/	215	597	930	1830
/ɑ/	210	650	958	1208



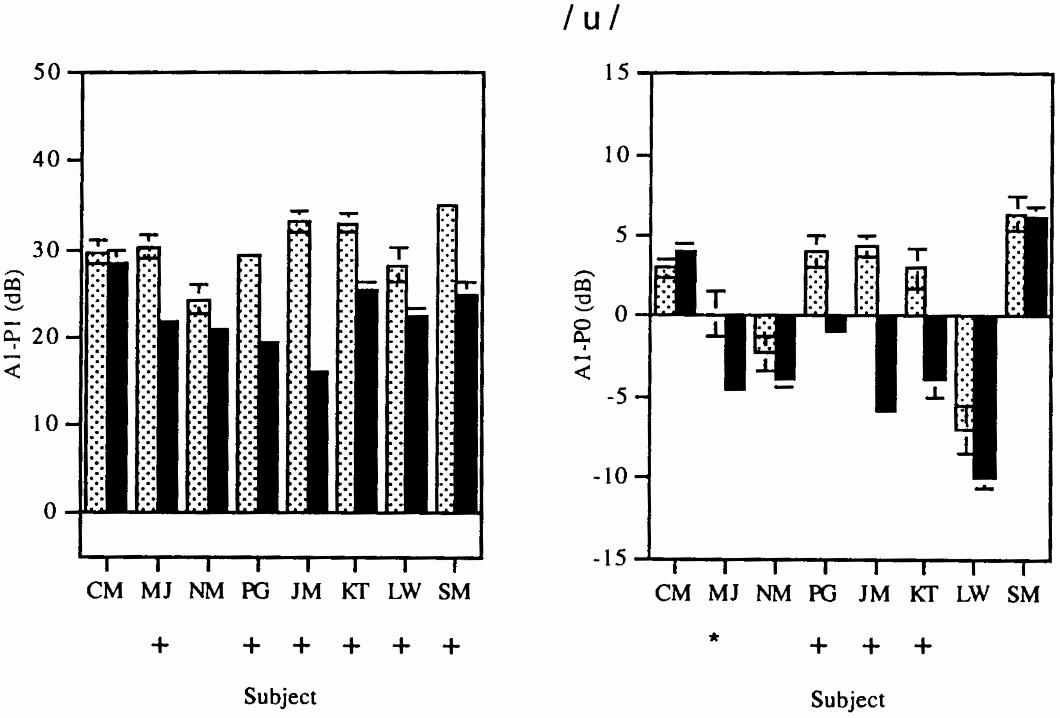
(a) (b)

Figure 3.3: (a) The correlate A1-P1 was measured at the beginning, the middle, and the end of the non-nasal /i/ (dotted bars) and nasal /i/ (solid bars). (b) The correlate A1-P0 was also measured at the beginning, the middle, and the end of the non-nasal /i/ (dotted bars) and nasal /i/ (solid bars). The symbol '+' along the abscissa indicates $p < 0.01$ and '' indicates $p < 0.05$ between the non-nasal and nasal values for the given speaker. The symbol '-' shows that the value of the non-nasalized vowels is actually greater than that of the nasalized vowels.*

$p < 0.05$ from a one-sided T-test between the frequencies of the nasalized and non-nasalized vowels. Comparisons of F_{P1} and F_{P0} between nasalized vowels and non-nasalized vowels mostly were not statistically significant. Comparisons of F1 and F2 showed statistically significant differences between nasalized and non-nasalized vowels, however for some vowels the formants are higher for the nasalized vowels and in other vowels, the formants are lower for the nasalized vowels.

The measurements of A1-P1 and A1-P0 made at the various points within the vowel showed no significant difference, so that only their averages were examined for eight speakers with three repetitions each. The first harmonic amplitude was used for P0 in females and the second harmonic amplitude was used for P0 in males because of the lower male fundamental frequency. Figures 3.3a-b to Figs. 3.8a-b show the average A1-P1 and A1-P0 measured in the vowels /i/, /u/, /ɛ/, /ʌ/, /æ/, and /ɑ/, respectively, for each of the speakers. The dotted bars show values for the non-nasal vowels and the solid bars show values for the nasal vowels. The symbol '+' along the abscissa indicates $p < 0.01$ and '*' indicates $p < 0.05$ from a one-sided T-test between the non-nasal and nasal values for the given vowel and speaker. The error bars are also indicated. In some cases they are too small to be seen on the graph.

For /i/ and /u/, as shown in Fig. 3.3a and Fig. 3.4a, respectively, most of the speakers showed statistically significant differences between A1-P1 measured in the non-nasal vowels, (A1-P1)$_{\text{-nasal}}$, and that of the nasal vowels, (A1-P1)$_{\text{nasal}}$, while fewer speakers showed statistically significant differences for A1-P0, as shown in Fig. 3.3b and Fig. 3.4b. For /i/, due to the low first formant



(a) (b)

Figure 3.4: (a) The correlate A1-P1 was measured at the beginning, the middle, and the end of the non-nasal /u/ (dotted bars) and nasal /u/ (solid bars). (b) The correlate A1-P0 was also measured at the beginning, the middle, and the end of the non-nasal /u/ (dotted bars) and nasal /u/ (solid bars). The symbol '+' along the abscissa indicates $p < 0.01$ *and '*' indicates* $p < 0.05$ *between the non-nasal and nasal values for the given speaker.*

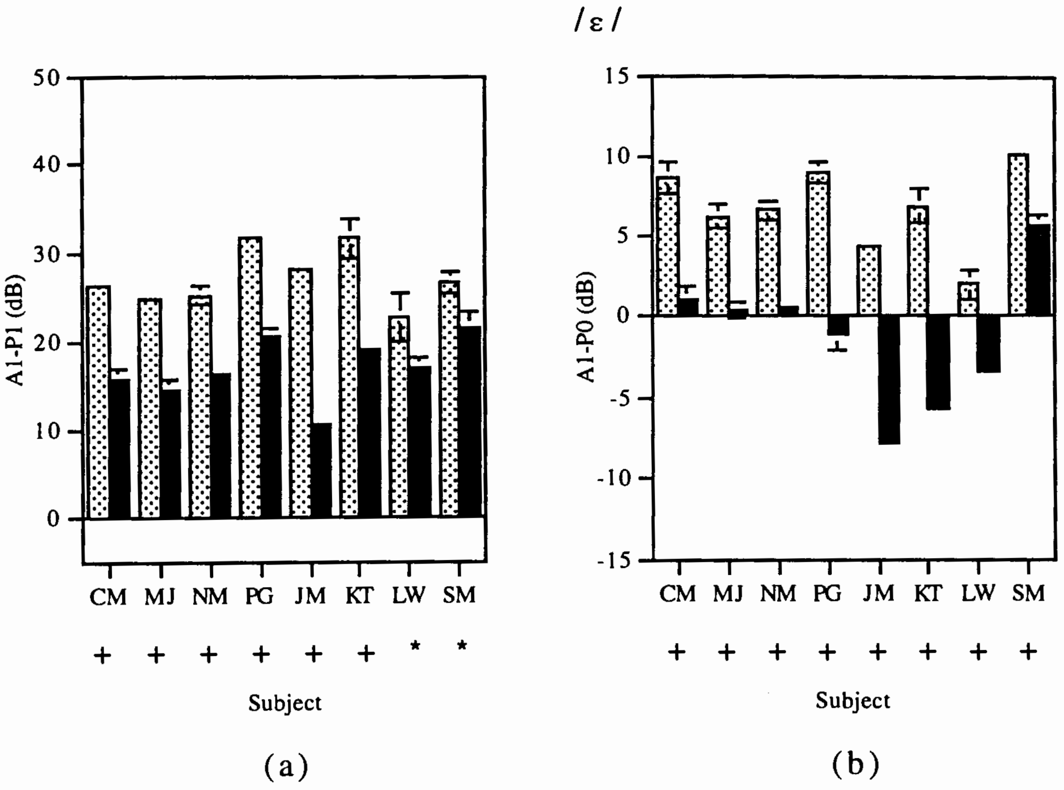
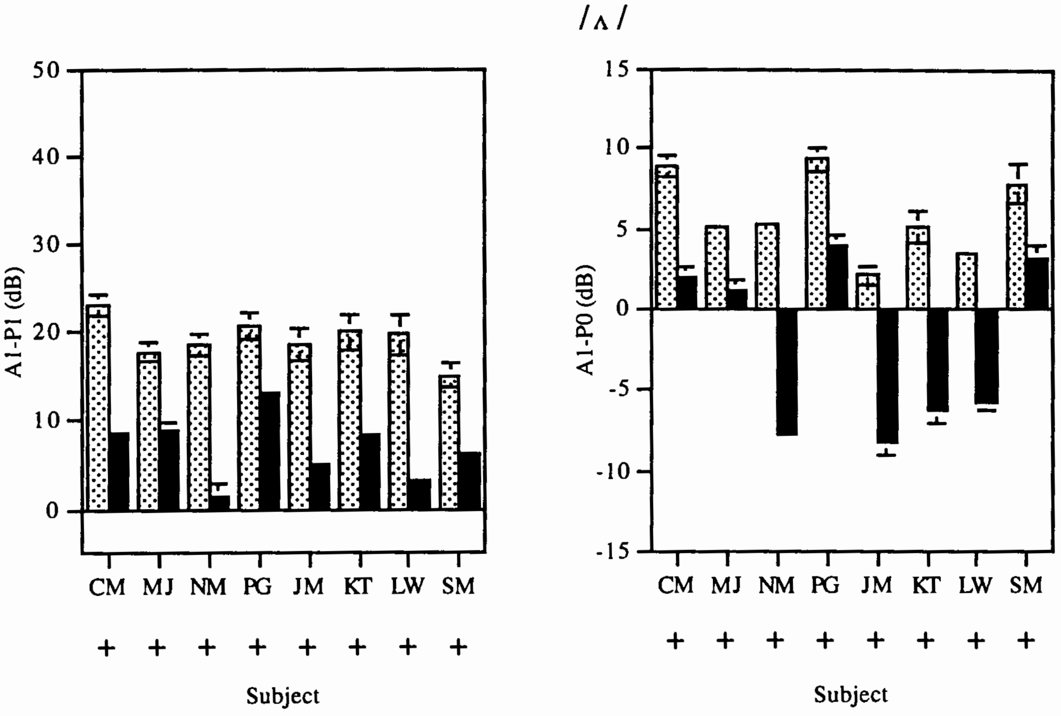


Figure 3.5: (a) The correlate A1-P1 was measured at the beginning, the middle, and the end of the non-nasal /ε/ (dotted bars) and nasal /ε/ (solid bars). (b) The correlate A1-P0 was also measured at the beginning, the middle, and the end of the non-nasal /ε/ (dotted bars) and nasal /ε/ (solid bars). The symbol '+' along the abscissa indicates $p < 0.01$ and '' indicates $p < 0.05$ between the non-nasal and nasal values for the given speaker.*

F1 as shown in Table 3.1, only the first harmonic could be used to measure P0 even though the extra peak may coincide with the first formant. The vowel /u/ also has a relatively low F1, which may influence the measurement of A1-P0, making A1-P0 a less robust measure of nasalization than A1-P1.

For /ε/ and /ʌ/, as shown in Fig. 3.5a and Fig. 3.6a, respectively, there was statistically significant difference between $(A1\text{-}P1)_{\text{-nasal}}$ and $(A1\text{-}P1)_{\text{nasal}}$ for all of the speakers. In some cases, the second formant F2 in /ʌ/ was fairly close to the frequency of P1. As a result, the nasal peak and the second formant influence one another making the measurement of A1-P1 less reliable. In the non-nasalized vowel, however, A1-P1 was more robust toward the end of the vowel preceding /d/ where F2 shifts up in frequency as shown by the higher F2 average in Table 3.2. Also, for /ε/ and /ʌ/, there was statistically significant difference between $(A1\text{-}P0)_{\text{-nasal}}$



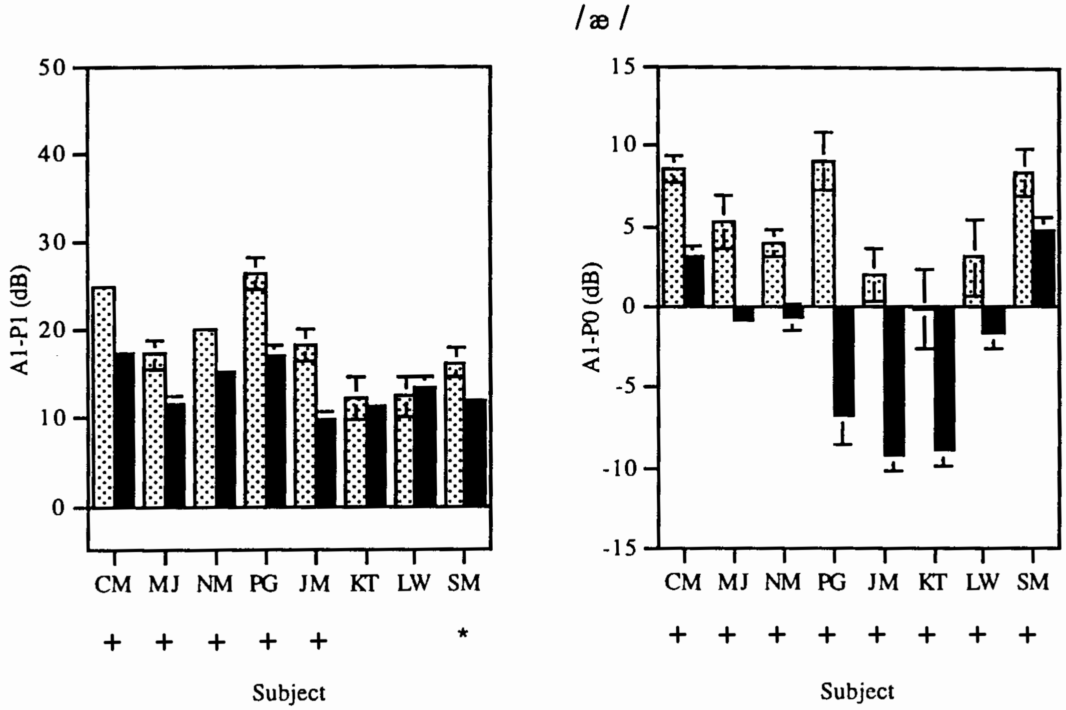
(a) (b)

Figure 3.6: (a) The correlate A1-P1 was measured at the beginning, the middle, and the end of the non-nasal /ʌ/ (dotted bars) and nasal /ʌ/ (solid bars). (b) The correlate A1-P0 was also measured at the beginning, the middle, and the end of the non-nasal /ʌ/ (dotted bars) and nasal /ʌ/ (solid bars). The symbol '+' along the abscissa indicates $p < 0.01$ and '' indicates $p < 0.05$ between the non-nasal and nasal values for the given speaker.*

and $(A1\text{-}P0)_{nasal}$ for all of the speakers, as shown in Fig. 3.5b and 3.6b, respectively, since F1 for these two vowels occur at a high frequency, farther away from F_{P0}.

For /æ/ and /ɑ/, as shown in Fig. 3.7b and Fig. 3.8b, respectively, all of the speakers showed a statistically significant difference between $(A1\text{-}P0)_{\text{-}nasal}$ and $(A1\text{-}P0)_{nasal}$ while fewer speakers showed a statistically significant difference for A1-P1, as indicated by Fig. 3.7a and Fig. 3.8a. As shown in Table 3.1, F1 was at a much higher frequency than F_{P0} so that A1-P0 was more reliable. However, F1 occurring at a high frequency in the low vowels would influence the measurement of A1-P1. In addition, F2 in /ɑ/ occurred at a low frequency close to F_{P1} making the measurement of A1-P1 less robust.

In summary, spectral analysis of /i/, /ɛ/, /æ/, /u/, /ɑ/, and /ʌ/ at the beginning, middle, and end of the vowels was performed for

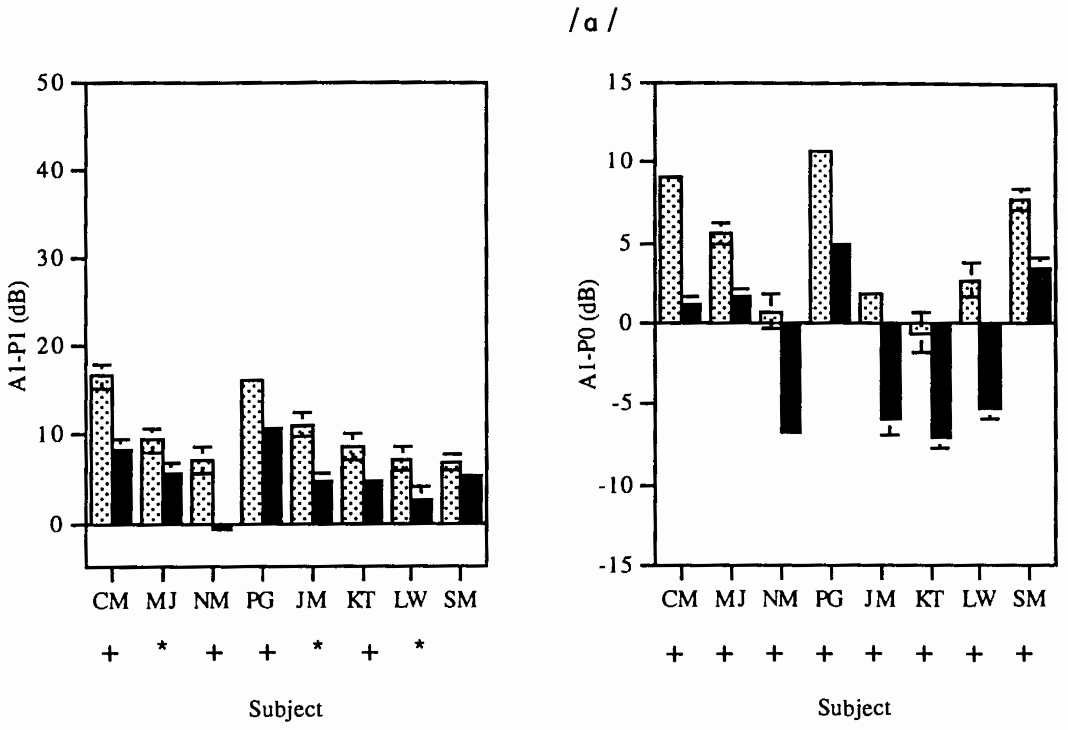


(a) (b)

Figure 3.7: (a) The correlate A1-P1 was measured at the beginning, the middle, and the end of the non-nasal /æ/ (dotted bars) and nasal /æ/ (solid bars). (b) The correlate A1-P0 was also measured at the beginning, the middle, and the end of the non-nasal /æ/ (dotted bars) and nasal /æ/ (solid bars). The symbol '+' along the abscissa indicates $p < 0.01$ and '' indicates $p < 0.05$ between the non-nasal and nasal values for the given speaker.*

eight English speakers with three repetitions each. A more statistically significant difference of A1-P1 than A1-P0 between non-nasalized vowels and nasalized vowels was found for the high vowels /i/ and /u/ since F1 at a low frequency would influence the measurement of P0. On the other hand, a more statistically significant difference of A1-P0 than A1-P1 was found for the low vowels /æ/ and /ɑ/ since F1 at a high frequency would influence the measurement of P1. For /ɛ/ and /ʌ/, both A1-P1 and A1-P0 can be used to distinguish non-nasalized and nasalized vowels.

The difference Δ(A1-P1) between the average of $(A1\text{-}P1)_{-nasal}$ and the average of $(A1\text{-}P1)_{nasal}$ among speakers was calculated for the non-low vowels. The same was done for the difference Δ(A1-P0) between the average of $(A1\text{-}P0)_{-nasal}$ and the average of $(A1\text{-}P0)_{nasal}$ for the non-high vowels. Table 3.3 shows the results from the averages of eight subjects. The mean for Δ(A1-P1) ranged from



(a) (b)

Figure 3.8: (a) The correlate A1-P1 was measured at the beginning, the middle, and the end of the non-nasal /ɑ/ (dotted bars) and nasal /ɑ / (solid bars). (b) The correlate A1-P0 was also measured for the non-nasal /ɑ / (dotted bars) and nasal /ɑ/ (solid bars). The symbol '+' indicates $p < 0.01$ and '' means $p < 0.05$ between the non-nasal and nasal values for the given speaker.*

Table 3.3 Mean, minimum (min.), and maximum (max.) differences between the average A1-P1 of non-low vowels and A1-P0 of the non-high vowels measured in non-nasal and nasal English vowels for eight subjects.

	Δ(A1-P1) (dB)			Δ(A1-P0) (dB)		
Vowel	mean	min.	max.	mean	min.	max.
/i/	15	5	23	-	-	-
/u/	11	7	17	-	-	-
/ɛ/	10	4	17	8	5	13
/ʌ/	12	8	17	8	4	13
/æ/	-	-	-	8	4	16
/ɑ/	-	-	-	6	4	8

10 dB to 15 dB with the minimum ranging from 4 dB to 8 dB and the maximum from 17 dB to 23 dB. /i/ in general had higher Δ(A1-P1) than the other vowels. Although A1 is expected to decrease less for /i/ with its lower F1 and smaller A_{VP}, P1 is expected to increase more than that of the other vowels even with a smaller A_{VP}, as shown in Table 2.1. Theoretically, it is possible for A1-P1 to decrease by 23 dB when losses and coupling to the main nasal passages are considered. The mean for Δ(A1-P0) ranged from 6 dB to 8 dB with the minimum ranging from 4 dB to 5 dB and the maximum from 8 dB to 16 dB. The maximum value in Table 3.3 corresponds to the theoretical calculations. Theoretically, it is possible for A1-P0 to decrease by 13 dB due to losses and coupling to sinuses. If the open quotient were manipulated, A1-P0 may decrease further by 5 dB. The theoretical values are generally larger than the changes in A1-P1 and A1-P0 in nasalized vowels from non-nasalized vowels since the theory assumes a large velopharyngeal coupling during the nasalized vowels. This may not occur when the vowel is nasalized due to coarticulation. Also, the open quotient of the glottal waveform may not be increased to raise P0. Furthermore, the peak amplitudes can only be approximated by using the amplitude of the closest harmonic, which varies with the fundamental frequency of the vowel.

3.4 French Speakers: Nasalized and Non-nasalized Portions of Nasal Vowels

3.4.1 Method

In French, which has nasal-non-nasal distinction, nasal vowels do not have corresponding oral vowel counterparts in which non-nasalized and nasalized vowels have similar speech articulator positions except for the velum and the lateral pharyngeal walls, as in English. In order to obtain non-nasalized and nasalized vowel counterparts, eight normal native-French speakers - three males (DH, DN, JS) and five females (AG, CL, CT, MT, VV) - said words with a stop followed by one of the four French nasal vowels /ɛ̃/, /ɑ̃/, /ɔ̃/, and /œ̃/, e.g. bain, ban, bon, and quelqu'un, in the carrier phrase "dites ______ pour moi." Three repetitions were recorded for each token. The beginning of the vowel should be produced with the v-p port closed for pressure build-up in the stop while later in the vowel the velum should be lowered and the pharyngeal walls relaxed to nasalize the vowel with minimal change in the position of the other speech articulators. Although some languages have oral vowel counterparts to nasal vowels, we chose to study the French vowels, since most of the past studies on languages with nasal-non-nasal distinction were done on French nasal vowels.

The frequency of the first and second formant, F1 and F2, respectively, and the frequency of the extra peaks P1 and P0, F_{P1} and F_{P0}, respectively, were measured at the beginning and end of the vowel. The maximum deviations of A1-P1 and A1-P0 from the beginning of the vowel to 20 ms into the vowel, to the middle of the vowel, and to the end of the vowel were measured for each speaker and repetition.

3.4.2 Results and Discussion

Table 3.4 shows the average frequencies measured from the nasalized portion of the vowel and Table 3.5 shows the average

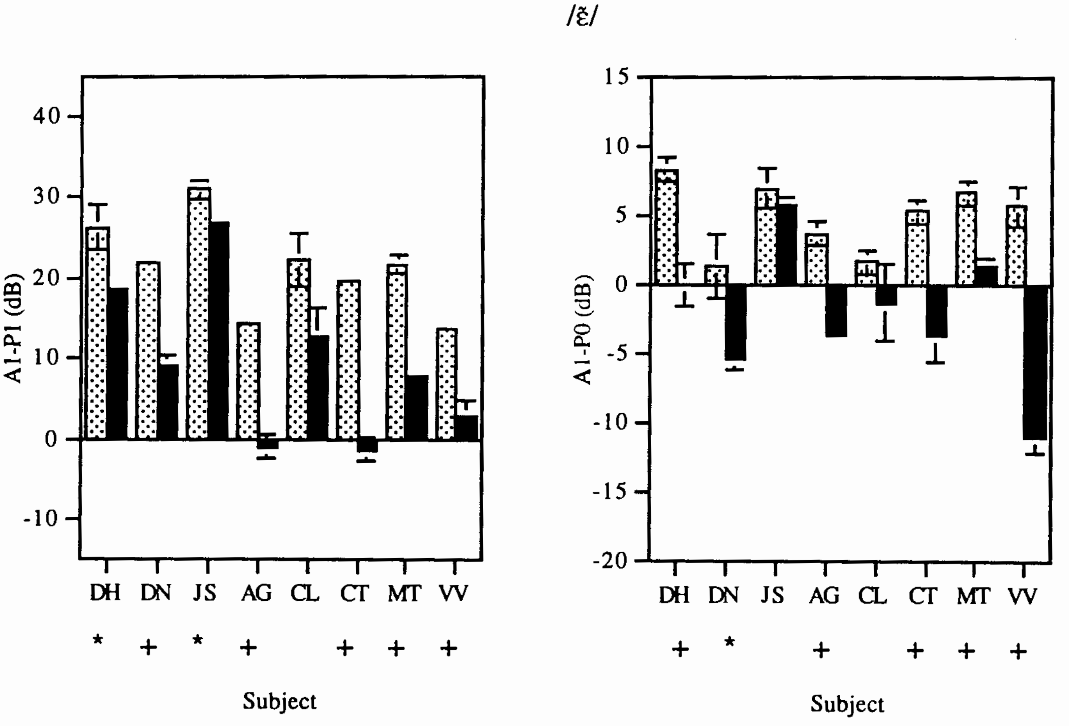
Table 3.4: Average frequency of the nasal peaks and the first two formants measured at the end of French nasalized vowels of eight speakers with three repetitions each. F_{P0} and F_{P1} are the frequency of the nasal peaks with amplitude P0 and P1, respectively. F1 and F2 are the first and second formants, respectively.

Nasalized (end)				
Vowel	F_{P0} (Hz)	F1 (Hz)	F_{P1}(Hz)	F2 (Hz)
/ɛ̃/	236*	536*	958	1302+
/ɑ̃/	256+	555+	883	1019
/ɔ̃/	238+	475+	1029	828
/œ̃/	216+	481	874+	1277+

Table 3.5: Average frequency of the nasal peaks and the first two formants measured at the beginning, non-nasalized portion, of French nasalized vowels of eight speakers with three repetitions each. F_{P0} and F_{P1} are the frequency of the nasal peaks with amplitude P0 and P1, respectively. F1 and F2 are the first and second formants, respectively.

Non-nasalized (beginning)				
Vowel	F_{P0} (Hz)	F1 (Hz)	F_{P1}(Hz)	F2 (Hz)
/ɛ̃/	224	485	950	1417
/ɑ̃/	238	460	874	1049
/ɔ̃/	222	435	1042	841
/œ̃/	242	507	930	1732

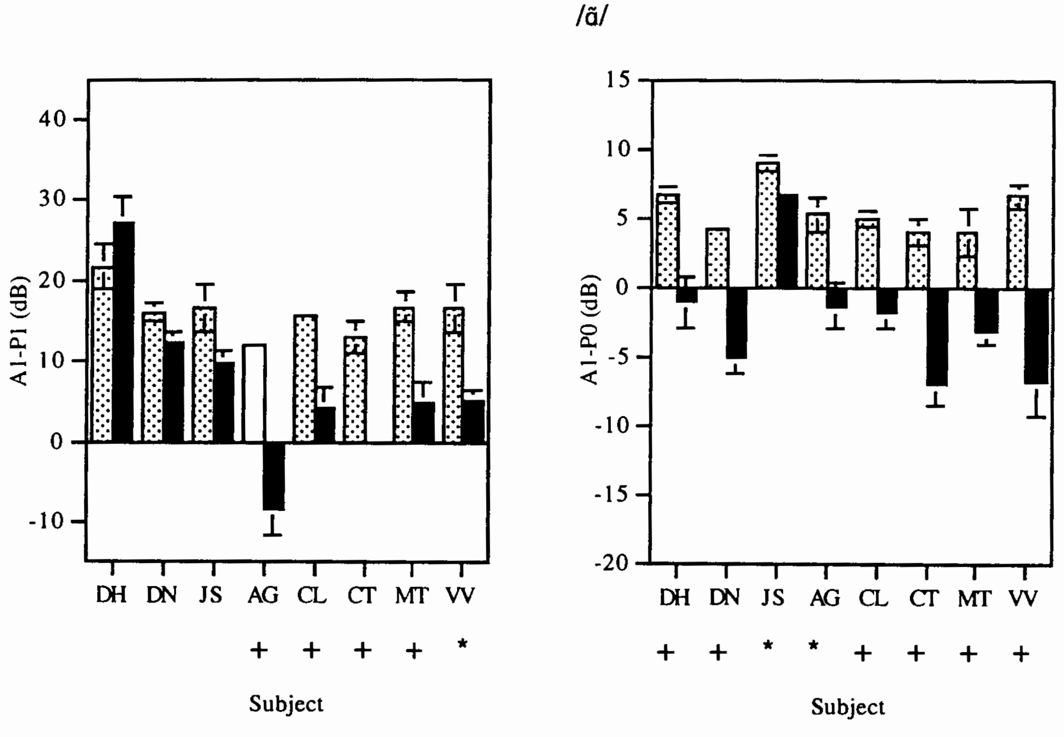
frequencies measured from the non-nasalized portion of the vowel. For all of the nasalized vowels, average Fp_0 ranged from 216 Hz to 256 Hz with an average across vowels of 237 Hz, and average Fp_1 ranged from 874 Hz to 1029 Hz with an average across vowels of 936 Hz. The frequency of the first formant indicates that the nasal vowels are non-low and non-high vowels. Comparison between the nasalized and non-nasalized portions of the vowel showed that there is a statistically significant difference for Fp_0 in all of the vowels. This difference was due to changes in the fundamental frequency rather than the use of different harmonic amplitudes for P0. For Fp_1 only one vowel showed statistically significant difference which may also be due to changes in the fundamental frequency. With nasalization, there was a significant lowering of F2 as two of the vowels were nasalized but there was a significant raising of F1 as three of the vowels were nasalized. These changes in formant frequencies were influenced by nasalization as well as mouth closure for making the beginning stop.



(a) (b)

Figure 3.9: (a) The correlate A1-P1 was measured at the beginning of /ɛ̃/ (dotted bars) and the end of /ɛ̃/ (solid bars). (b) The correlate A1-P0 was also measured at the beginning of /ɛ̃/ (dotted bars) and the end of /ɛ̃/ (solid bars). The symbol '+' along the abscissa indicates $p < 0.01$ and '' indicates $p < 0.05$ between the non-nasal and nasal values for the given speaker.*

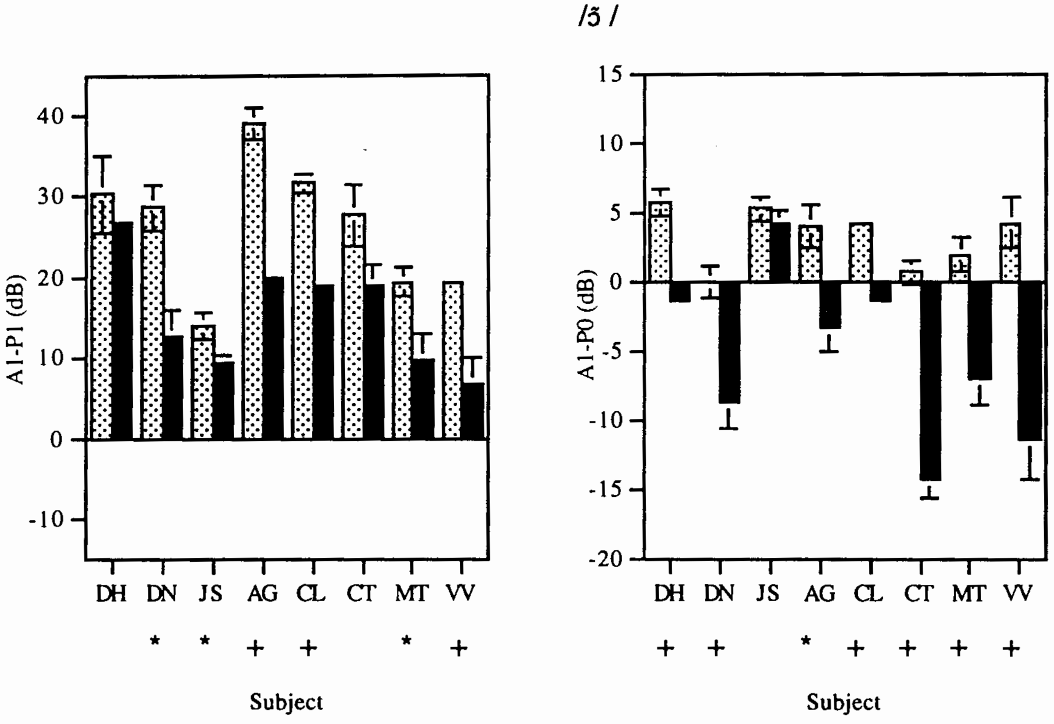
In general, the correlates A1-P1 and A1-P0 measured at the end of the vowel showed greatest deviation from those at the beginning than the values measured at the other two locations in the vowel. According to the definition of P0, the second harmonic amplitude was used for measuring P0 for two male speakers while the first harmonic amplitude was used for measuring P0 for female speakers and one male speaker. Figures 3.9a-b through Figs. 3.12a-b show the average A1-P1 and A1-P0 measured in the nasal vowels, /ɛ̃/, /ɑ̃/, /ɔ̃/, and /œ̃/, respectively, for each of the eight speakers. The dotted bars show values for the non-nasalized portion of the vowel following the stop and the solid bars show values for the maximally nasalized portion at the end of the vowel. The symbol '+' along the abscissa indicates $p < 0.01$ and '*' indicates $p < 0.05$ between the non-nasal and nasal values for the given vowel and speaker. Since the sample size is small, a technique suggested by Sachs (1984) using the arithmetic means and the arithmetic mean of the ranges was used for the one-sided T-test.



(a) (b)

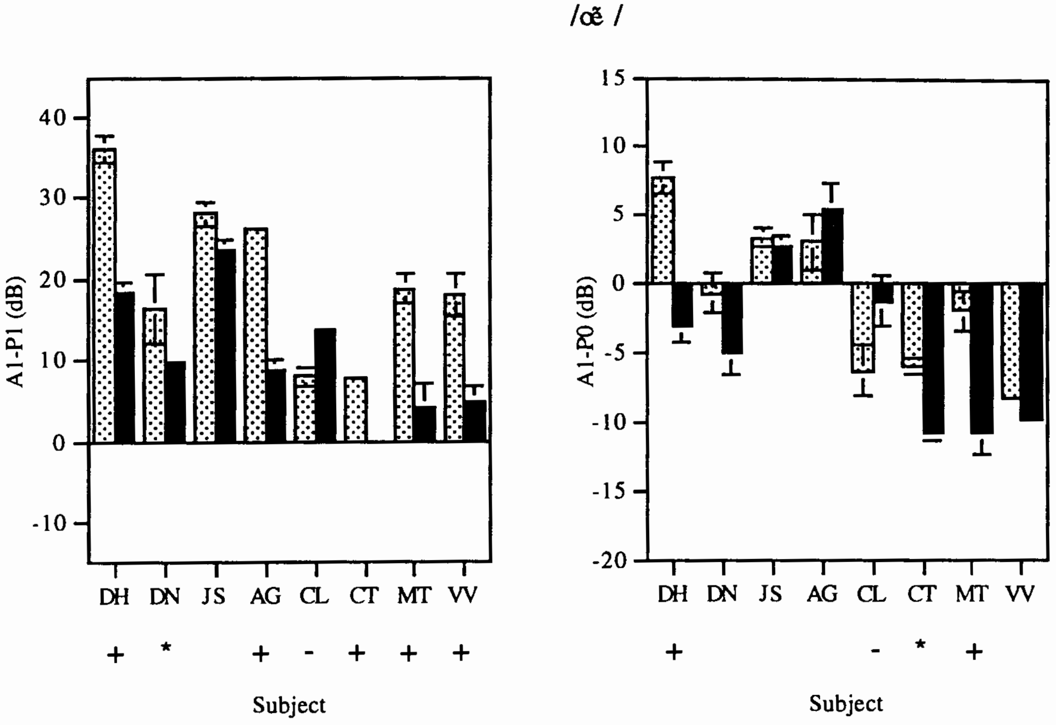
Figure 3.10: (a) The correlate A1-P1 was measured at the beginning of /ɑ̃/ (dotted bars) and the end of /ɑ̃/ (solid bars). (b) The correlate A1-P0 was also measured at the beginning of /ɑ̃/ (dotted bars) and the end of /ɑ̃/ (solid bars). The symbol '+' along the abscissa indicates p < 0.01 and '' indicates p < 0.05 between the non-nasal and nasal values for the given speaker.*

For /ɛ̃/, as shown in Figs. 3.9a, there was a statistically significant difference between A1-P1 measured in the non-nasal portion of the vowel (A1-P1)$_{\text{-nasal}}$ and that of the nasal portion (A1-P1)$_{\text{nasal}}$. There was also a statistically significant difference between (A1-P0)$_{\text{-nasal}}$ and (A1-P0)$_{\text{nasal}}$, as shown in Fig. 3.9b. For all speakers except CL, the measurements were larger at the beginning of the vowel with minimal nasalization than into the vowel where nasalization occurred, in agreement with A1-P1 and A1-P0 being inversely related to nasalization. For /ɑ̃/, and /ɔ̃/, most of the speakers showed statistically significant differences between (A1-P0)$_{\text{-nasal}}$ and (A1-P0)$_{\text{nasal}}$, as indicated by Fig. 3.10b and Fig. 3.11b, while for A1-P1, the differences were less consistently significant, as shown in Fig. 3.10a and Fig. 3.11a. Table 3.4 indicates that F2 was very close to F_{p1} for /ɑ̃/, and /ɔ̃/, making A1-P1 more difficult to measure. For /œ̃/, most of the speakers showed statistically



(a) (b)

Figure 3.11: (a) The correlate A1-P1 was measured at the beginning of /ɔ̃/ (dotted bars) and the end of /ɔ̃/ (solid bars). (b) The correlate A1-P0 was also measured at the beginning of /ɔ̃/ (dotted bars) and the end of /ɔ̃/ (solid bars). The symbol '+' along the abscissa indicates $p < 0.01$ and '' indicates $p < 0.05$ between the non-nasal and nasal values for the given speaker.*



(a) (b)

Figure 3.12: (a) The correlate A1-P1 was measured at the beginning of /œ̃/ (dotted bars) and the end of /œ̃/ (solid bars). (b) The correlate A1-P0 was also measured at the beginning of /œ̃/ (dotted bars) and the end of /œ̃/ (solid bars). The symbol '+' along the abscissa indicates $p < 0.01$ and '' indicates $p < 0.05$ between the non-nasal and nasal values for the given speaker. The symbol '-' shows that the value of the non-nasalized vowels is actually smaller than that of the nasalized vowels.*

significant differences between (A1-P1)$_{\text{-nasal}}$ and (A1-P1)$_{\text{nasal}}$, as shown in Fig. 3.12a, while for A1-P0, the differences were less consistently significant, as shown in Fig. 3.12b. For speaker CL, A1-P1 and A1-P0 were greater for the nasal portion of the vowel than the non-nasal portion.

The difference between (A1-P1)$_{\text{-nasal}}$ and (A1-P1)$_{\text{nasal}}$, Δ(A1-P1) and the difference between (A1-P0)$_{\text{-nasal}}$ and (A1-P0)$_{\text{nasal}}$, Δ(A1-P0), were calculated. Table 3.6 shows the results from the averages of eight subjects. The mean for Δ(A1-P1) ranged from 9 dB to 12 dB with the minimum ranging from -6 dB to 4 dB and the maximum from 18 dB to 21 dB, depending on the vowel type. Theoretically, A1-P1 can decrease as much as 23 dB when losses due to coupling to the nasal passages are considered which corresponds to the maximum Δ(A1-P1) that was measured. The mean for Δ(A1-

P0) ranged from 3 dB to 9 dB with the minimum ranging from -5 dB to 2 dB and the maximum from 11 dB to 17 dB, depending on the vowel type. The low minimum range was mainly due to speaker JS and also, for /œ̃/, F2 shifting to a lower frequency into the vowel may have decreased Δ(A1-P0). Theoretically, it is possible for A1-P0 to decrease by as much as 13 dB due to losses and coupling to sinuses with a further decrease of 5 dB if the open quotient is increased. As for the minimal pairs in English, the maximum differences in the nasalized portion and the non-nasalized portion of the vowels are below the theoretical values since the theory assumes a large velopharyngeal coupling during the nasalized vowels which may not always occur at the end of the nasal vowel. The peak amplitudes can only be approximated by using the amplitude of the closest harmonic, which varies with the fundamental frequency of the vowel.

Table 3.6: Mean, minimum (min.), and maximum (max.) differences between the average A1-P1 measured in non-nasal and nasal portions of French vowels for eight subjects are shown. The differences for A1-P0 are also listed.

	Δ(A1-P1) (dB)			Δ(A1-P0) (dB)		
Vowel	mean	min.	max.	mean	min.	max.
/ɛ̃/	12	4	21	7	1	17
/ɑ̃/	9	-6	20	8	2	13
/ɔ̃/	11	3	19	9	1	16
/œ̃/	9	-6	18	3	-5	11

3.5 Summary

Spectral analysis was carried out for English vowels with minimal pairs of nasalized and non-nasalized vowels and for the nasalized and non-nasalized portions of the French nasal vowels. According to the theory of speech production, coupling to the nasal cavity introduces losses, which lower the amplitude of the first formant, A1 (dB). The side branch also introduces above the first formant a nasal pole-zero pair which raises the amplitude of the extra peak P1 (dB). The average frequency measured across speakers and vowels for this extra peak was 966 Hz for English speakers and 936 Hz for French speakers. In addition, coupling to the paranasal sinuses, which introduces a pole-zero pair at low frequencies and/or a large open quotient of the glottal waveform, would raise the amplitude of the extra peak P0 (dB). In general, the amplitude of the first harmonic was used for P0 in females and the second harmonic amplitude was used for males. The average frequency measured across speakers and vowels for this extra peak was 216 Hz for English speakers and 237 Hz for French speakers.

The acoustic correlate A1-P1 was used to quantify the spectral characteristic in the non-low vowels while A1-P0 was used to quantify the non-high vowels. Usually, the values from the non-nasal vowels $(A1\text{-}P1)_{\text{-nasal}}$ and $(A1\text{-}P0)_{\text{-nasal}}$ were statistically larger than the corresponding values from nasal vowels $(A1\text{-}P1)_{\text{nasal}}$ and $(A1\text{-}P0)_{\text{nasal}}$, showing the fact that A1-P1 and A1-P0 are inversely related to nasalization. Across English speakers, the difference between the means of $(A1\text{-}P1)_{\text{-nasal}}$ and $(A1\text{-}P1)_{\text{nasal}}$ ranged from 10 dB to 15 dB while the difference between the means of $(A1\text{-}P0)_{\text{-nasal}}$ and $(A1\text{-}P0)_{\text{nasal}}$ ranged from 6 dB to 8 dB, depending on vowel types. Across French speakers, the difference between the means of $(A1\text{-}P1)_{\text{-nasal}}$ and $(A1\text{-}P1)_{\text{nasal}}$ ranged from 9 dB to 12 dB while the difference between the means of $(A1\text{-}P0)_{\text{-nasal}}$ and $(A1\text{-}P0)_{\text{nasal}}$ ranged from 3 dB to 9 dB. The maximum differences between the non-nasal vowels and nasal vowels could be explained by theoretical calculations. Since the nasal and non-nasal values were measured in minimal pairs in English but in different portions of nasal vowels in French, it was not feasible to make rigorous direct comparisons between the two languages.

Chapter 4

Perceptual Experiments on Synthetic Stimuli

By establishing the relationships between the acoustic correlates A1-P1 and A1-P0 of nasalization to the perceived nasality, these measures may be used directly to indicate the degree of nasality. Synthesis was used to determine the perceptual effects of systematically manipulating these acoustic parameters.

4.1 Effects of Manipulating A1-P1

A1-P1 as a measure of nasality was tested in which a synthesis technique was used to systematically manipulate specific acoustic parameters relevant to nasalization-- first formant bandwidth and the frequency of the nasal zero. The synthetic stimuli were then presented to listeners in a nasality perception experiment. The corpus consisted of 10 vowels in the context of bVt, synthesized based on utterances of an adult male speaker.

4.1.1 Stimuli

The non-nasalized vowels in the form bVt were synthesized by mimicking "beat", "bit", "bait", "bet", "bot", "but", "boot", "bat", "bought", and "boat" spoken by an adult male speaker. Synthesis was carried out by using the Klatt formant synthesizer, KLSYN88. Spectral matching at 10-ms intervals was used to obtain the fundamental frequency contour and the first five formant frequencies and bandwidths, with acoustic theory of speech production used as a guide. The duration of each word was 250 ms to 270 ms. The initial burst was simulated with 10 ms of frication noise at a 60 dB amplitude of frication (AF). The amplitude of voicing (AV) was fixed at 60 dB or 62 dB throughout the vowel with a rise time of 20 ms at the beginning and a fall time of 20 ms at the end of the word. The speaker did not produce the final /t/ release for all of the vowels so /t/ was not synthesized. For the copy-synthesis of the original recording, all other parameters were kept at the default values (Klatt and Klatt, 1990). Table 4.1 shows the average fundamental frequency, f0, the average frequency of the first formant, F1, the average frequency of the second formant, F2, and the bandwidth of the first formant, B1, for the non-nasal reference which is the synthesized version of the original.

Table 4.1: Synthesis parameter values for the average fundamental frequency (f0), the average first formant frequency (F1), the average second formant frequency (F2) and bandwidth of the first formant (B1) of the original speech.

Word	avg. f0 (Hz)	avg. F1 (Hz)	avg. F2 (Hz)	B1 (Hz)
beat	94	295	2400	60
bit	99	395	1900	60
bait	94	385	2300	60
bet	90	700	1800	150
bot	88	800	1260	150
but	92	700	1400	120
boot	95	350	1300	60
bat	90	800	1700	150
bought	91	550	980	120
boat	101	480	875	60

Theory and past synthesis studies of vowel nasalization provided some indication of the ranges of parameters to use in the synthesis of nasal vowels. Theoretically, with nasalization, the first formant is expected to be lowered by as much as 11 dB, so that the bandwidth would be more than 3 times that of the non-nasal reference. For some of the non-nasalized vowels, e.g. /ɛ/, /ɑ/, and /æ/, B1 was 150 Hz so that for nasalized vowels, B1 would have to increase to more than 450 Hz, which involves adding 300 Hz to B1 of the non-nasal vowel. Also, Chen's (1995) synthesis of vowels of two hearing-impaired children who were perceived to be nasal has shown that the first-formant bandwidth can be as large as 500 Hz at a given time within the vowel with a maximum average of 410 Hz. A pole-zero pair above the first formant was also necessary to match the spectra closely for these vowels. The average bandwidths of the extra pole and zero, BNP and BNZ, were each about 250 Hz. The zero was usually above the pole by as much as 300 Hz. In an informal synthesis study involving nasalized vowels for an adult speaker, this spacing was found to be as large as 400 Hz.

Starting with the non-nasal reference, to generate nasal stimuli, the bandwidth of the first formant was systematically varied and a nasal pole-zero pair above the first formant was introduced. From the theory and past synthesis experiments, the bandwidth of the first formant, B1, was increased by 100 Hz, 200 Hz, or 300 Hz. An extra pole was introduced at 950 Hz, which was around the average frequency at which the nasal peak occurred in English and French vowels (see Sec. 3.3 and 3.4). An extra zero was introduced at a frequency spacing of 0 Hz, 100 Hz, 200 Hz, 300 Hz, or 400 Hz above that of the extra pole. The distance in frequency between the pole and the zero was varied to manipulate the prominence of the extra peak. In order to synthesize the gradual introduction of vowel nasalization following the stop, B1 and the frequency of the extra pole and zero, FNP and FNZ, respectively, were increased linearly for the first 50 ms from the value of the non-nasal reference to the target value and then remained constant for the rest of the vowel. This time variation pattern was based on the study of the timing of the velar movement by Moll and Daniloff (1970). The bandwidths of the nasal pole and zero were both 250 Hz so that when they occurred at the same frequency, they canceled each other, leading to deletion of the extra peak. The values of B1 and FNZ were varied simultaneously so that for each vowel, 20 different variations were synthesized (4 values of bandwidth and 5 values of pole-zero spacing).

4.1.2 Procedure

Each item in the listening experiment was a pair of stimuli with the reference (synthesized version with no nasalization) followed in 0.4 s by one of its variations. Each pair was presented once with 2.5 s of silence between consecutive pairs. The intended bVt was also indicated to the listener in a written form. Four repetitions of all of the pairs were randomized and recorded onto TDK D60 cassettes using a Yamaha model 1000 cassette recorder and a custom software program for creating randomized tests of stimuli. The stimuli were presented with a mixture of vowel types through binaural headphones in a sound-treated room.

The listeners were tested individually. Each was asked to compare the second member of the pair with the first member (the reference) on a scale from -1 to 3 by writing the response (-1, 0, 1, 2, or 3) on a blank for the test item. If the listener judged the test stimulus to be less nasal than the reference, the response would be '-1.' A '0' response indicated that there was no detectable difference in nasality. Responses of '+1', '+2', '+3' indicated that the second stimulus sounded increasingly more nasalized than the first stimulus. An asymmetrical scale was used since it is expected that most of the responses would be positive and the task is easier with fewer choices. Realizing that using an asymmetrical scale may cause bias towards response of greater nasality, some precaution was made by instructing the subjects that they should not attempt to distribute their responses evenly. Before each test, instruction was given orally and in written form on the response sheet. To train the listeners for nasality perception before the test, they were presented words judged to be highly nasal and non-nasal by two trained phoneticians. They were able to distinguish the nasal and non-nasalized words correctly. To accustom the listeners to the task, the first ten pairs of the test were used solely for practice and not for scoring.

4.1.3 Listeners

Five male and three female right-handed native speakers of English participated in the experiment. They were undergraduates at MIT with ages under 25. The subjects had no known speech or hearing impairments. All were naive listeners without any past training in perception of nasality. It has been shown that there is no significant difference in the 50% crossover points of the nasality identification functions between speakers of languages with non-nasal-nasal vowel opposition, as in Gujarati, Hindi, and Bengali, and speakers of languages with no such distinction, as in English (Hawkins and Stevens, 1985). This suggests that there would be no significant difference in nasality perception between naive and non-naive listeners so that naive listeners may be used. Furthermore, most of the English listeners one encounters are naive listeners.

To determine how consistently each listener performed the task, the perceptual judgments for the four repetitions of each test token were analyzed. For a given token, the listener was able to choose from five categories (-1, 0, 1, 2, and 3). If the judgments were consistent, the listener would have used only one of the five categories for the four repetitions of the same token; the judgments would be the least consistent if the listener used four of the five categories. Table 4.2 shows, according to subjects, the frequency a given number of categories was used for making the judgments on

Table 4.2: Listener consistency in making judgments on stimuli with manipulations of A1-P1. The number of categories used for four repetitions of a given token is from one to four. The categories may be adjacent to each other or spread apart. The number of times a given category was chosen in presenting twenty versions of each of the ten vowels is listed.

	Number of categories for four repetitions					
Subject	One	Two adjacent	Two spread	Three adjacent	Three spread	Four
A	65	107	5	19	4	0
B	42	114	11	30	3	0
C	46	107	7	32	6	1
D	39	122	13	23	2	1
E	38	110	13	30	5	4
F	68	72	9	39	5	3
G	33	104	9	43	9	2
H	70	92	6	30	2	0

four repetitions of the tokens. The categories used may be adjacent to each other, e.g. '0' and '1', or they may spread, e.g. '0' and '2'. Most of the tokens received judgments using two adjacent categories with about 3/4 of the tokens using only one or two categories. Listener consistency was further examined according to vowel type as shown in Table 4.3. For some vowels, e.g. /ɑ/ and /u/, the judgments were less consistent, than other vowels with three or four categories chosen at a greater frequency.

In another analysis of listener consistency, the first and second repetitions of a given token were compared by using measures of association in ordered contingency tables. Since the repetitions were measured on an ordinal scale, the test made use of the ordered relationship to determine the direction of association between variables. The null hypothesis tested was that the two repetitions are independent with no association, implying that the listener is highly inconsistent. The Goodman and Kruskal gamma coefficient (γ) was calculated. This measure required the total number of concordant minus the number of discordant pairs. Assume 'X' is the nasality judgment on the first repetition of a given token and 'Y' is the judgment on the second repetition of the same token. Observation of token i, (X_i, Y_i), and observation of token j, (X_j, Y_j), are concordant when $(X_i-X_j)(Y_i-Y_j)$ is positive so that the difference between X components has the same sign as the difference between

Table 4.3: The effect of vowel type on listener consistency in making judgments on synthetic stimuli with manipulations of A1-P1. The number of categories used for four repetitions of a given token is from one to four. The categories may be adjacent to each other or spread apart. The number of times a given category was chosen in presenting twenty versions to each of the eight listeners is listed.

	Number of categories for four repetitions					
Vowel	One	Two adjacent	Two spread	Three adjacent	Three spread	Four
i	45	76	12	24	2	1
I	41	89	4	22	3	1
e	31	101	3	25	0	0
ε	36	85	13	22	4	0
ɑ	40	75	6	26	8	5
ʌ	39	89	5	23	4	0
u	31	86	8	27	7	1
æ	51	68	13	23	4	4
ɔ	45	78	8	27	2	0
o	42	86	1	27	2	2

the Y components. Discordant means that the product is negative. A '1' indicates a perfect positive relationship, so that when X increases, Y also increases; a '-1' indicates a perfect negative relationship. The coefficient γ can achieve the values -1 and 1 since the actual number of tied pairs was used. Similarly, the coefficient was also calculated for the third and the fourth repetitions. Table 4.4 shows the values for γ of the first half comparing the first two repetitions and the second half comparing the last two repetitions. The proximity of γ to +1 and the similarity between values for the first half and the second half for a given subject indicate that subjects were fairly consistent in making their judgments over time. Listeners C and G were relatively less consistent than the other listeners, which corresponded to the results in Table 4.2. Listener consistency, may be improved if the stimuli of the same vowel type are presented in isolation.

Listener's reliability was examined by determining the judge-to-judge correlations as shown in Table 4.5. The average of the judge-to-judge correlations, shown in the last column, is the reliability r of each judge. The reliability varied from 0.59 to 0.68 indicating that no judge significantly disagreed with the other judges.

Table 4.4: Goodman and Kruskal Gamma coefficient γ of comparing judgments made on repetitions 1 and 2 (comparison 1) and those on repetitions 3 and 4 (comparison 2) for a given synthetic token with A1-P1 manipulated.

Subject	Comparison	γ
A	1	0.972
	2	0.929
B	1	0.948
	2	0.893
C	1	0.876
	2	0.897
D	1	0.920
	2	0.909
E	1	0.934
	2	0.917
F	1	0.889
	2	0.923
G	1	0.827
	2	0.838
H	1	0.953
	2	0.956

Table 4.6 shows the listeners' mean reliability by vowel types. Some vowels received a higher reliability, e.g., /I/ and /e/, while other vowels received a much lower reliability, e.g., /ɑ/ and /æ/. The reliability values suggest that listeners disagree more on the nasality perception of low vowels which require larger velopharyngeal opening than high vowels. Table 4.6 also shows the effective reliability which is the reliability of the total set of judges. By using the Spearman-Brown formula (Rosenthal and Rosnow, 1991), the effective reliability may be estimated by:

$$R = \frac{nr}{1+(n-1)r} \qquad 4.1$$

where R is the effective reliability, n is the number of judges, and r is the mean correlation among all of the judges. Because of the large n, the effective reliability of judges for all of the tokens, 0.93, is much larger than the mean reliability of 0.63.

Table 4.5: Judge-to-judge correlation with the reliability r of a single judge from nasality perceptual judgments of synthetic token with manipulations on A1-P1.

Subj.	A	B	C	D	E	F	G	H	r
A	1.00	0.57	0.74	0.60	0.60	0.58	0.48	0.63	0.60
B	0.57	1.00	0.65	0.71	0.63	0.64	0.72	0.56	0.64
C	0.74	0.65	1.00	0.71	0.69	0.69	0.57	0.71	0.68
D	0.60	0.71	0.71	1.00	0.71	0.68	0.66	0.61	0.67
E	0.60	0.63	0.69	0.71	1.00	0.55	0.69	0.52	0.63
F	0.58	0.64	0.69	0.68	0.55	1.00	0.55	0.70	0.63
G	0.48	0.72	0.57	0.66	0.69	0.55	1.00	0.48	0.59
H	0.63	0.56	0.71	0.61	0.52	0.70	0.48	1.00	0.60

Table 4.6: Mean reliability (r) of judges in perceiving nasality of synthetic vowels with manipulation of A1-P1 by vowel type. The calculated effective reliability (R) which is the reliability of the total set of judges by using the Spearman-Brown formula is also shown.

Vowel	Mean reliability (r)	Effective reliability (R)
i	0.66	0.94
I	0.81	0.97
e	0.82	0.97
ɛ	0.58	0.92
ɑ	0.34	0.80
ʌ	0.64	0.93
u	0.77	0.96
æ	0.51	0.89
ɔ	0.55	0.91
o	0.78	0.97
all	0.63	0.93

In summary, listener consistency for each judge was examined by comparing the number of categories the listener used for four repetitions of the same token. The Goodman and Kruskal gamma coefficient was also calculated for the first and second repetitions as well as the third and fourth repetitions. The listeners were less consistent in judging /ɑ/ and /u/. In general the listeners were consistent with two of the listeners less consistent than the other listeners. Listener reliability showed that no judges greatly disagreed with the other judges, except when judging the low vowels.

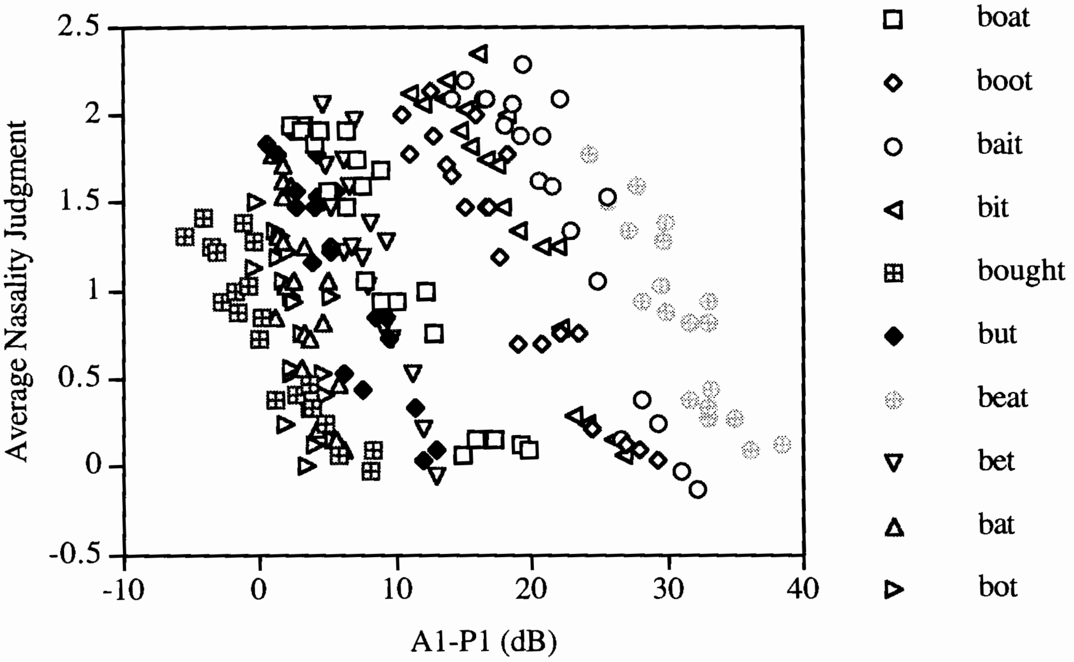


Figure 4.1: Scatter plot of the average nasality judgment versus A1-P1 measured in the middle of ten synthetic vowels and their derivatives.

With the large number of judges, the effective reliability is much greater than the mean reliability.

4.1.4 Results and Discussion

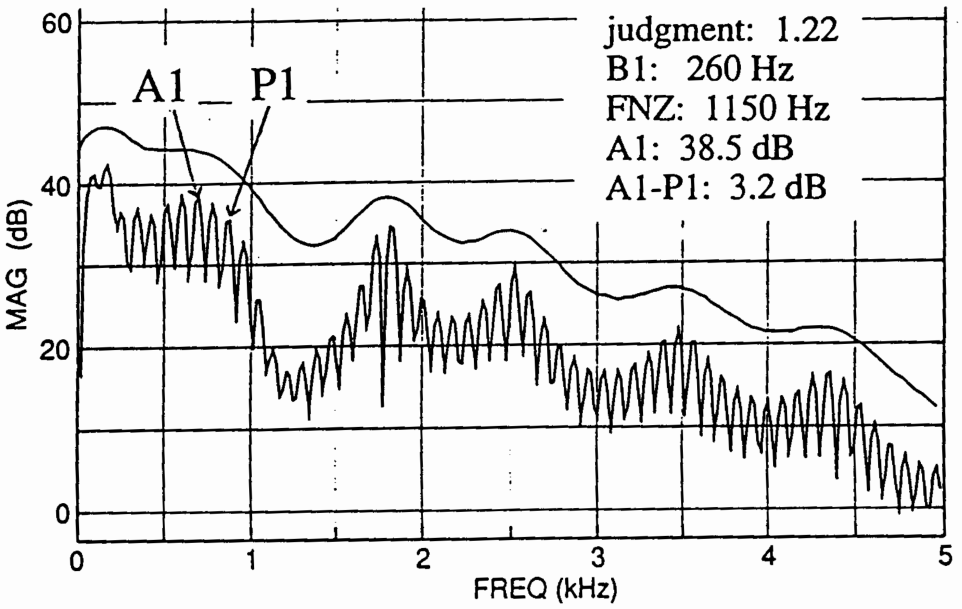
To test the hypothesis that A1-P1 is a measure of nasality as well as nasalization, the correlation between the average nasality judgments and A1-P1 were determined for each vowel type as reflected in the scatter plot in Fig. 4.1. The correlation coefficient of nasality judgments and A1-P1, $\rho_{A1\text{-}P1}$, is shown in Table 4.7. For all of the vowels, $\rho_{A1\text{-}P1}$ was statistically significant; for /ɑ/ and /æ/ the correlation was relatively lower than the others. The correlation coefficient of average nasality judgments and A1, ρ_{A1}, is also shown in Table 4.7. In general, A1-P1 was demonstrated to be a better parameter than A1 in assessing the degree of nasality. The reason is illustrated by the examples in Fig. 4.2, showing spectra taken from the midpoints of two synthetic versions of /ɛ/. Although the bottom spectrum has a wider first formant, the value of A1 is actually higher than expected due to the effect of the extra pole at a frequency close to the first formant. The values of A1 in the two stimuli are fairly

*Table 4.7: Correlation coefficients of average nasality judgments of synthetic stimuli with A1-P1 (ρ_{A1-P1}), and A1 (ρ_{A1}) measured from the spectra of the vowel. The symbols ** and * indicates p = 0.001 and p = 0.005, respectively.*

Word	ρ_{A1-P1}	ρ_{A1}
boat	-0.94**	-0.97**
boot	-0.94**	-0.91**
bait	-0.93**	-0.97**
bit	-0.93**	-0.96**
bought	-0.93**	-0.88**
but	-0.92**	-0.71**
beat	-0.90**	-0.79**
bet	-0.90**	-0.58*
bat	-0.79**	-0.35
bot	-0.67**	-0.31

close but those of A1-P1 differ, influencing the difference in the average nasality judgments.

The relationship between the nasality judgments and the synthesis parameters FNZ and B1 were also examined. Figure 4.3 shows the mean nasality perception judgments for all of the vowels plotted as a function of FNZ, with increase of B1 from that of the non-nasal reference as the parameter. To explain how the spectral characteristic of varying FNZ affects perception results, spectra from the middle of the synthetic vowel in "bat" with B1 as 250 Hz, i.e. increase in B1 of 100 Hz, and FNZ at 950 Hz, 1050 Hz, and 1350 Hz were obtained, as shown in Figures 4.4a-c, respectively. Even though FNZ increases from 950 Hz to 1050 Hz, the effect of the zero on the harmonics adjacent to F1 is still great so that the vowel was actually perceived to be less nasal, which corresponds to the slight increase of A1-P1, as shown in Fig. 4.4a and Fig. 4.4b. From Fig. 4.3, this effect of FNZ seems greater when the bandwidth of the first formant is large. However, as FNZ moves up in frequency, the effect of the nasal pole dominates that of the zero so that the harmonics adjacent to F1 are raised so that the first formant would have an overall wider



(a)

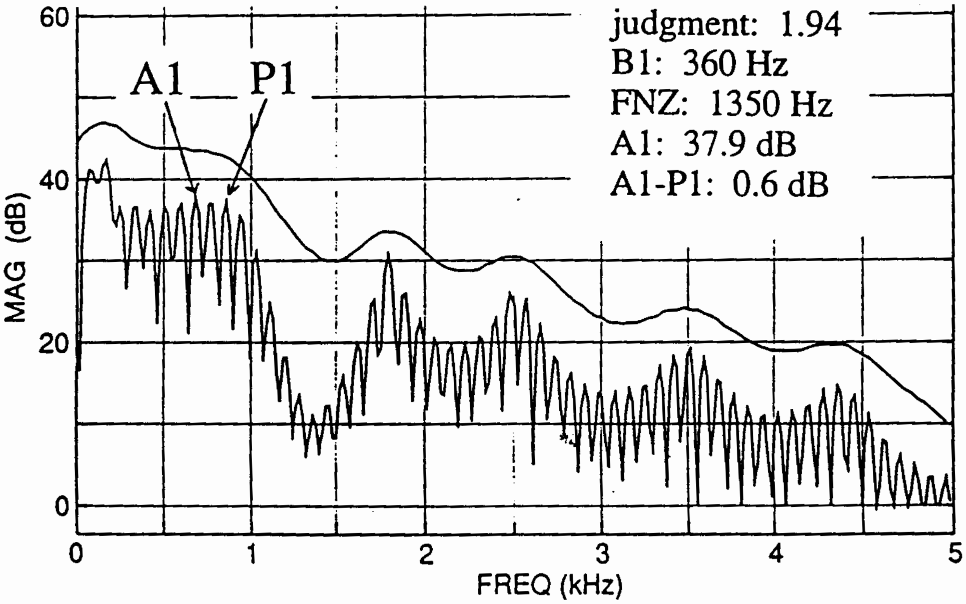


Figure 4.2: Spectra of the synthetic vowel in "bet" for two conditions of nasalization, illustrating the comparison between A1-P1 and A1 as the measuring parameter for nasality perceptual judgments. (a) With B1 equal to 260 Hz and FNZ equal to 1150 Hz, the average nasality judgment was 1.22. (b) With B1 equal to 360 Hz and FNZ equal to 1350 Hz, the average nasality judgment was 1.94. Judgment is the average value of eight listeners judging nasality on a scale from -1 to 3.

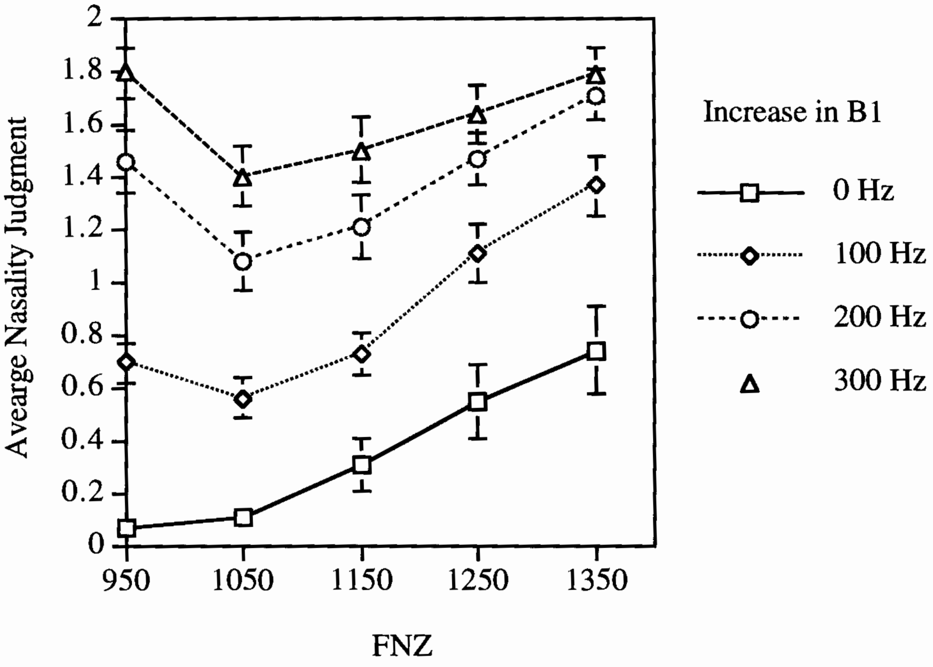
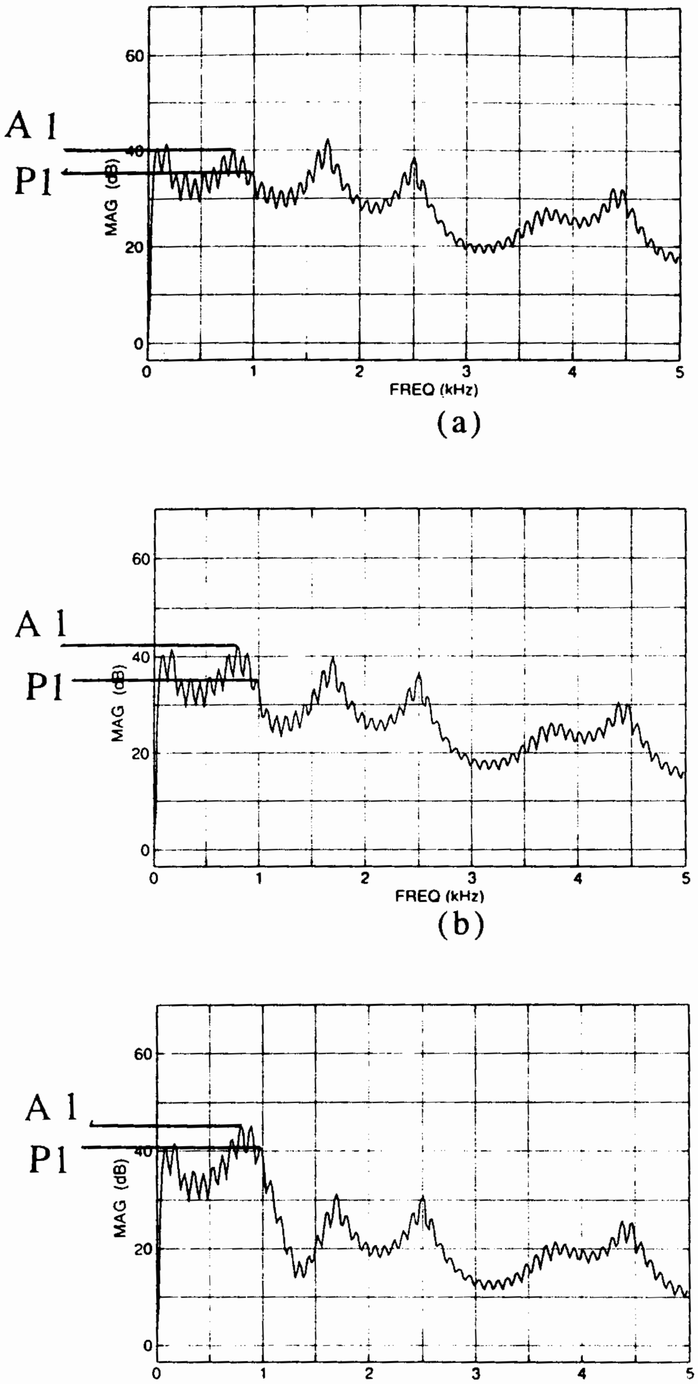


Figure 4.3: The mean nasality perceptual judgments for all of the synthesized vowels are plotted as a function of the frequency of the nasal zero, FNZ, with various increases in bandwidth of the first formant, B1, from that of the non-nasal reference. The nasal pole, FNP, was constant at 950 Hz.

bandwidth causing the vowel to sound more nasal. From Figs. 4.4b and 4.4c, A1-P1 is smaller for FNZ at 1350 Hz than for FNZ at 1050 Hz. The above observation indicates that for FNZ close to FNP, the zero may have a greater effect on the nasal peak and cause A1-P1 to increase. As FNZ moves further away from FNP, the nasal pole has greater effect on the nasal peak and therefore A1-P1 decreases.

Focusing on the effect of greater nasal coupling with FNZ greater than 1050 Hz, we see that FNZ moving from 1050 Hz to 1350 Hz increases nasality perception, as shown in Fig. 4.3. The average nasality judgment increases the most for a given increase in FNZ for stimuli with B1 increase of 100 Hz, giving a steeper slope. The slope is flatter for stimuli with a B1 increase of 0 Hz and 200 Hz, and it is flattest for stimuli with B1 increase of 300 Hz. There is a trend indicating that as B1 increases, the perception of nasality increases, although the effect is less when B1 is already large. The result also indicates a ceiling effect of varying B1 and FNZ for stimuli that already have a wide B1.

The nasality judgments were subjected to a repeated-measures analysis of variance on 3 vowel types (high, mid, and low) x 4 B1 variations (0 Hz, 100 Hz, 200 Hz, and 300 Hz greater than the non-



(a)

FNZ= 950 Hz

A1-P1 = 5 dB

(b)

FNZ= 1150 Hz

A1-P1 = 8 dB

(c)

FNZ= 1350 Hz

A1-P1 = 5 dB

Figure 4.4: The spectra taken from the middle of the synthetic vowel in "bat" with the bandwidth of the first formant at 250 Hz, i.e. increase in B1 by 100 Hz and (a) FNZ at 950 Hz, (b) FNZ at 1050 Hz, and (c) FNZ at 1350 Hz are shown in order to examine the smaller mean nasality perception for stimuli with FNZ at 1050 Hz.

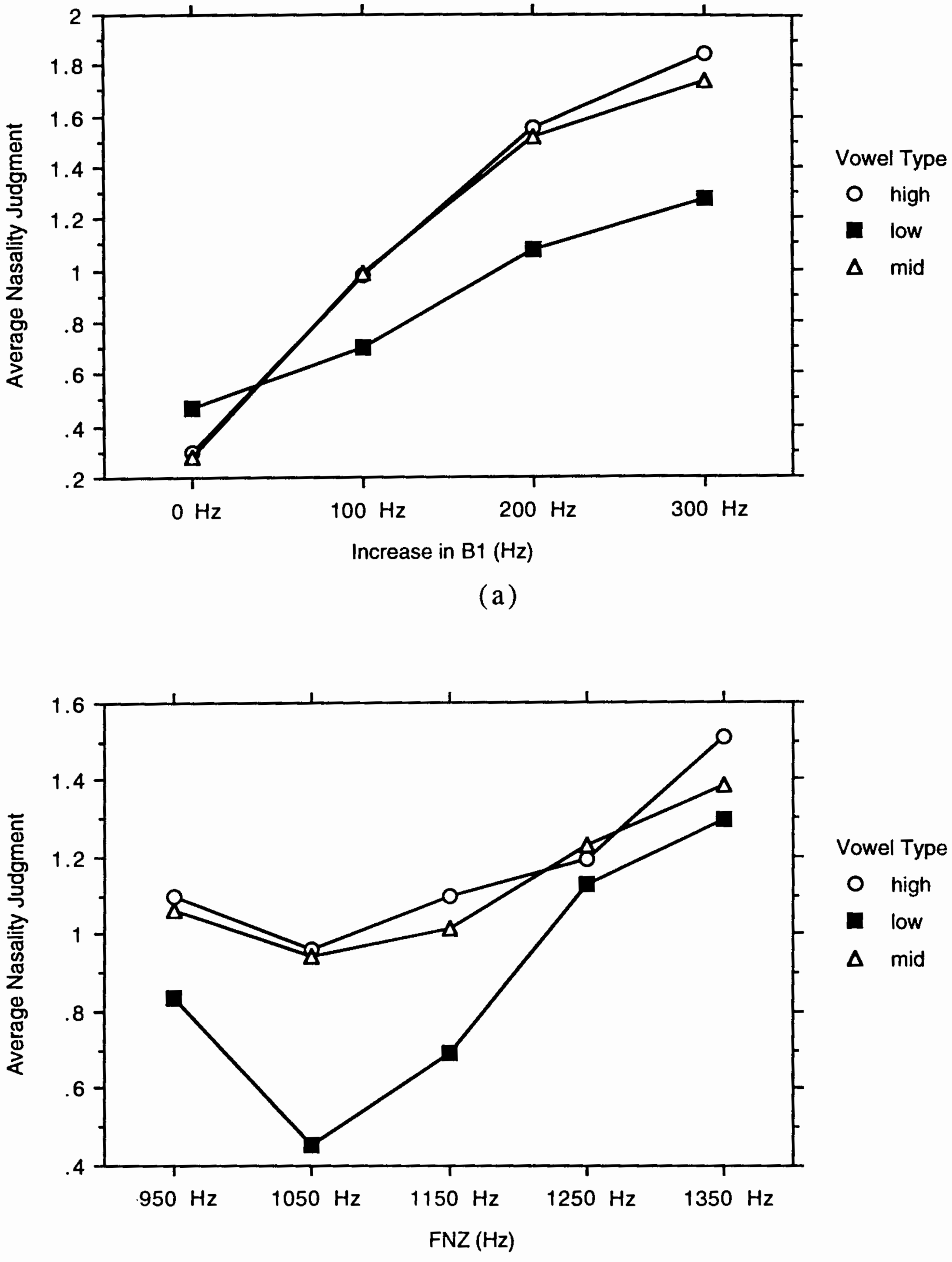
nasal stimulus) x 5 FNZ variations (950 Hz, 1050 Hz, 1150 Hz, 1250 Hz, and 1350 Hz) and 8 judges. The high vowels consisted of the synthetic stimuli of "beat", "bit", and "boot"; the mid vowels consisted of "bait", "bet", "but", "boat", and "bought"; and the low vowels consisted of "bat" and "bot". There was a significant effect for vowel types, [F(2,14) = 4.40, $p < 0.033$] and a highly significant effect for B1 [F(3,21) = 125.36, $p < 0.0001$] and for FNZ, [F(4,28) = 11.44, $p < 0.0001$]. The B1 x FNZ interaction was also highly significant, [F(12,84) = 4.51, $p < 0.0001$].

The vowel type x B1 x FNZ interaction was not significant, [F(24, 168) = 1.17, $p < 0.27$]. This observation allowed one to analyze the effect of B1 and FNZ on vowel type separately. The vowel type x B1 interaction was highly significant , [F(6,42) = 8.87, $p < 0.0001$]. Figure 4.5a and Fig. 4.5b show the effect of increasing B1 and FNZ, respectively, on nasality judgments for the three groups of vowels: high, mid and low. In Fig. 4.5a, the high- and mid-vowel groups behaved similarly; the low-vowel group had slightly higher nasality perception when B1 was narrow but as B1 increased, nasality judgments were lower than those of the high- and mid-vowel groups. From Figs. 2.7 and 2.8, for a given velopharyngeal opening size A_{vp}, greater losses are introduced for low vowels due to their high first formant frequency. Therefore, the low vowels may require a wider B1 to be perceived as nasal than other vowel types. The vowel type x FNZ interaction was highly significant, [F(8,56) = 2.93, $p < 0.0084$]. In Fig. 4.5b, FNZ also had similar effect on high and mid vowels. For low vowels, the nasality judgment varied over a wider range for FNZ up to 1250 Hz since the high F1 was closer to FNZ so that the zero had greater influence on the F1 prominence. The effect on nasality depends on the location of FNZ. As FNZ moved above 1250 Hz, it was far enough from F1 that the judgment of nasality was similar for the different vowel types. For a given FNZ, the nasality judgments were lower for low vowels than the other vowel types, again reflecting the fact that low vowels require a larger velopharyngeal opening to be perceived as nasal.

4.2 Effects of Manipulating A1-P0

4.2.1 Stimuli

The synthetic stimuli used in the A1-P1 experiment (Sec. 4.1.1) were further manipulated to determine the perceptual effects of systematically changing the parameters that influence A1-P0. Only



(a)

(b)

Figure 4.5: (a) The effect of manipulating B1 in synthesis on nasality judgments averaged across FNZ and (b) the effect of manipulating FNZ on the nasality judgments averaged across B1 for three groups of vowels: high, mid and low.

six vowels were used: /i/, /u/, /ɛ/, /ʌ/, /æ/, and /ɑ/. The synthetic version of the original words (Group 0) was also used as the non-nasal reference in this experiment. A second group of stimuli (Group 1) was generated by setting the values of B1 and FNZ to ones that received the median value of nasality judgments from the A1-P1 experiment. These changes affect A1-P0 by lowering A1 with P0 being unchanged. A1-P0 was further varied (Group 2) by changing P0 while keeping the values of the synthesizer parameters B1 and FNZ unchanged from those of Group 1, so that further changes in perception of nasality would be solely due to the alteration of P0. The open quotient OQ of the glottal waveform (using the KLGLOTT88 source in the synthesizer KLSYN88) was manipulated to vary P0 since it increases the amplitude of the first harmonic, which is often used to measure P0. The open quotient was manipulated so that A1-P0 would span the range of nasal and non-nasal vowels of French speakers who had larger $(\mathrm{A1\text{-}P0}_{\text{non-nasal}})-(\mathrm{A1\text{-}P0}_{\text{nasal}})$ than that of English speakers, as observed from acoustic analysis. For each vowel type, the range was determined by using the French vowel that had the most similar F1 and F2 location to the vowel. For /i/, /ɛ/, and /æ/ a range of 7 dB according to that of /ɛ̃/ was used; for /u/ a range of 9 dB was used according to that of /ɑ̃/; for /ɑ/ and /ʌ/ a range of 9 dB according to that of /ɔ̃/ was used. This required an OQ of 50, 60, and 80 percent for synthesizing /i/; an OQ of 50, 70, and 90 percent for synthesizing the other five vowels so that the values of A1-P0 would be evenly spaced. Stimuli in Group 0 and Group 1 had the OQ fixed at the default value of 50 percent; for Group 2, OQ was set at 50 for the first 50 ms and then increased linearly to 60, 70, 80, or 90 percent toward the end of the vowel. (A1-P0 was observed to monotonically decrease throughout the vowel in the French utterances.)

4.2.2 Procedure

The same procedure as that for the A1-P1 experiment (see Sec. 4.1.2) was used with these test stimuli. The pairs contained the same references (synthesized version with no nasalization) followed by one of its variations. Each item in the test was also indicated to the listener in a written form. Four repetitions of all of the vowel types were randomized and presented.

The listeners were tested individually and the same categories as in the A1-P1 experiment were used for the judgments. Each listener was asked to compare the second member of the pair with the first member (the reference) on a scale from -1 to 3 by writing the response (-1, 0, 1, 2, or 3) on a blank for the test item. Before each test, instruction was given orally and written on the response sheet. Also, the listeners were specifically told that there is no 'correct' answer and that there was no attempt to create an even distribution of answers on the test scale. Before the test, it was ascertained that they were able to distinguish the nasal and non-nasalized words. The first 24 repetitions of the test pairs were used solely for practice and not for scoring.

4.2.3 Listeners

Three male and six female native speakers of English participated in the experiment. No listener had speech or hearing impairments. They were students or staff at the MIT Speech Communication group, so they were familiar with the perception of nasality. Their ages were under 40. This group of listeners were different from the group that participated in the experiment on the manipulation of A1-P1 since most of the previous listeners were no longer available.

The same technique as in Sec. 4.1.3 was used to determine how consistent each listener was in performing the judgments. Table 4.8

Table 4.8: Listener consistency in making judgments on stimuli with manipulations of A1-P0. The number of categories used for four repetitions of a given token is from one to four. The categories may be adjacent to each other or spread apart.

	Number of categories for four repetitions					
Subject	One	Two adjacent	Two spread	Three adjacent	Three spread	Four
A	6	13	1	2	2	0
B	3	10	2	8	1	0
C	7	10	1	6	0	0
D	10	9	2	3	0	0
E	5	15	0	4	0	0
F	5	12	1	5	1	0
G	6	11	1	6	0	0
H	6	13	1	2	1	0
I	6	8	2	7	0	1

Table 4.9: The effect of vowel type on listener consistency in making judgments on synthetic stimuli with manipulations of A1-P0. The number of categories used for four repetitions of a given token is from one to four. The categories may be adjacent to each other or spread apart.

	Number of categories for four repetitions					
Vowel	One	Two adjacent	Two spread	Three adjacent	Three spread	Four
i	10	16	3	5	2	0
ɛ	6	20	2	6	1	1
æ	8	17	2	9	0	0
u	17	15	1	3	0	0
ɑ	6	18	1	11	0	0
ʌ	7	16	1	12	2	0

shows the frequency a given number of categories was used by each subject for making judgment on the four repetitions of each token. The categories used may be adjacent to each other or they may be spread apart. Most of the tokens received judgments using two adjacent categories, with about 2/3 of the tokens using only one or two adjacent categories. Listener B seemed to be the least consistent with a greater number of judgments using three categories and fewer judgments using only one category. Listener consistency was further examined according to vowel type, as shown in Table 4.9. Unlike the previous experiment, /u/ received judgments that were the most consistent, with one and two adjacent categories chosen with greater frequency, while the other vowels received judgments with about the same consistency. The γ coefficient was not calculated since too few samples were involved in this experiment. To improve the consistency of the judgments, the pairs may be presented in groups by vowel types to make the task easier.

Listener reliability was also examined by determining the judge-to-judge correlation, as shown in Table 4.10. The reliability of a single judge (r) varied from 0.16 to 0.83. Except for the listener with r of 0.16, the judges showed higher reliability than those in the experiment on A1-P1. This result may be due to the fact that the listeners in this experiment were more experienced in perceiving nasality, or the task for this experiment with fewer tokens was easier to perform than the experiment on A1-P1. Table 4.11 shows listener reliability by vowel types. The overall reliability was 0.76. Some vowels had high reliability, e.g. /u/ and /ɛ/, while other

Table 4.10: Judge-to-judge correlation with the reliability r of a single judge from perceptual judgments of nasality of synthetic token with manipulations on A1-P0.

Subj.	A	B	C	D	E	F	G	H	I	r
A	1.00	0.25	0.74	0.81	0.87	0.79	0.87	0.82	0.80	0.81
B	0.25	1.00	0.31	0.32	-0.05	0.00	0.17	0.14	0.11	0.16
C	0.74	0.31	1.00	0.71	0.60	0.60	0.76	0.48	0.53	0.63
D	0.81	0.32	0.71	1.00	0.79	0.73	0.85	0.68	0.66	0.75
E	0.87	-0.05	0.60	0.79	1.00	0.89	0.90	0.81	0.85	0.82
F	0.79	0.00	0.60	0.73	0.89	1.00	0.85	0.84	0.76	0.78
G	0.87	0.17	0.76	0.85	0.90	0.85	1.00	0.77	0.78	0.83
H	0.82	0.14	0.48	0.68	0.81	0.84	0.77	1.00	0.69	0.73
I	0.80	0.11	0.53	0.66	0.85	0.76	0.78	0.69	1.00	0.72

vowels had low reliability, e.g. /ɑ/ and /æ/, with /ɑ/ receiving a negative reliability. This result also suggests that listeners disagreed more on the nasality perception of low vowels, which require larger velopharyngeal opening than high vowels. The effective reliability shown in Table 4.11 depended on the number of judges and the mean correlation among all of the judges, as shown in Eq. 4.1. Because of the large number of judges, the effective reliability of judges for all of the tokens, 0.96, was much larger than the mean reliability, 0.76.

Table 4.11: Mean reliability (r) of judges in perceiving nasality of synthetic vowels with manipulation of A1-P0 by vowel type. The calculated effective reliability (R), which is the reliability of the total set of judges by using the Spearman-Brown formula, is also shown.

Vowel	Mean reliability (r)	Effective reliability (R)
i	0.81	0.97
ɛ	0.94	0.99
ɑ	-0.03	-0.30
æ	0.49	0.88
u	0.97	0.99
ʌ	0.85	0.98
all	0.76	0.96

4.2.4 Results and Discussion

The correlation coefficients between the average nasality judgments and A1-P0 are shown in Table 4.12. The correlation

Table 4.12: Correlation coefficients, ρ_{A1-P0}, *of average nasality judgments of synthetic stimuli with A1-P0 measured from the spectra of the vowel are listed. The low vowels have the highest absolute value of the correlation.*

Word	ρ_{A1-P0}
Beat	-0.79
Bet	-0.70
Bat	-0.91
Boot	-0.72
Bot	-0.99
But	-0.91

was statistically significant for all of the words, with the absolute value of the correlation of "bat", "bot", and "but" being relatively higher. This finding suggests that A1-P0 is a better measure for low vowels, which have high F1 so that P0 may be observed more easily than P1.

Figure 4.6 shows the average nasality judgment for a given change of A1-P0 from that of the non-nasal reference $(A1\text{-}P0)_{ref}$. The first set of data (solid data points) with no alteration of A1-P0 is from Group 0 with stimuli identical to the non-nasal references. They received average judgments close to zero, indicating that no difference in nasality was perceived between the reference and the test token. The next set of data (half solid data points) with A1-P0 reduced by lowering A1 via widening the first formant bandwidth B1 while also increasing FNZ (Group 1) shows that the vowels with lower F1, including the high vowels, received relatively greater nasality judgments than those with high F1, including the low vowels. This occurred even for a small change in A1-P0, e.g. 1 dB for /i/. The difference between the average nasality judgments for Group 1 and Group 0 was highly statistically significant except for /ɑ/ with $p = 0.17$. This indicates that more than altering A1-P1 may be necessary for low vowels to be perceived as nasal.

Another way to increase nasality perception is to have a more prominent P0. The rest of the data points (unfilled data points) in Fig. 4.6 are from Group 2, with increased OQ to further alter A1-P0. In general, when the OQ was raised to change A1-P0 slightly, e.g. 2 dB to 4 dB, little effect was seen in the nasality judgments. In fact, the difference between the nasality judgment for Group 1 and these data points for Group 2 was not statistically significant. However, as the OQ was increased further, it altered A1-P0 enough

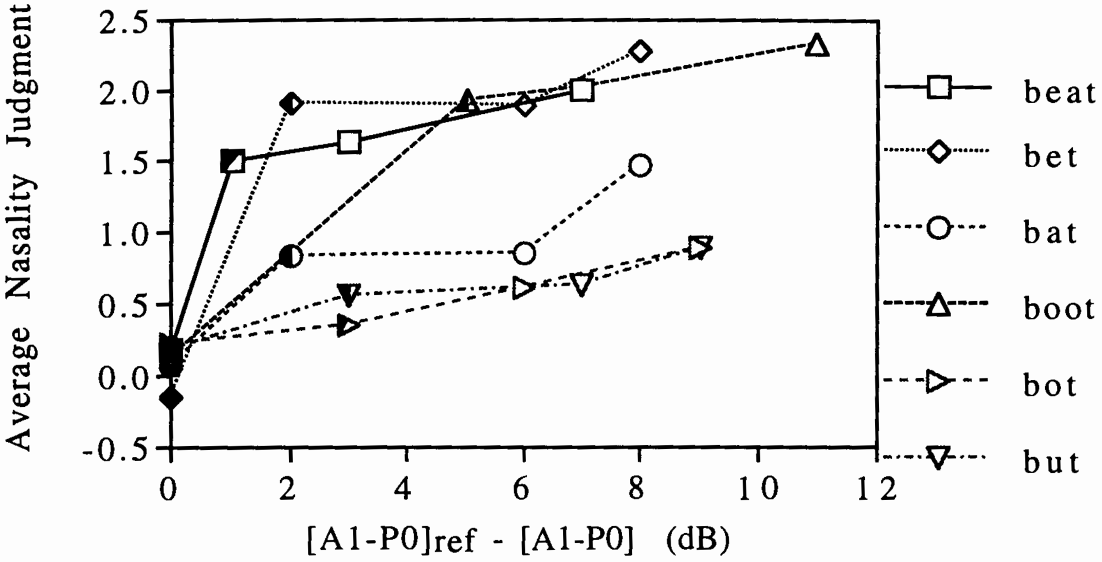


Figure 4.6: The average nasality judgment across nine listeners on synthetic vowels is plotted for a given change of A1-P0 from that of the non-nasal reference $[A1\text{-}P0]_{ref}$. *Group 0 consists of filled data points with test stimuli identical to the non-nasal references. Group 1 consists of half-filled data points with A1-P0 reduced by lowering A1. Group 2 consists of unfilled data points with A1-P0 reduced by raising P0 in the stimuli from Group 1.*

to change nasality perception. The difference between the nasality judgment for the data points in Group 1 and the data points in Group 2 with greater OQ was highly statistically significant for all of the vowels. Since A1 and P1 manipulation has a relatively smaller effect on the nasality perception of low vowels, the effect of raising P0 on nasality perception may be more significant for low vowels than for other vowels. Greater coupling to the nasal passages and thus the paranasal sinuses can also raise P0, which would affect the overall prominence of the first formant, causing the vowels to be perceived as more nasal.

4.3 Summary

According to analysis of a model of vocal tract acoustics, A1-P1 is a better spectral measure of nasalization for the non-low vowels while A1-P0 is better for the non-high vowels. To further quantify the relationship of these acoustic correlates and nasality perception, synthesis was used to determine the perceptual effects of systematically manipulating specific acoustic parameters relevant to nasalization. In one experiment, versions of the synthesized non-nasal reference was generated by changing the first formant bandwidth up to 300 Hz greater than that of the non-nasal reference, and the frequency of the nasal zero was changed up to 400 Hz higher than that of the nasal pole, which was fixed at 950 Hz. In another experiment with the same non-nasal reference, one set of additional stimuli had the first-formant bandwidth and the frequency of the nasal zero equal to those of the token that received the median value of the average nasality judgment from the above experiment. Another set of stimuli was generated by further manipulation involving the open quotient ranging from 50 to 90 percent.

Statistically significant correlation coefficients between the average nasality judgments and A1-P1 as well as between the average nasality judgments and A1-P0 were found. For A1-P1, the correlation was relatively higher for the non-low vowels while for A1-P0, it was relatively higher for the non-high vowels. This finding supports the use of A1-P1 for vowels with low F1 and A1-P0 for vowels with high F1 in spectral analysis. For vowels with a low F1, P1 is well above F1 and can be observed more readily. Likewise, for vowels with a high F1, P0 is well below F1, and can be readily observed and measured.

Examination of the relationships of synthetic parameters related to A1-P1, i.e. B1 and FNZ, and nasality judgments indicates that as B1 widened, the perception of nasality increased, especially when B1 started out to be small. The effect of B1 widening on perception was less when FNZ was much higher than FNP. There was a ceiling effect of varying B1 and FNZ for stimuli that already had a wide B1. The high- and mid-vowel groups behaved similarly. In general, the low vowel group had lower nasality judgments than the high- and mid-vowel groups for a given B1 as well as for a given FNZ. The relationships of synthetic parameters related to A1-P0, i.e. B1 and OQ, with nasality judgments indicates that as B1 widened, high vowels received relatively greater nasality judgments than the low vowels. As the OQ of the glottal waveform was increased to 60 or 70 percent, little effect was seen on the nasality judgments. However, as the OQ was increased further to 80 or 90 percent, it altered A1-P0 enough to change nasality perception significantly. The effect was especially noticeable for the low vowels.

Chapter 5

Endonasal Sinus Surgery Effects on Nasal Vowels

Once the baseline of A1-P1 and A1-P0 was established for the normal speakers by acoustic analysis and synthesis, the acoustic correlates were further examined in the speech of people who required endonasal sinus surgery. It is likely that the acoustic signal of their nasal sounds would be affected by the anatomical changes from the surgery. The study of their speech before and after the surgery would provide additional data for refining the model for the production of nasal sounds as well as a technique for predicting how a patient's speech is likely to change as a result of nasal surgery.

5.1 Background

5.1.1 Sinusitis

Inflammation of the paranasal sinuses, known as sinusitis, is the number one chronic illness in the United States, affecting one out of eight Americans (Levine, 1990). Refer to Section 2.1 for the anatomy of the nasal cavity and sinuses, with Fig. 2.1 and Fig. 2.2 showing the lateral view and the coronal view, respectively.

5.1.1.1 Causes

Normally the sinuses have adequate drainage through the mucociliary pathway of the natural ostium. However, lack of ventilation due to obstruction at the ostiomeatal complex (OMC), which is a term given by Naumann (1965) to the anterior ethmoid region, especially the infundibulum at the entrance to the maxillary sinus and the frontal recess, can lead to sinusitis.

Tight spaces in the nasal cavity, which lead to obstruction and ultimately sinusitis (Zinreich *et al.* 1987 and Stammberger and Wolf, 1988), can be due to anatomical variations, polyps, protruding mucosa, and localized hyperplasia. Anatomical anomalies, occurring bilaterally as well as unilaterally, include deviated septum, medial deviation or bony hyperplasia of the uncinated process, abnormal anatomy of the maxillary ostium, “paradoxical curvature and pneumatization” of the middle turbinate (concha bullosa), Haller’s cells, and expansion of the ethmoidal bulla. Concha bullosa is believed to be due to anatomic alteration of the ethmoid air cell development (Lothrop, 1903). Haller’s cells are also due to the ethmoid air cells expanding along the maxillary sinus roof. The expansion of the ethmoid bulla would block the hiatus semilunaris, which leads to the ethmoidal infundibulum, since the anterior portion of the ethmoid bulla is just superior to the hiatus semilunaris.

The formation of the polyps involves epithelial necrosis caused by the obstruction of the middle meatus due to mucosal edema from allergy or nasal inflammation. Nasal polyps seem to originate from the nasal mucosa of the middle meatus and generally extend laterally into the anterior part of the middle meatus, according to a study of six patients by Larsen and Tos (1991). Polyps can also originate from the septum, the maxillary sinus, the lateral wall of the nose, anterior roof of the nose, and other places. The location of the polyps may be due to the fact that the maxillary sinus and ethmoid cells drain their secretions into the middle meatus so that the ostiomeatus is more exposed to pathological stimuli. Secondary polyps may develop due to blockage of the ostiomeatal complex by the nasal polyp.

Mucosa in the narrow spaces is often pressed together so that mucus is retained, therefore allowing virus and/or bacterial growth. This growth would lead to cilia damage and increased goblet cells in the nasal glands, causing mucosal thickening, cysts, and highly viscous mucous. Because mucous transport is slower than its production, ostiomeatal obstruction results, which can lead to chronic sinusitis (Kennedy *et al.*, 1985; Stammberger and Wolf, 1988). In a study of coronal CT scans, Bolger (1991) found that more than 80% of the patients with abnormal mucosa in the anterior ethmoid region have disease in the frontal or maxillary sinuses while normal mucosa in that region corresponds to normal sinuses.

5.1.1.2 Symptoms and Signs

Chronic sinusitis consists of several major symptoms that include cephalea, blocked nasal respiration, polyposis, pathologic secretions from the nose, and postnasal discharge. These classic symptoms may lead to secondary symptoms such as mycoses of the sinuses, mucoceles, and sinobronchial syndrome. There are also signs such as purulent nasal or postnasal drainage, erythematous boggy nasal mucosa, nasal polyposis, facial swelling, and recurrent or chronic otitis media with effusion. The most frequent complaints are nasal discharge and nasal congestion with headache or facial pain being less common. The symptoms and signs may persist for more than three months despite medical treatment with broad spectrum oral antibiotic, beclomethasone dipropionate nasal spray, and nasal decongestant spray.

5.1.1.3 Diagnostic Evaluation

There are various tools that the surgeon uses for diagnosis, including computed tomography (CT) scans, endoscopy, and a staging system. CT scan has been shown to be extremely reliable for evaluating paranasal sinus pathology (Som, 1985). It is done with the patient on the scanner bed in the supine position with head hyperextended. Five-mm sections and sometimes 3-mm sections are obtained from the anterior of the frontal sinus to the posterior of the sphenoid sinus. CT scan allows not only direct bilateral comparison but also precise location of the disease because of the highly contrasting densities due to air in the bony sinuses, fat within the orbit, and soft tissue surrounded by air in the nasal cavity. It is especially helpful in patients with nasal polyposis and allergic rhinitis since the telescopes cannot pass into the nasal cavity. Furthermore, computer tomography detects even subtle paranasal sinus and ostiomeatal complex anatomy variations, which may be overlooked by rhinoscopy or endoscopy. Thickening of the sinus mucosa shown on CT may indicate sinusitis secondary to obstruction of the ostiomeatal complex. In a detailed analysis of coronal CT scans by Bolger (1991), anatomic variations such as agger nasi cells, pneumatization of the middle turbinates, paradoxical curvature of the middle turbinates, Haller's cells, and pneumatization of the uncinate processes were detected in 65% of the cases, while mucosal abnormalities were detected in 83% of the cases. Computed tomography scan can also distinguish inflammatory disease from hyperplastic disease (Chakeres, 1985). However, it is limited in differentiating between fibrous tissue and inflammatory mucosal disease (Katsantonis *et al.*, 1990).

Endoscopy enables the surgeon to visualize the details of the ostiomeatal complex and paranasal sinuses in addition to the gross pathology. The first attempt of endoscopy in the nose and sinuses was done by Hirschmann in 1901 with a modified cystoscope (Draf, 1983) but it was not until more recently that the first systematic documentation of endoscopic findings was published in English by Messerklinger (1987). Endoscopy is performed with the nasal cavities sprayed with 3% ephedrine and anesthetized with 4% lidocaine spray. A 0-degree 4-mm telescope is used to examine the nasal cavity systematically, first along the floor into the choana to view the overall anatomy, the inferior meatus, and the nasopharynx. It is then angled 30-degrees from the floor to examine the anterior middle meatus, uncinate process, fontanelles, and the inferior middle meatus. It is also passed medial to the middle turbinate to see the superior turbinate and the sphenoid sinus ostium. In addition, the telescope enters the middle meatus in order to examine the ethmoidal bulla, hiatus semilunaris, maxillary sinus ostium, and the frontal recess area. The purpose of an endoscopic examination is to detect any inflammation, turbinate degeneration, tumor, drainage, polyps, anatomic variations, and accessory ostia. Endoscopy can also detect septal deflection and spurs, mucosal edema, turbinate hypertrophy, and adenoid hypertrophy which may not be seen on the CT scans (Vining *et al.*, 1993). Disease in the anterior ethmoid is indicated by edematous mucosa prolapsed into the infundibulum or an inflamed bulla; disease in the frontal recess is indicated by mucosal edema or polypoid mucosa at the attachment of the anterior middle turbinate to the lateral nasal wall. Disease of the maxillary sinus and infundibulum would be reflected by the edema or erythema in the region of the fontanelles and the infundibulum, respectively. Endoscopy can also correctly determine the soft tissue that obstructs the middle meatus.

A staging system for the extent of the sinus disease before the surgery would be useful in guiding treatment strategies and in indicating the prognosis of the disease process. The staging system should be clear, comprehensive, easily remembered and simple to apply. Kennedy (1992) classified the extent of the disease based on preoperative endoscopic examination: nonpolypoid disease, middle meatal polyposis, and diffuse polyposis. In addition, he classified the disease according to CT scans and findings during the surgery based on anatomic abnormality, the involvement of the ethmoid and other sinuses, the presence of polyps, and whether the disease is unilateral or bilateral. Metson (unpublished) developed a five-stage system based on the CT scan: normal with mucosal thickening < 2mm, unilateral disease/anatomic abnormality, bilateral ethmoid/maxillary disease, bilateral disease with frontal/sphenoid involvement, and pansinusitis. Other researchers had staged the disease based on the presence and absence of allergy and polyps (Sogg, 1989), the presence of local or diffuse disease and asthma (Lawson, 1991), and the extension of the polyps into the nose (Levine, 1990).

5.1.2 Surgery

Surgery is recommended to patients with debilitating symptoms that persist after extensive use of antibiotics, anti-allergic medications, or systemic decongestants, and whose endoscopy and/or CT indicate abnormalities such as obstructive mucosal or bony disease in the OMC or paranasal sinuses. However, surgery of the ostiomeatal complex should not be used as a primary therapy since in addition to anatomic obstruction, there may be other underlying factors of chronic sinusitis (Kennedy, 1992).

5.1.2.1 Surgical Technique

The surgery involves careful preparations as well as a systematic surgical technique. Prior to the surgery, attempts are made to minimize inflammation by using oral antibiotics 1 to 3 weeks prior to the surgery. Those with massive polyposis or hyperactive nasal mucosa are on 20 to 30 mg prednisone. Those that are already on steroids would increase their dosage to reduce the chance of bronchospasm and to shrink nasal polyposis. The CT scans should be used to review the anatomy and the extent of the disease prior to the surgery. They should also be used to note the slope and thickness of the roof of the ethmoid sinuses, the position of the uncinate processes, the width of the infundibulums, the shape of the medial orbital walls, the size of the maxillary sinuses, the position of the ethmoid cells, the position of the frontal sinuses, the distance between the maxillary sinus and the posterior ethmoid roof, the relationship of the optic nerve to the posterior ethmoid cells, the size of the sphenoid sinuses, and the position of the internal carotid arteries and the optic nerves.

It was only within the last ten years that endoscopic surgery of sinuses has become more popular in treating sinusitis. In fact, functional endoscopic sinus surgery, FESS, is becoming the procedure of choice in treating both adults and children with diseased sinuses (Lazar *et al.*, 1993). This procedure was introduced by Kennedy *et al.* (1985) by adapting the concept that the obstruction of the ostiomeatal complex, OMC, causes sinusitis. The objective is to allow drainage and ventilation of the paranasal sinuses by opening the OMC without affecting the large paranasal sinuses unless cysts or polyps are present. The treatment of the cause instead of the consequences of sinusitis is feasible since the sinuses are expected to heal once the ventilation and drainage has been reestablished. For chronic or relatively advanced sinusitis, however, a complete removal of the ethmoid cells is necessary since they are lined with mucosa that can be hyper-reactive (Friedman *et al.*, 1986).

Two well-known approaches of FESS are the Wigand and Messerklinger procedures. The Wigand approach begins with a sphenoidotomy and moves anteriorly to the frontal recess (Wigand, 1981) while the Messerklinger approach begins at the ethmoid bulla and moves anteriorly to the frontal recess as well as moving posteriorly for ethmoidectomy and sphenoidotomy if needed (Messerklinger, 1987). The Wigand procedure is more commonly performed in patients with sphenoid sinusitis, while the Messerklinger approach is performed in patients with chronic and recurrent acute sinusitis, fungal sinus disease, and drainage of frontal sinus mucoceles. The performance of the surgery with the Messerklinger approach is described in depth below.

During the surgery the patient is in a supine position with the head slightly elevated and turned to the right facing the surgeon. Nasal endoscopic exam is performed to examine the septum, middle turbinate, posterior nasal airway, and adenoids. Various telescopes are available for better viewing. A video camera is attached to the endoscope so that the minute details can be seen on the video monitor. The 4-mm 0 degree telescope is used for the ethmoid and sphenoid sinuses, the 4-mm 30 degree scope is used for the frontal recess while the 4-mm 70 degree scope grants a view of the maxillary sinus.

Optimal vasoconstriction is necessary to minimize bleeding and to improve visualization. This is done by administering a topical decongestant spray such as Neo-Synephrine or Afrin. In general, local anesthesia with intravenous sedation and anesthesia monitoring is used. The patient's nose is packed along the lateral wall for at least 10 minutes with neurosurgical cottonoid pledgets that were soaked in a 4% cocaine solution for the effect of vasoconstriction and nerve block. The medial infundibular wall, anterior middle turbinate, and ethmoidal bulla are injected with 1% lidocaine hydrochloride and 1:100,000 epinephrine for topical anesthesia and further vasoconstriction. For a surgery on the sphenoid sinus, injections of greater palatine foramen, septum, inferior and middle turbinates, and posterior attachment of the middle turbinate are made.

The surgeon needs to make careful decisions on which anatomic structures, i.e. septum, turbinates, and sinuses, are to be preserved and how others are to be modified. Since clear surgical view is vital, if septum deviation obstructs the view greatly, a septoplasty may be necessary. Resection of the deviated portion is done by hemitransfixion incision and the septum is reconstructed. A 1- or 2-cm opening is often created in the middle meatus for better drainage. The surgery involves resecting and removing the uncinate process adjacent to the anterior portion of the middle turbinate attachment. Resection of the lateral half of the middle turbinate may occur if the concha bullosa inhibits the access to the middle meatus. The middle turbinate should be preserved for the protection of the cribriform plate, to prevent scarring that affects drainage, and to maintain a landmark in later surgeries (Parsons and Phillips, 1993).

The maxillary sinus ostium is opened when there is stenotic opening or maxillary sinus disease. Locating the ostium involves palpation or identifying the uncinate process as a landmark since the opening is lateral to the inferior portion of the residual uncinate process. The ostium is widened by dissecting into the posterior fontanelle and by the removal of any residual uncinate process. The size of the opening is dependent on the extent of the disease. If necessary, the maxillary ostium may be extended into the anterior fontanelle. Superior extension should be avoided to prevent violation of the orbit; anterior widening should also be avoided to prevent injury to the nasolacrimal duct. Sometimes the ostium cannot be found because the uncinated process may be hypoplastic or indistinguishable from the lateral nasal wall or the maxillary sinus ostium simply is not apparent. Then, direct puncture into the maxillary sinus may be necessary to obtain an opening. Sometimes, inferior antrostomy may precede the middle antrostomy for better identification of the middle antrostomy. Through the larger ostium, the maxillary sinus may be viewed by using a 70 degree telescope and any mycotic mass or pus may be removed. The procedure on the maxillary sinus should be performed prior to the manipulation of the ethmoid sinus which may obscure vision due to bleeding.

Ethmoidectomy of the diseased region indicated by CT scan is performed anteriorly to the agger nasi cells and posteriorly through the basal lamina of the middle turbinate. To get access to the posterior ethmoid cells, ethmoid bulla is removed and the ground lamella of the middle turbinate is infractured. The roof of the ethmoid and the skull base should be identified to prevent any complications. Dissection of the posterior ethmoid cells should not be done past the face of the sphenoid sinus to avoid injuring the optic nerve.

The frontal recess is explored only if there is historical or clinical evidence of disease in the frontal sinus region. Endoscopic operation is recommended for patients with either obstruction of the frontal recess or of the anterior ethmoid that limits sinus drainage or for those with mucosal disease in the medial aspect of the frontal sinus so that direct endoscopic visualization is possible. However, endoscopic operation is not recommended for those patients with small frontal sinus ostia that cannot be enlarged, or with disease in the lateral mucosa, or with hypertrophic mucosa obstructing the entire sinus (Schaefer and Close, 1990). It is important to limit the dissection within the region bounded by the middle turbinate, the orbit and the ethmoid roof. During the frontal recess dissection, infundibulotomy is performed by resecting the uncinate process so that the ethmoid bulla is apparent. The bulla should stay intact to protect the anterior ethmoid artery and ethmoid dome while acting as a landmark for the frontal sinus opening (Loury, 1993). Next is the removal of the anterior wall of the agger nasi cells which allows access to the frontal recess and ostium. To obtain a direct visualization of the frontal recess, bone with overlying mucosa and any soft tissue obstructing the frontal recess are removed. If there is inflammation or purulent infection in the frontal recess, cultures should be obtained and the surgery halted until the inflammation or infection has subsided. The ostium is anterior to the anterior ethmoid artery and lateral to the insertion of the middle turbinate. It is usually medial to the openings from the supraorbital ethmoid and agger nasi cells. If it cannot be seen, the recess is gently probed with a suction cannula which would fall into the sinus (Metson, 1992). Metson suggested that if the ostium is closed or stenotic (less than 2 mm), it can still be enlarged by removing the anterior bone while not damaging the ostial mucosa circumferentially, which could lead to restenosis. The opening may be stented with a 4-mm Silastic drainage catheter into the frontal sinus and sutured to the lateral nasal wall or the septum for 1 to 8 weeks.

The sphenoid sinus is to be opened for treating sinusitis or mucocele. When the middle turbinate limits the accessibility of the sphenoid sinus, it may be removed. The anterior wall of the sphenoid sinus should be seen after opening of the posterior ethmoid sinuses and removing any polypoid tissue. The sphenoid ostium is enlarged inferiorly, laterally, and medially while avoiding damaging the sinus roof that may result in cerebrospinal fluid (CSF) leak. Inferior dissection is limited to avoid injury to the sphenoipalatine artery. At the end of the surgery, 40 mg of methylprednisone is placed in the surgical site and the anterior surface of the inferior turbinate while antibiotic steroid ointment is applied in the ostiomeatal complex.

5.1.2.2 Complications of Surgery

Although the use of endoscopes significantly reduces morbidity due to improved visualization of the sinus anatomy and disease, there are still some complications. Surgical complications include cerebrospinal fluid (CSF) rhinorrhea, blindness, intraoperative orbital hematoma, and intracranial infections. The complications are influenced by the anatomic complexities, the extent of the disease, the presence of a prior surgery, and the experience of the surgeon. There are various precautions that the surgeon can take to prevent complications. They include being familiar with the anatomical landmarks, having intensive training, and being strict about hemostasis to maintain good visualization.

It is important for the surgeon to be aware of the high-risk areas where complications are likely to occur. Locating crucial landmarks such as the nasofrontal duct, dome of the ethmoid, anterior ethmoid artery, and sphenoid anterior wall, is essential. According to Ohnishi *et al.* (1993), there are five high-risk areas just in the ethmoid sinus alone: the lamina papyracea, the roof of the ethmoid sinus near the anterior ethmoid artery, the lateral lamella of the cribriform plate, the ethmoid roof near the posterior ethmoid artery, and the area between the sphenoid and posterior ethmoid sinuses. The lamina papyracea is the thin lateral wall of the ethmoid sinus that convexes into the surgical site. Simultaneous eye palpation and endoscopic exam of the lateral wall allows immediate detection of the entrance into the lamina papyracea. Its violation may cause ophthalmic muscle injury, intraorbital hemorrhage or infections, and even blindness. There are protrusions of the bony wall that contains the anterior ethmoid artery, especially at the roof of the anterior ethmoid sinus which is slightly more yellowish and particularly sensitive to pain. Its injury may lead to orbital hematoma in the lateral wall and CSF leakage or intracranial infections in the medial wall. The lateral lamella of the cribriform plate often bulges into the sinus cavity. Its injury would cause CSF leakage and also meningitis. Although the bony canal of the posterior ethmoid artery is thicker, sometimes the artery may travel beneath the bony plate. Its injury would cause more bleeding than if the damage is done to the anterior ethmoid artery. Between the ethmoid and sphenoid sinuses, the optic nerve is obscured and the internal carotid artery within the cavernous sinus bulges into the sinus cavity. Injury to either the nerve or the artery would lead to significant damage.

There are several ways that the surgeons could receive training to better equip themselves for the surgery. From gross and gross-sectioned anatomical studies, data on the distance between important structures of the medial orbital wall are available (Kirschner *et al.*, 1961 and Rontal *et al.*, 1979). They may be used to quantify the location of the landmarks and avoid damages to important structures. Also, to avoid complications, careful examination of the coronal CT scans should be used to determine the level of the protrusion. In addition, to minimize the risks, the general standard of training in sinus surgery should be improved with special training of otolaryngologists. Experience in traditional sinus surgery is important since excessive bleeding may require an alternative to the endoscopic surgery. Early recognition and treatment of the complications are essential in minimizing disability or even to prevent death.

It is crucial to visualize the important structures, such as the optic nerve and the internal carotid artery behind the bony walls during the surgery. Bleeding often occurs in acutely inflamed areas with extensive nasal polyps and with high blood pressure. Good hemostasis is necessary for clear vision of the structures during the surgery. Minimizing inflammation with antibiotics and stabilizing the mucosa with steroids should be done prior to the surgery. Blood pressure may be reduced by the anesthetist. Local anesthesia seems to minimize blood loss during the operation (Kennedy, 1992). Trauma to the turbinates and the vascular mucosa of the lateral nasal wall should be minimized during the surgery to avoid excessive bleeding. If significant bleeding inhibits good visualization, the surgery should be stopped.

Minor complications consist of synechiae, persistent or recurrent polyposis, bleeding, ecchymosis, maxillary ostium stenosis, dacryocystorhinitis, severe ear pain or headache. Synechiae between the middle turbinate and lateral nasal wall is often found during postoperative follow-up (Kennedy, 1985 and Schaefer *et al.*, 1989). This may cause obstruction of the ostiomeatal complex that can lead to sinusitis. Synechiae may be prevented by using a spacer between the middle turbinate and the lateral wall, by preserving the mucous membrane on the lateral surface of the middle turbinate, by removing the agger nasi cells and the mucosa membrane at the anterior attachment of the middle turbinate, by removing loose fragments of bone and mucous membrane, and by frequent postoperative cleansing (Schaefer *et al.*, 1989). Post-operative use of antibiotic-containing ointment would also help to prevent synechiae. The adhesions can be lysed, with significant adhesions requiring nasal splints and oral steroids. Bleeding may require packing the nose for 1 to 2 hours with cottonoid pledges soaked in a 4 % cocaine solution. If orbital fat is exposed, eye massage can be used to prevent intraorbital pressure increase and ecchymosis. To prevent ostia restenosis, intrasinus polypoid disease should be removed as completely as possible. Dacryocystorhinitis patients may require treatment by an ophthalmologist.

Although there are potential complications even with well-trained surgeons, the incidence of complications is very low. According to Lawson (1991), it is possible to reduce the complications of "total" ethmoidectomy to 1.1%. According to Stammberger (1986), more than 2,500 endoscopic ethmoid operations were performed without any serious complications. Wigand reported intraoperative CSF leaks in 2% of the patients (Wigand, 1981). Even so, patients who undergo endoscopic sinus surgery should be properly selected.

5.1.3 Post-Operation

5.1.3.1 Recovery

Seven to ten days after the surgery, the cavities are cleaned under endoscopic control to remove ointment, crust, blood clots, granulation tissue, or adhesions. Blood and mucus are aspirated from the nasal cavity, the maxillary, and the frontal sinuses. Saline nasal irrigation and cleaning are done to avoid postoperative osteitis and adhesions or recurrent ostiomeatal obstruction. Instillation of an antibiotic-cortisone ointment may be given for at least 10 days to patients without mycotic infection. Those with mycotic infection require antimycotic instillation. Most of the time, the use of topical steroids is continued and tapered. Depending on the extent of the disease and the endoscopic examination, other medication may be used. Evidences of mucosal hypertrophy, polyposis, inflammation, discharge, adhesions, crusting, and stenosis of maxillary or frontal sinus ostia are determined and treated. The patient returns for follow-up and cleaning in 4 to 6 weeks when the normal epithelialization occur. The capacity for the paranasal sinus mucous membrane and damaged cilia to regenerate is very high.

5.1.3.2 Success Rate

The success rate of surgery is influenced by many factors, such as the cause of the sinus disease, other disease involved, past surgery, and what was carried out in the surgery. In a study by Lazar *et al.* (1993) consisting of 773 patients who had received functional endonasal sinus surgery (FESS), 83% were successfully treated for chronic or recurrent sinusitis. The surgery was successful in relieving symptoms immediately in the patients with anatomical variations or suppurative infection but unsuccessful in some patients with polyps or hyperplastic sinusitis. Most patients whose symptoms persisted after the surgery had reactive airway disease, or other systematic disease, or had previous sinus surgery. The effect of nasal polypectomy varies widely, depending on the severity and its past treatment, the extent of the sinusitis, presence of other systemic factors, and the performance of the sinus surgery.

From a 4-year follow-up, the success rate of preventing recurrent symptomatic polyposis, adhesions between the middle turbinate and the lateral wall, or maxillary ostium restenosis is 91% (Schaitkin *et al.*, 1993). The latter two become symptomatic within the first 6 months of surgery while recurring polyps may occur up to 3 years after the surgery (Lawson, 1991).

Frontal sinus surgery provides complete relief of frontal discomfort within 1 week of the operation. The healed ostia stabilize with a diameter of 2 to 4 mm and remains patent more than 24 months post-operatively (Metson, 1992). According to Loury (1993), it is more difficult to maintain the frontal sinus opening due to residual ethmoidal disease in the area of the frontal recess since aggressive dissection in the region is hindered by the small space and the proximity of the anterior ethmoid artery, skull base, and the lamina papyracea. However, post surgical scarring of the nasofrontal recess may not cause obstruction to the frontal sinus 5 to 15 years after the surgery (Hardt and Montgomery, 1976).

The success rate for ethmoidectomy is difficult to compare since there is a lack of information of the extent of the disease (focal versus diffuse), type of disease (infectious versus polyps), and the involvement of systemic diseases (asthma, allergies, aspirin sensitivity) in the reports. Also, adjunctive medical treatment may not have been taken into account. Even the definition of success varies widely, from symptomatic relief of airway obstruction or rhinorrhea to the absence of the disease with an endoscopic exam. Furthermore, prolonged follow-up of patients three years post-operatively, is scarce. Incomplete ethmoid sinus dissection, often in the anterior region, has been a major reason for revision. Therefore, it is important to completely remove the ethmoid cells, especially in chronic hyperplastic rhinosinusitis which is often a diffuse process.

Recurrence of the initial condition or surgical sequelae such as adhesions and ostium stenosis may cause obstruction that requires further surgery to prevent sinusitis. Endoscopic examination is crucial in identifying persistent disease before it becomes symptomatic so that it could be treated medically and by local debridement. Computed tomography scan is used to identify the sites of recurrence and to determine the need for further surgical intervention.

5.1.3.3 Effects on Speech

The surgery may be successful in relieving symptoms immediately in patients. However, the patients and/or their family and friends have often noticed a change in the patient's speech post-operatively, according to informal comments by the patients to the surgeon. It is likely that anatomical changes such as opening blocked sinuses and removing the turbinates would affect the acoustic signal. Although patients have often noticed a change in their speech after surgery, the exact change has not been determined by formally trained phoneticians nor quantified by speech analysis. A comprehensive review of clinical, speech, and acoustic literature has uncovered no study that has been done on the effect of endonasal sinus surgery on speech and only one study on the acoustic effects of opening and closing the maxillary sinus ostia with epinephrine in subjects without nasal or paranasal disorders (Masuda, 1992). Masuda found that varying the maxillary sinus opening has a significant effect on speech spectra, introducing a zero between 1 to 2 kHz in the nasal transfer function.

5.2 Acoustic Analysis

Speech changes were quantified through detailed spectral analysis of nasal sounds recorded by patients pre- and post-operatively. The observed changes in the speech were related to the

Table 5.1: Patient history of the subjects whose speech was analyzed. Their sex (M, F), age, and symptoms of sinusitis prior to surgery are listed.

Subj.	Age	Symptomatic Illness
M1	42	Bilateral facial pressure at maxillary and frontal sinuses Almost daily headaches with post-nasal drainage Environmental allergies requiring immunotherapy (5 years) Chronic sinusitis at least 5 years, refractory to medication
M2	43	Bilateral constant nasal airway obstruction for many years Severe frontal headaches during plane descent Chronic sinusitis; refractory to medication
M3	42	Bilateral nasal congestion with almost no airway Loud snoring for 3 years; worse for past 1.5 years No sense of smell for more than 1 year Chronic sinusitis; refractory to medical therapy for 4 years
F1	34	Frontal headaches and maxillary pressure with colds Nasal airway obstruction when lying down; post nasal drip Chronic sinusitis for past five years Severe symptoms for past 3 months; refractory to medication
F2	57	Facial pain initially in the maxillary region to entire face Bilateral, alternating nasal airway obstruction Symptoms worsened when lying down Numerous environmental allergies; post nasal drip Chronic sinus infection caused by colds or flying > 4 sinusitis episodes over the past year refractory to medication.

anatomical variations through modeling of the nasal cavities according to information from CT scans obtained before the surgery and the surgical notes.

5.2.1 Speakers

Subjects were recruited from patients who were going through endonasal sinus surgery by Dr. Ralph Metson at the Massachusetts Eye and Ear Infirmary. Only patients with no previous nasal operation and no evidence of velopharyngeal incompetence were studied. The speech of three male and two female native English speakers was analyzed in depth. Table 5.1 shows the age of the patients and a description of the symptoms each patient faced before the surgery. The patients were between ages 34 to 57 years old. All of them had chronic sinusitis that was refractory to medication. The more common symptoms were headaches and airway obstructions. Other symptoms include facial pressure, post nasal drip, snoring, and lost sense of smell. For some of the patients, the symptoms occurred more often with colds or flying and worsened when lying down.

Table 5.2 indicates the abnormal findings from the patients' nasal endoscopy and/or CT scan done prior to the surgery. The last

Table 5.2: Abnormal findings from the patients' nasal endoscopy and/or CT scans done prior to the surgery. The CT staging was based on a method developed by the surgeon.

Subj.	Septal deviation	Abnormal bony growth	Polyps location	Fluid level (pus)	CT Stage
M1	right posterior	left agger nasi	FS ES	left MS cyst ES	II
M2	left anterior	bilateral IT	ES	MS cyst	II
M3	left		cavity all sinuses		II
F1	right posterior superior	bilateral MT	cavity, ES	all sinuses	IV
F2	left middle right post.	Haller's cells			I

ES: ethmoid sinus FS: frontal sinus MS: maxillary sinus IT: inferior turbinate
MT: middle turbinate

column shows the staging developed and used by Metson based on the CT scan: (0) normal with mucosal thickening < 2mm, (I) unilateral disease/anatomic abnormality, (II) bilateral ethmoid/maxillary disease, (III) bilateral disease with frontal/sphenoid involvement, and (IV) pansinusitis. Subject M1 had a bulging agger nasi, which is the ethmoid cell extension in front of the neck of the middle turbinate, nearly obstructing the anterior middle meatus region and partially obstructing the frontal recess region, and a right septal deviation that precluded adequate visualization of the middle meatus. Both may have induced the chronic sinusitis and polyps formation by causing close contact of adjacent mucosa. M2 had septal deviation and inferior turbinate hypertrophy which caused nasal airway obstruction. The bilateral OMC was also occluded by maxillary cyst leading to sinusitis and reflecting in the frontal headaches. M3 had multiple large obstructing polyps which induced chronic sinusitis, loud nasal breathing, and the poor sense of smell. All three male subjects with ethmoid/maxillary diseases were categorized in stage II. Subject F1 had right septal deviation which precluded endoscopic visualization of the middle turbinate (MT) and large obstructing nasal polyps with enlarged middle turbinates that induced pansinusitis. According to the CT staging, she was in stage

Table 5.3: Additional procedure performed during the surgery other than the general procedure of endonasal sinus surgery described in the text.

Subj.	Septoplasty	Resection	Polyp removal	Pus evacuation	SO diameter
M1	right	left agger nasi	ES, FR	left MS	right (4 mm)
M2	left	bilateral IT (anterior 2/3)	ES, right SO		right (6 mm)
M3	left	right MT	MM, FR all sinuses right SO		right (4 mm) left (8 mm)
F1		bilateral MT	FR, MM		right (5 mm) left (4 mm)
F2	right				

IT: inferior turbinate MT: middle turbinate ES: ethmoid sinus FR: frontal recess
SO: sphenoid ostium MM: middle meatus MS: maxillary sinus

IV. F2 had septal deviations on both sides due to a nose fracture which was never repaired at the age of 17. Due to the fracture, the left middle turbinate was not well visualized and the right ethmoid region was difficult to access. The Haller's Cells probably led to her maxillary pain and infections. Her anatomic abnormality placed her in stage I.

All of the subjects went through endonasal sinus surgery, which in general followed the procedure described in Sec. 5.1.2.1. They had bilateral uncinated process incised and removed, ethmoidectomy which involved widely opening the ethmoid bulla and basilar lamella and the removal of thickened mucosa and bony septa of the anterior and posterior ethmoid sinuses. The maxillary ostium was enlarged to 10 mm in diameter and the frontal recess was cleaned and opened. Table 5.3 shows additional procedures involved for each patient. All of the patients, except F1, had septoplasty which included hemitransfixion incision, mucoperichondrial flap elevation, an approximately 3x4 cm^2 piece of deviated quadrangular cartilage freed from the surrounding soft tissue and removed, the mucoperiosteom elevated from the deviated bony septum and removed, and the hemitransfixion incision closed. For M2, a 3 mm wide strip of deflected caudal septum was also resected. All of the patients except F2 had resection of bony tissues, e.g. agger nasi, middle turbinate or inferior turbinate, and polyp removal which involved clearing the mass due to localized swelling of the mucosa. The right middle turbinate of M3 was resected due to extensive polypoid degeneration even though it was not detected before the surgery. The sphenoid ostium was opened to a diameter of 4 mm to 8 mm for all of the patients except F2.

Most of the patients recovered fairly well from the surgery. For M1, 10 days after the surgery, the middle meatus was only partially obstructed by an uncinate process, the maxillary pus was evacuated, and the nasal cavity was widely patent and cleaned of crusts on the left; on the right, MT was somewhat lateralized and there was crusting at the septoplasty site. Two months later, the patient had left nasal congestion and small polyps medial to the left MT in the superior meatus, which later disappeared. For M2, seven days after the surgery, the crust was cleared from the IT but he was not breathing too well through his nose. One month later, the nasal region remained swollen but was widely patent and breathing was much improved. The IT resection healed well but the MT's were edematous and lateralized. For M3, in 7 days, he was recovering excellently and the debris was cleaned from the sinuses. In two months, the nasal airway was good with no polyps present. For F1, in 10 days, the sinuses appeared widely patent but there was cyst or polypoid disease at the maxillary ostia. In a month and a half, there was normal appearing mucosa along the ethmoid roofs free of polypoid disease with much improved nasal breathing. For F2, in 10 days, the septal suture was still in place, the MT's were somewhat lateral, and the maxillary ostia were visible and patent. In a month, the septum was at the midline with the nose and sinuses patent.

5.2.2 Method

Recordings were made during the patient's clinic visit about one week prior to the surgery. The patient's speech was also recorded about a week after the surgery and again more than a month after the surgery at the end of the clinical visit. In each recording session, the subject was asked to say the same utterances as the English speakers mentioned in Sec. 3.3.1 in the form 'CVC', where V is one of six vowels and C's are either nasal consonants or stop consonants, embedded in a carrier phrase. The task was done three times within one session at a comfortable speaking level in order to assess the consistency of the data. An Electro-Voice model 635A omnidirectional dynamic microphone vertically placed at six inches from the subject's lips was used to transduce the acoustic signal, which was recorded on a PMD model 201 portable recorder.

Since the recordings were done in the clinic, they took place in an audiology booth or a room with carpeted floors and sufficient wall and ceiling padding to reduce reverberation. Detailed acoustic analysis was performed using the same method as in Sec. 3.1.

5.2.3 Results and Discussion

Any spectral differences of nasal consonants between before and after the surgery were observed and quantified. Nasal and non-nasal vowels from before and after the surgery were also quantified by using the acoustic correlates A1-P1 and A1-P0. It was expected that the vowels in the stop context would not change much with surgery but those in the nasal context would show alterations due to changes of the nasal cavity anatomy from the surgery.

5.2.3.1 Nasal Consonants

Figure 5.1a and Fig. 5.1b show the spectra obtained from the middle of /m/ spoken by subject M1 before the surgery and after the surgery, respectively. A qualitative comparison of the spectra in /m/ and /n/ from before the surgery showed generally a raised first formant amplitude, A1, and a lowered peak amplitude around 1 kHz, $P1_n$, after the surgery. Both A1 and $P1_n$ were measured from the harmonic with the maximum amplitude in the vicinity of the peaks. The increase in A1 after the surgery may be due to the removal of tissue, e.g. middle turbinates and polyps, which would decrease the total surface area and also eliminate constrictions in the nasal airway after the surgery. As the total surface area decreases and the constriction is eliminated, there would be less loss and therefore, the first formant bandwidth would be narrowed while its amplitude is raised. The peak $P1_n$ is at the second natural frequency of the nasal consonant system. Its frequency can be calculated roughly from the total length of the pharyngeal, oral, and nasal cavities, and it is expected to occur around 1 kHz. The amplitude of this peak may be lowered by the zero in /m/ which occurs in the vicinity of 1 kHz due to the length between the mouth closure and the velopharyngeal opening. In addition, greater coupling to the sphenoid sinus from the surgery may introduce a zero around 1 kHz (see Sec. 2.2.2) to further lower $P1_n$. The removal of tissues, such as the turbinates, polyps, and ethmoid sinus septa, would enlarge the nasal passage volume.

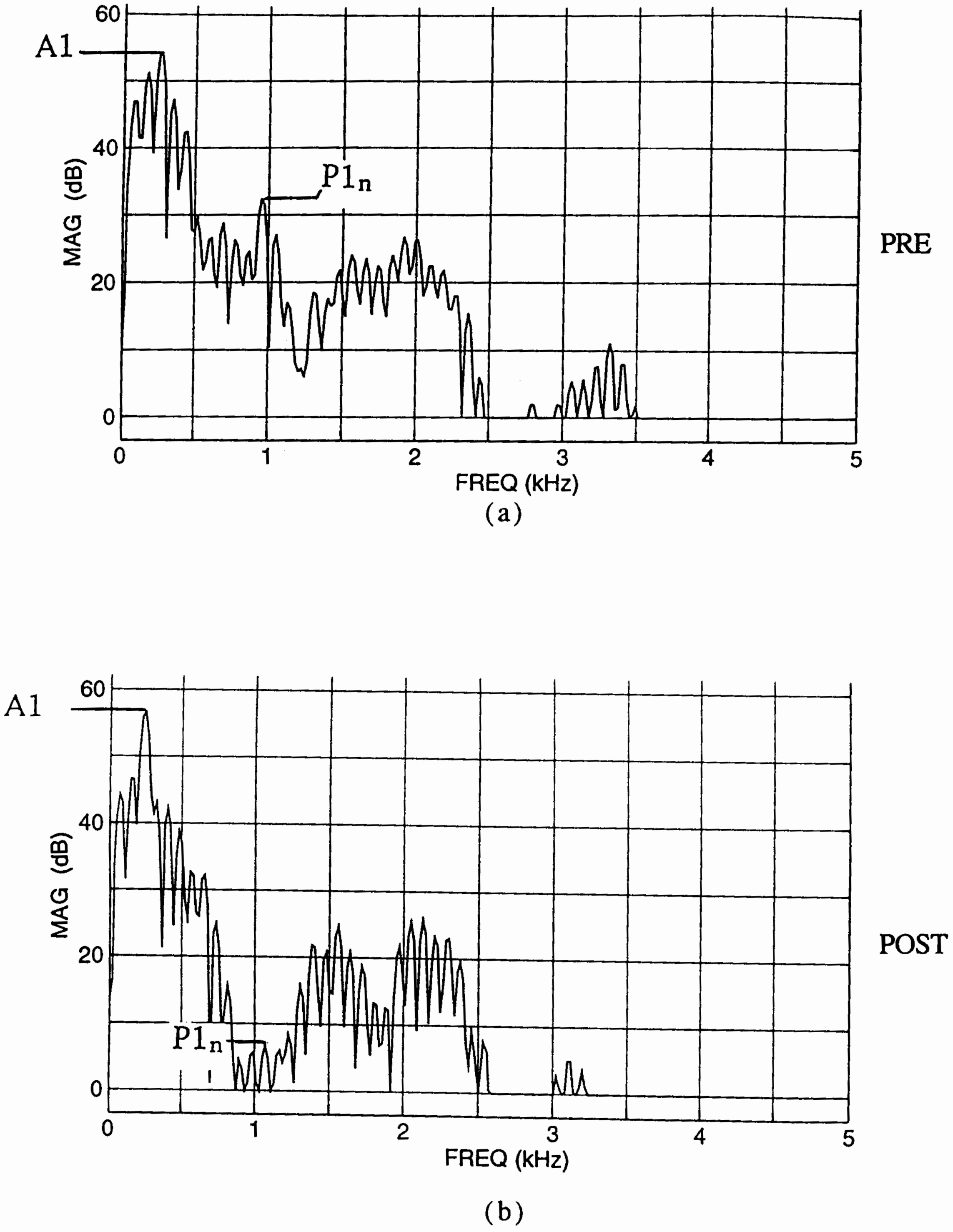


Figure 5.1: The spectra obtained from the middle of /m/ spoken by subject M1 (a) before the surgery and (b) after the surgery are presented. A1 is the amplitude of the first formant and $P1_n$ is the amplitude of the nasal peak around 1 kHz.

According to the perturbation theory (Stevens, in preparation), increased volume in that region would shift the nasal pole to a lower frequency than the frequency where the nasal peak would occur if the volume did not change so that measuring $P1_n$ at the original frequency would yield a lower amplitude. Since A1 was raised and $P1_n$ was lowered after the surgery, $A1-P1_n$ was used as an acoustic correlate to quantify nasalization in nasal consonants. It was expected that $A1-P1_n$ would increase after the surgery.

There was little change in spectrum throughout the nasal consonant. Measurements were made from the middle of /m/ and /n/, those preceding and following the vowel, in different repetitions and vowel contexts. Since little difference was observed for the various vowel contexts, all of the tokens were analyzed together for each of the five patients. Also, the spectral changes from pre- and post-surgery had similar characteristics for /m/ and /n/. Figure 5.2 shows the average of $A1-P1_n$ measured in /m/ and /n/ spoken before the surgery, one week after the surgery, and more than one month after the surgery. The symbol '+' indicates $p < 0.01$ between the values from before and after the surgery. The standard error bars are shown. For each of the patients, the average $A1-P1_n$ increased significantly after the surgery. For four of the patients, the average $A1-P1_n$ was larger more than one month after the surgery than just one week after the surgery, indicating that over time, healing causes additional change in $A1-P1_n$. For M3, $A1-P1_n$ did not vary much between measures made one week and two months later may be because he recovered even one week after the surgery. For the males, the average $A1-P1_n$ increased between 11 dB and 23 dB; for females, it increased between 9 dB and 11 dB.

To determine if the significant increase of $A1-P1_n$ occurs for normal speakers recorded at different times, two males and one female with normal nasal cavity without surgery were recorded on three different days saying the same utterances as the patients. $A1-P1_n$ was measured in the middle of /m/ and /n/ for all of the repetitions and vowel contexts. Figure 5.3 shows the average values for the three sessions arranged in ascending order for each speaker to allow comparison between pairs. The symbol '+' indicates $p < 0.01$ and '*' indicates $p < 0.05$ between the lowest average $A1-P1_n$ (session 1) and that of the other two sessions. The average $A1-P1_n$ difference among sessions for males was between 1 dB and 3 dB and for the female was 2 dB which were much less statistically significant than the differences for the patients.

In comparing the absolute values of $A1-P1_n$ for the male speakers in Fig. 5.2, the averages before the surgery, around 20 dB,

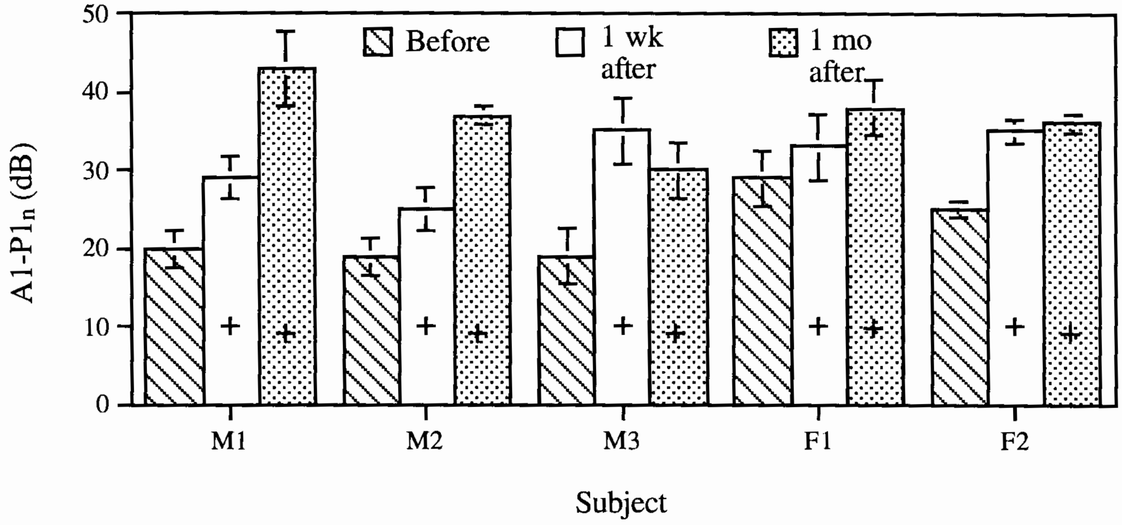


Figure 5.2: The average of A1-P1$_n$ measured in /m/ and /n/ across vowel contexts and repetitions spoken by patients before endonasal sinus surgery (stripped bars), one week after the surgery (white bars), and more than one month after the surgery (dotted bars). The symbol '+' indicates $p < 0.01$ between the values from before the surgery and those from after the surgery.

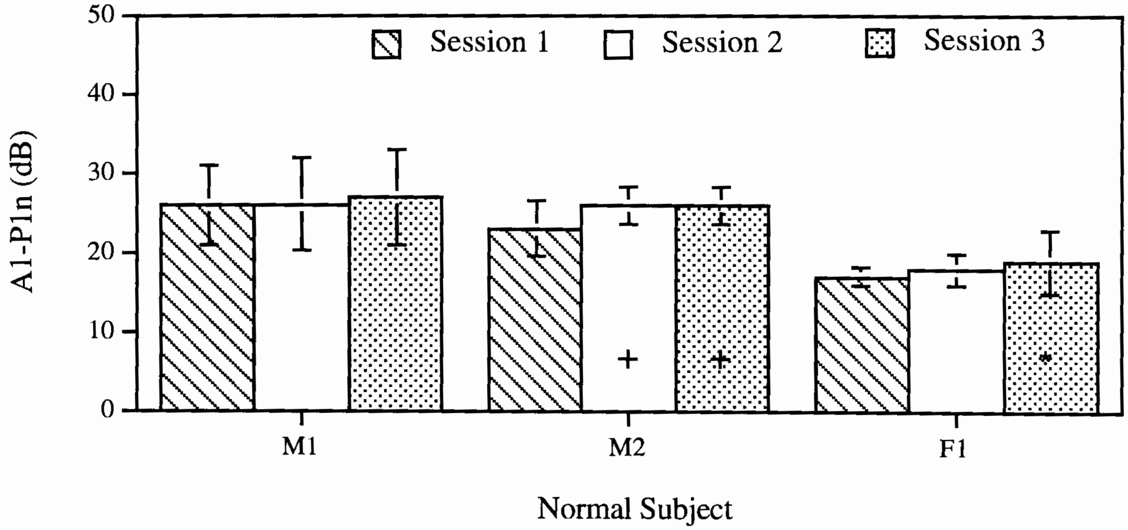


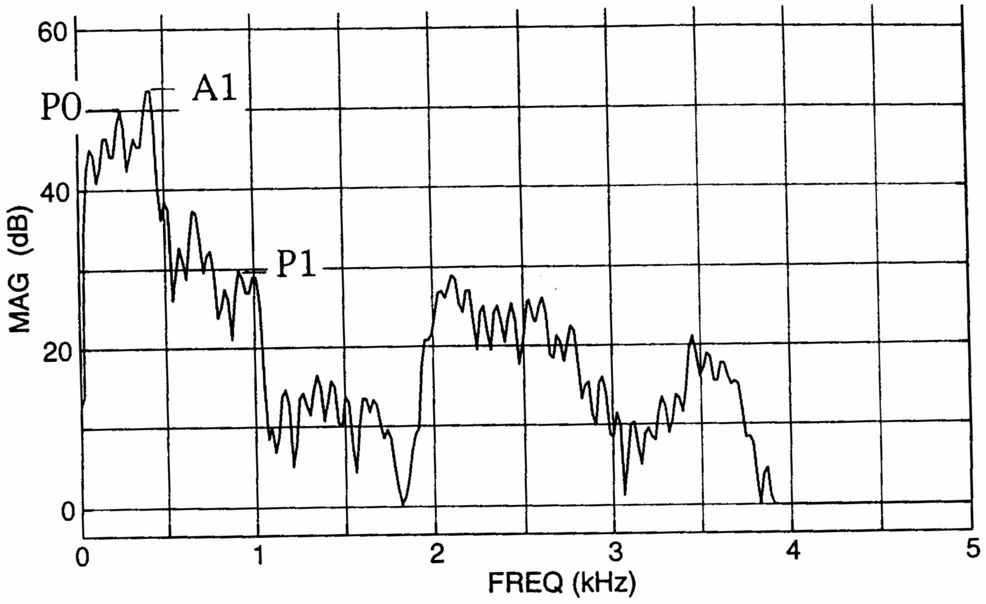
Figure 5.3: The average of A1-P1$_n$ measured in /m/ and /n/ across vowel contexts and repetitions spoken by normal speakers for the three sessions arranged in ascending order for each speaker. The symbol '+' indicates $p < 0.01$ and '' indicates $p < 0.05$ between the values of session 1 (stripped bars) with the lowest average A1-P1$_n$ and the values of session 2 (white bars) and session 3 (dotted bars).*

were smaller than those of normal male speakers in Fig. 5.3, around 25 dB, while the averages from after the surgery were over 30 dB. This indicates that before the surgery, the spectra of the nasal consonants had lower A1 due to greater losses than the speakers with normal anatomy while after the surgery, the spectra had higher A1 and lower $P1_n$ due to reduced losses and greater coupling to the sinuses with enlarged volume in the nasal passages. For the females, $A1-P1_n$ was greater before and after the surgery than that of the normal speaker. This may be due to speaker variation because of differences in voicing and anatomy.

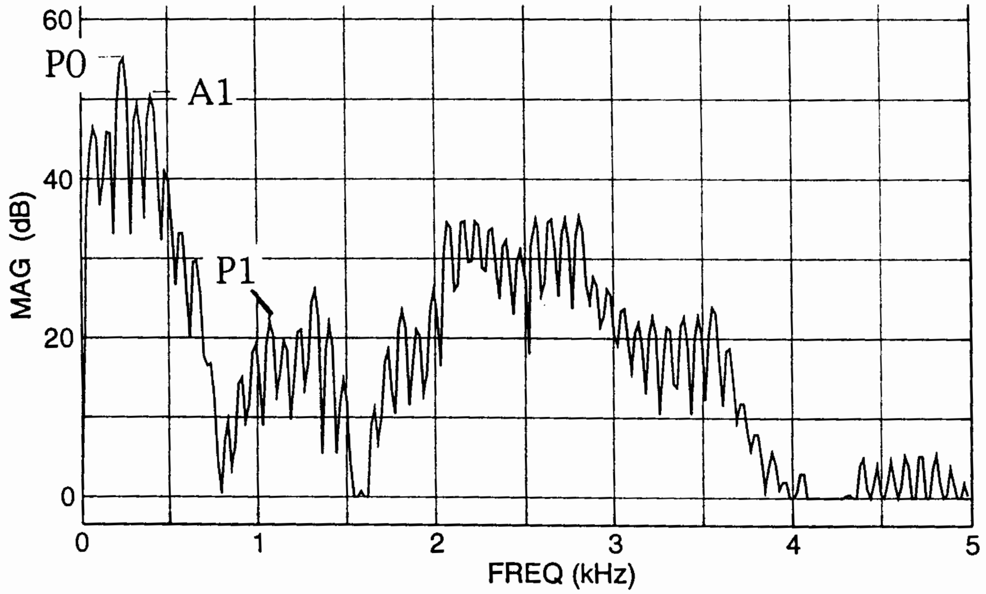
5.2.3.2 Nasal Vowels

Vowel nasalization of normal English and French speakers causes several spectral changes: lowering of the first formant amplitude A1 due to losses, raising of nasal peak P1 around 950 Hz due to coupling to the main nasal passage, and raising of nasal peak P0 around 250 Hz due to coupling to the paranasal sinuses and possibly increased open quotient of the glottal waveform (See Sec. 3.3.2). Figure 5.4a and Fig. 5.4b show the spectra obtained from a nasalized /æ/ by M1 spoken before the surgery and after the surgery, respectively. The amplitude of the first formant, A1, along with P1 and P0 are labeled. Table 5.4 shows the frequency of the nasal peaks, F_{P0} and F_{P1}, averaged across vowel types and the three recording sessions for each speaker since the differences among sessions were mainly not significant. These frequencies were found to be independent of vowel type for normal English and French speakers (See Sec. 3.3.2 and Sec. 3.4.2). The average F_{P0} for the male speakers who have gone through surgery was 249 Hz and for the females was 363 Hz; the average F_{P1} for the males was 863 Hz and for the females was 948 Hz. The frequencies for females were higher than males. This difference may be due to the higher fundamental frequency affecting the harmonic used for the peak amplitude. It may also be due to females having smaller anatomical dimensions, which would result in higher peak frequencies. The difference in F_{P0} among speakers of the same sex was statistically significant, which may be due to the different anatomy of the sinuses of each speaker. Only the difference in F_{P1} between M2 and M3 was statistically significant since F_{P1} is dependent mainly on the main nasal passageway which may be more similar among the speakers.

The acoustic correlates that have been used in analyzing the English and French speakers, A1-P1 and A1-P0, were applied to the vowels spoken by these patients. A1-P1 was used to analyze the



(a)



(b)

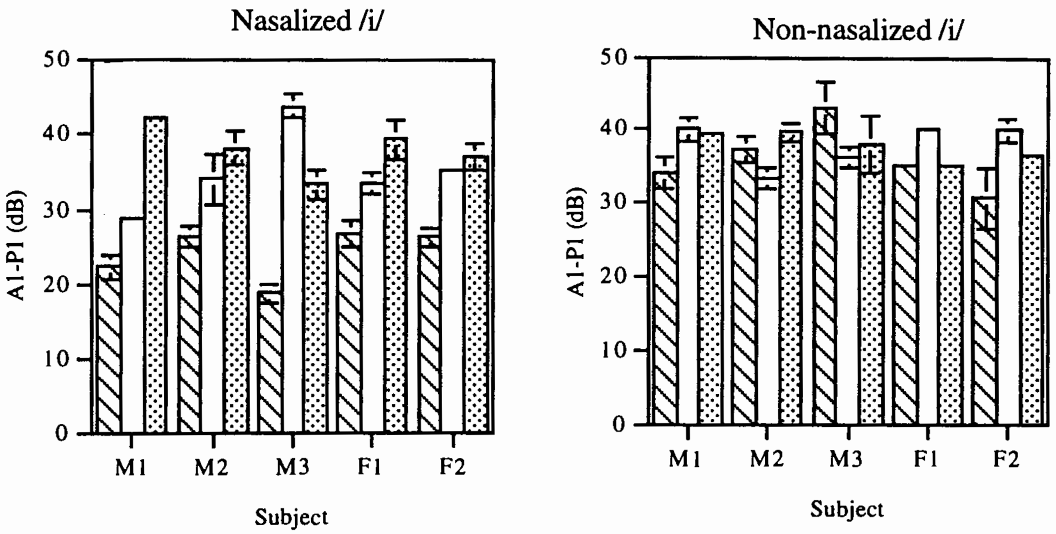
Figure 5.4: The spectra obtained from a nasalized /æ/ spoken by M1 (a) before the surgery and (b) after the surgery. The peak amplitude of the first formant, A1, the peak amplitude of nasal peak around 1 kHz, P1, and the nasal peak at low frequencies, P0 are labeled.

Table 5.4: Average frequency of the nasal peaks measured in nasalized vowels of five English speakers who have gone through endonasal sinus surgery. F_{P0} and F_{P1} are the frequency of the nasal peaks with amplitude P0 and P1, respectively.

Subject	F_{P0} (Hz)	F_{P1} (Hz)
M1	262	858
M2	268	833
M3	216	898
F1	382	924
F2	343	972

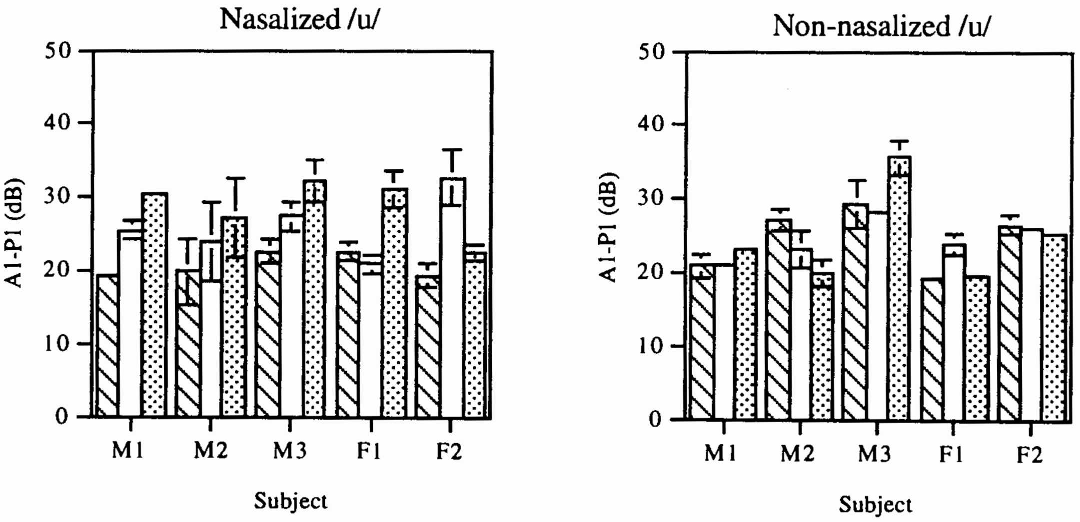
non-low vowels and A1-P0 was used for the non-high vowels. Measurements were made at the beginning and end of the vowels. Figure 5.5a and Fig. 5.6a show the averaged A1-P1 in dB for the nasalized high vowels /i/ and /u/, respectively, across repetitions and locations within the vowel. Generally, the average A1-P1 increased one week and one month after the surgery compared to before the surgery for all speakers. The change for /i/ more than one month after the surgery from before the surgery was from 12 dB to 21 dB for males and 10 dB to 13 dB for females while for /u/ the change was from 6 dB to 11 dB and 4 dB to 10 dB. Figure 5.5b and Fig. 5.6b show A1-P1 for the non-nasalized vowels /i/ and /u/, respectively. There was no clear trend of A1-P1 increasing after the surgery as in the nasal vowels, with a decrease of 5 dB to an increase of 5 dB for males and an increase of 0 dB to 6 dB for females in /i/. For /u/ there was a decrease of 8 dB to an increase of 6 dB for males and a decrease of 2 dB and an increase of 1 dB for females. The results suggest that the alterations in the nasal cavity greatly affected A1-P1 of nasalized high vowels.

Figure 5.7a and Fig. 5.8a show A1-P0 in dB for the nasalized low vowels, /æ/ and /ɑ/, respectively, across repetitions and locations within the vowel. In general, one month after the surgery the average A1-P0 decreased while there was no trend for one week after the surgery. For /æ/ the change of A1-P0 more than a month after the surgery from before the surgery ranged from an increase of 1 dB to a decrease of 12 dB for males and from a decrease of 6 dB to 12 dB for females. For /ɑ/ the decrease was between 2 dB and 10 dB for males and 4 dB to 5 dB for females. Figure 5.7b and Fig. 5.8b show A1-P0 for the non-nasalized low vowels, /æ/ and /ɑ/, respectively. For /æ/, the change in A1-P0 ranged from a decrease of 2 dB to 7 dB for males and a decrease of 0 dB to 4 dB for females.



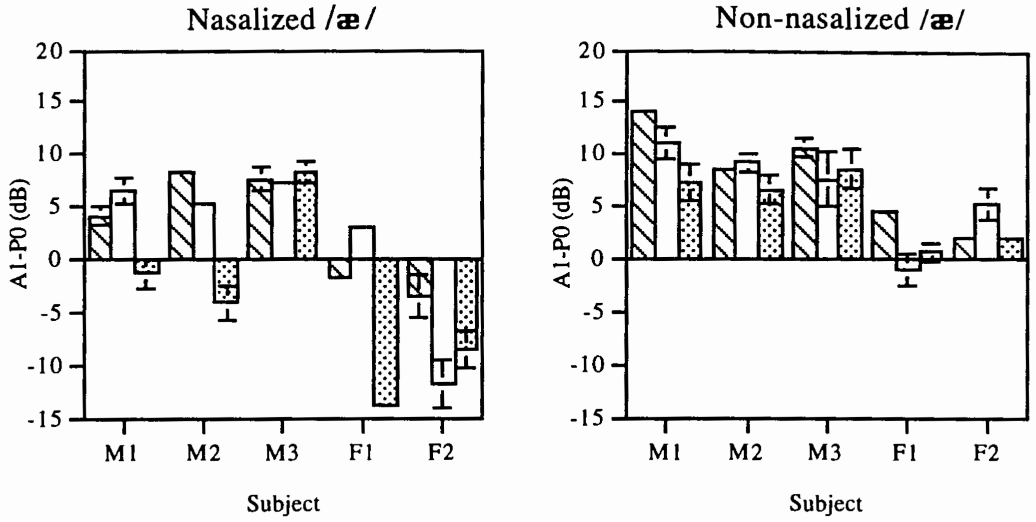
(a) (b)

Figure 5.5: (a) The correlate A1-P1 was measured at the beginning and the end of nasal high vowel /i/ with three repetitions spoken before the surgery (stripped bars), one week after the surgery (white bars), and more than one month after the surgery (dotted bars). (b) The correlate A1-P1 was also measured at the beginning and the end of non-nasalized /i/.



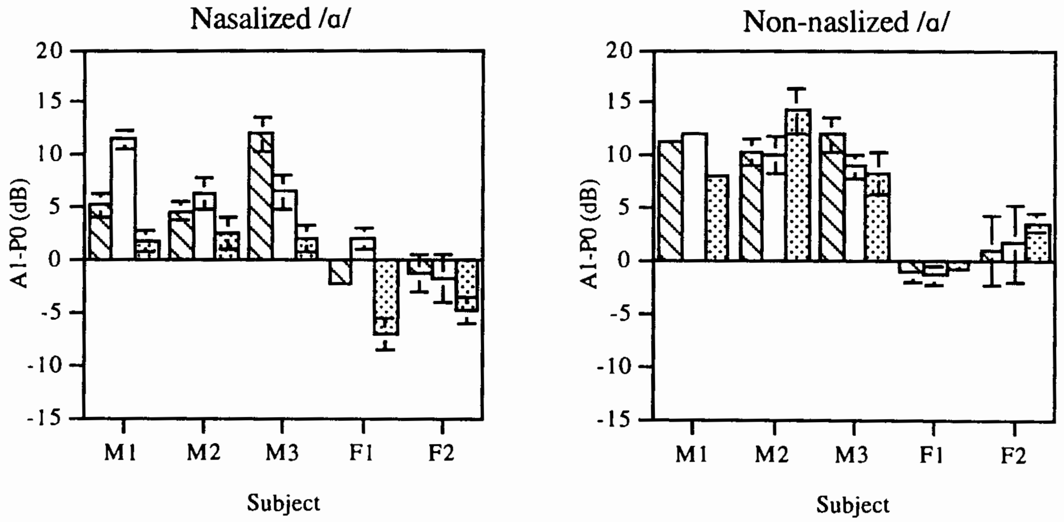
(a) (b)

Figure 5.6: (a) The correlate A1-P1 was measured at the beginning and the end of nasal high vowel /u/ with three repetitions spoken before the surgery (stripped bars), one week after the surgery (white bars), and more than one month after the surgery (dotted bars). (b) The correlate A1-P1 was also measured at the beginning and the end of non-nasalized /u/.



(a) (b)

Figure 5.7: (a) The correlate A1-P0 was measured at the beginning and the end of nasal low vowel /æ/ with three repetitions spoken before the surgery (stripped bars), one week after the surgery (white bars), and more than one month after the surgery (dotted bars). (b) The correlate A1-P0 was also measured at the beginning and the end of non-nasalized /æ/.



(a) (b)

Figure 5.8: (a) The correlate A1-P0 was measured at the beginning and the end of nasal low vowel /ɑ/ with three repetitions spoken before the surgery (stripped bars), one week after the surgery (white bars), and more than one month after the surgery (dotted bars). (b) The correlate A1-P0 was also measured at the beginning and the end of non-nasalized /ɑ/.

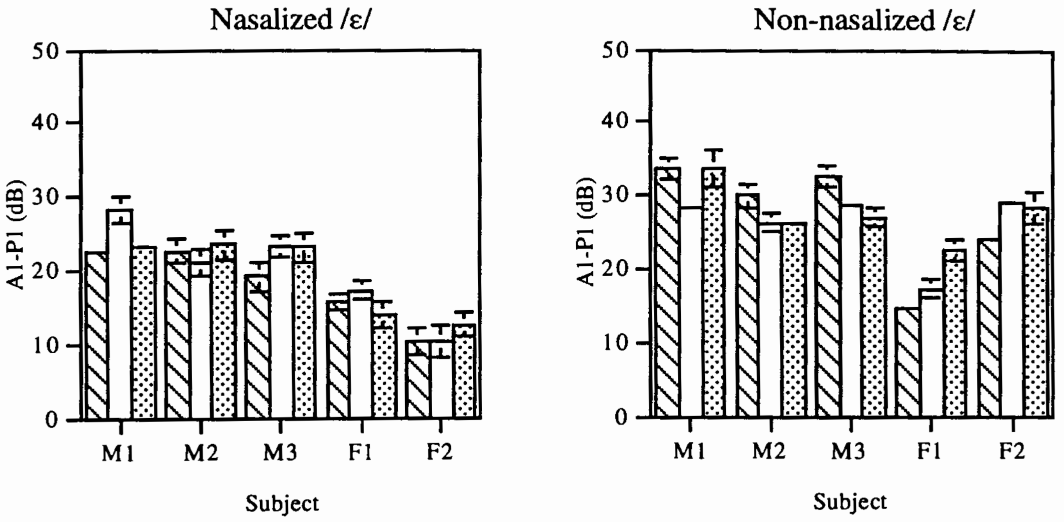
For the non-nasalized /ɑ/, the change in A1-P0 ranged from an increase of 5 dB to a decrease of 4 dB for males and an increase between 0 dB to 3 dB for the females. In general, A1-P0 decreased less for the non-nasalized vowels than for the nasalized vowels after surgery. The greater decrease of A1-P0 in the nasalized vowels compared to that of the non-nasalized vowels after the surgery also suggests that the alteration of the nasal cavity affected the acoustics of nasalized vowels.

Figures 5.9a and 5.9b show A1-P1 for the nasalized and non-nasalized /ɛ/, respectively. There was a small increase of 1 dB to 4 dB in the average A1-P1 a month after the surgery from before the surgery for males and a decrease of 2 dB to an increase of 2 dB for females. For the non-nasalized vowels of the males, the change ranged from a decrease of 0 dB to 5 dB while for the females the change ranged from an increase of 4 dB to 8 dB, which is actually a greater deviation than for the nasalized vowels. The results indicate that A1-P1 does not reflect the effect of the surgery on the nasalized vowels.

On the other hand, A1-P0, as shown in Figure 5.10a and Fig. 5.10b for the nasalized and non-nasalized /ɛ/, respectively indicates A1-P0 reflected the alterations in the nasal anatomy more consistently than A1-P1. For four of the speakers, there was a decrease in the average of A1-P0 of nasalized vowels more than a month after the surgery. There was an increase of 1 dB to a decrease of 9 dB for males and a decrease of 4 dB to 6 dB for females. The lowering of the average A1-P0 in the non-nasalized vowels ranged from 0 dB to 5 dB for males while A1-P0 increased by 6 dB or stayed the same after the surgery for the females.

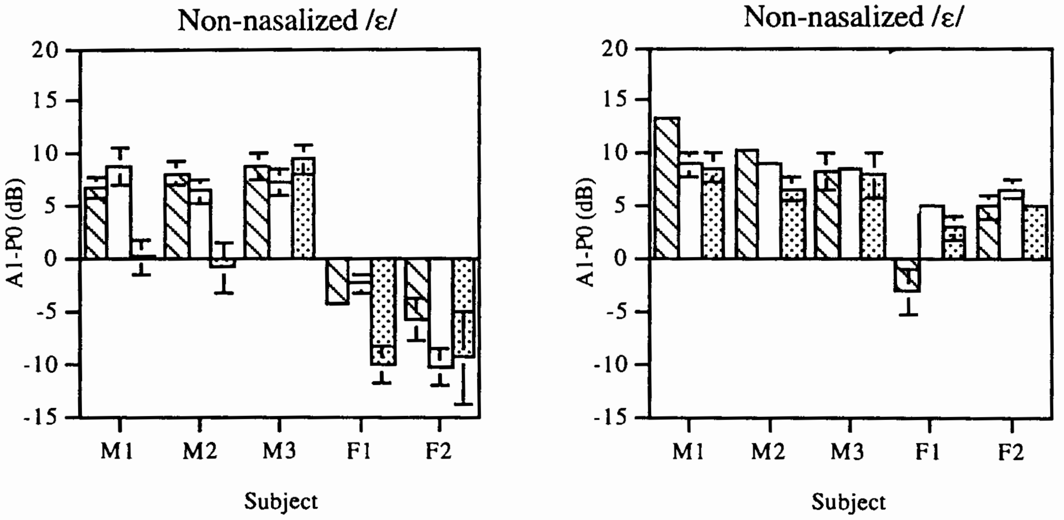
Figure 5.11a and Fig. 5.11b show A1-P1 for the nasalized and non-nasalized /ʌ/, respectively. The average A1-P1 in nasalized vowels decreased in three of the speakers but increased in two of the speakers more than a month after the surgery. There also was no trend in the non-nasalized vowels. Figure 5.12a and Fig. 5.12b show A1-P0 measured in nasalized and non-nasalized /ʌ/. The average A1-P0 in the nasal vowels either remained the same or decreased a month after the surgery by as much as 8 dB for males and 10 dB for females. A1-P0 in the non-nasalized vowel behaved similarly by decreasing as much as 6 dB for males and 4 dB for females; these were smaller changes than those of the nasal vowels. The results indicate that A1-P0 reflected the alterations in the nose more consistently than A1-P1, at least for the females.

From the observations of the individual vowels, A1-P1 is a better measure in revealing effects from the surgery for high vowels,



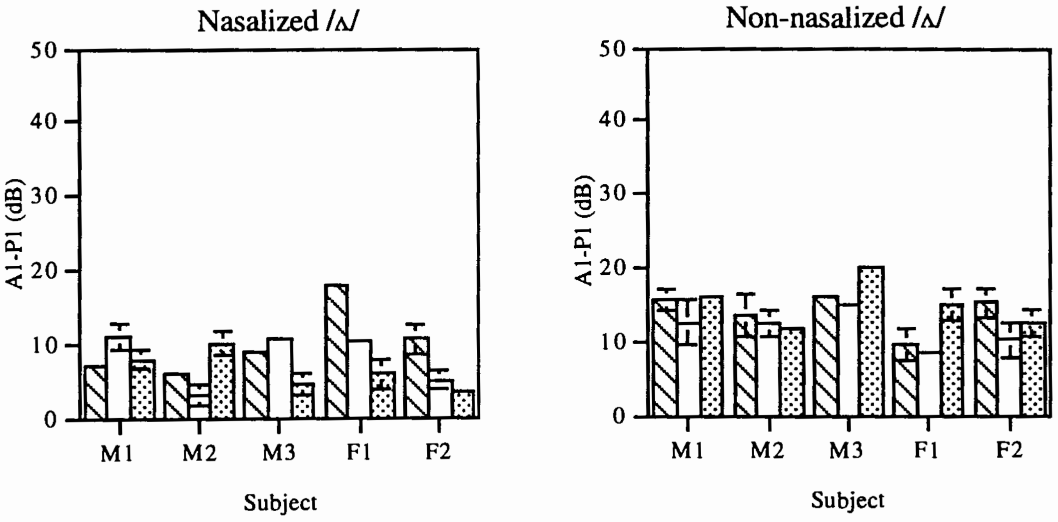
(a) (b)

Figure 5.9: (a) The correlate A1-P1 was measured at the beginning and the end of nasal mid vowel /ɛ/ with three repetitions spoken before the surgery (stripped bars), one week after the surgery (white bars), and more than one month after the surgery (dotted bars). (b) The correlate A1-P1 was also measured at the beginning and the end of non-nasalized/ɛ/.



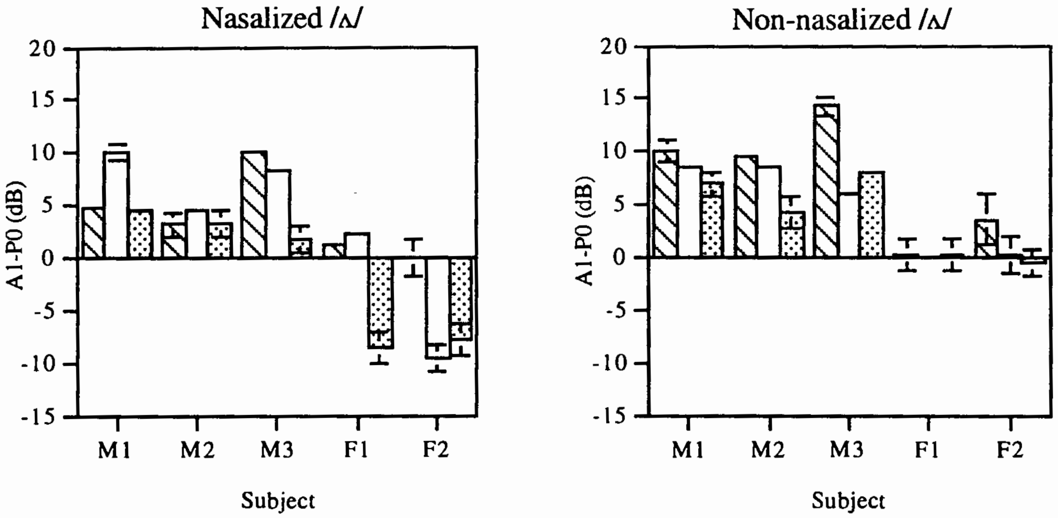
(a) (b)

Figure 5.10: (a) The correlate A1-P0 was measured at the beginning and the end of nasal mid vowel /ɛ/ with three repetitions spoken before the surgery (stripped bars), one week after the surgery (white bars), and more than one month after the surgery (dotted bars). (b) The correlate A1-P0 was also measured at the beginning and the end of non-nasalized /ɛ/.



(a) (b)

Figure 5.11: (a) The correlate A1-P1 was measured at the beginning and the end of nasal mid vowel /ʌ/ with three repetitions spoken before the surgery (stripped bars), one week after the surgery (white bars), and more than one month after the surgery (dotted bars). (b) The correlate A1-P1 was also measured at the beginning and the end of non-nasalized /ʌ/.



(a) (b)

Figure 5.12: (a) The correlate A1-P0 was measured at the beginning and the end of nasal mid vowel /ʌ/ with three repetitions spoken before the surgery (stripped bars), one week after the surgery (white bars), and more than one month after the surgery (dotted bars). (b) The correlate A1-P0 was also measured at the beginning and the end of non-nasalized /ʌ/.

/i/ and /u/, while A1-P0 is a better measure for the non-high vowels, /æ/, /ɑ/, /ɛ/, and /ʌ/. To determine if the changes were statistically significant, A1-P1 was analyzed for /i/ and /u/ averaged across repetitions and locations within the nasal vowel as shown in Figure 5.13a. The symbol '+' along the abscissa indicates $p < 0.01$ between A1-P1 from before the surgery and after the surgery. For all of the speakers, A1-P1 increased significantly more than one month after the surgery while for four of the speakers, A1-P1 increased significantly one week after the surgery. The changes were more consistent a month or more after the surgery than just a week after the surgery, presumably due to the healing process. The average A1-P1 more than one month after the surgery from before the surgery increased between 10 dB to 15 dB for males and 10 dB to 12 dB for females after the surgery. Figure 5.13b shows the values of A1-P1 from the corresponding non-nasalized /i/ and /u/ adjacent to stop consonants. None of the differences between A1-P1 from before the surgery and after the surgery was statistically significant. The change in the average A1-P1 ranged from a decrease of 2 dB to an increase of 3 dB for males, and for females it increased between 1 dB to 2 dB, much less than that of the nasalized vowels.

In comparing the average values of A1-P1, most of the speakers had an average A1-P1 in the nasal vowel of around 20 dB before the surgery which was lower than the average in the corresponding non-nasal vowel which was as large as 36 dB. On the other hand, the averages from after the surgery in the nasal vowels were either greater or about the same as those in the non-nasal vowels. This indicates that before the surgery, the spectra of the nasal vowels had lower A1 and/or higher P1 due to coupling to the nose than the non-nasal vowels while after the surgery, the spectra had higher A1 and/or lower P1. For these speakers with similar or greater values after the surgery in the nasal than in the non-nasal vowels, it may be that an over-correction of the anatomy was involved so that nasal vowels would be perceived as non-nasal vowels.

A1-P0 was analyzed for /æ/, /ɑ/, /ɛ/, and /ʌ/ across repetitions and locations within the nasal vowel to determine statistical significance, as shown in Fig. 5.14a. The symbol '+' along the abscissa indicates $p < 0.01$ between A1-P0 from before the surgery and after the surgery. For all of the speakers, the average A1-P0 decreased significantly more than a month after the surgery, but one week after the surgery did not show such a trend. The decrease in A1-P0 was between 4 dB and 6 dB for males and 6 dB to

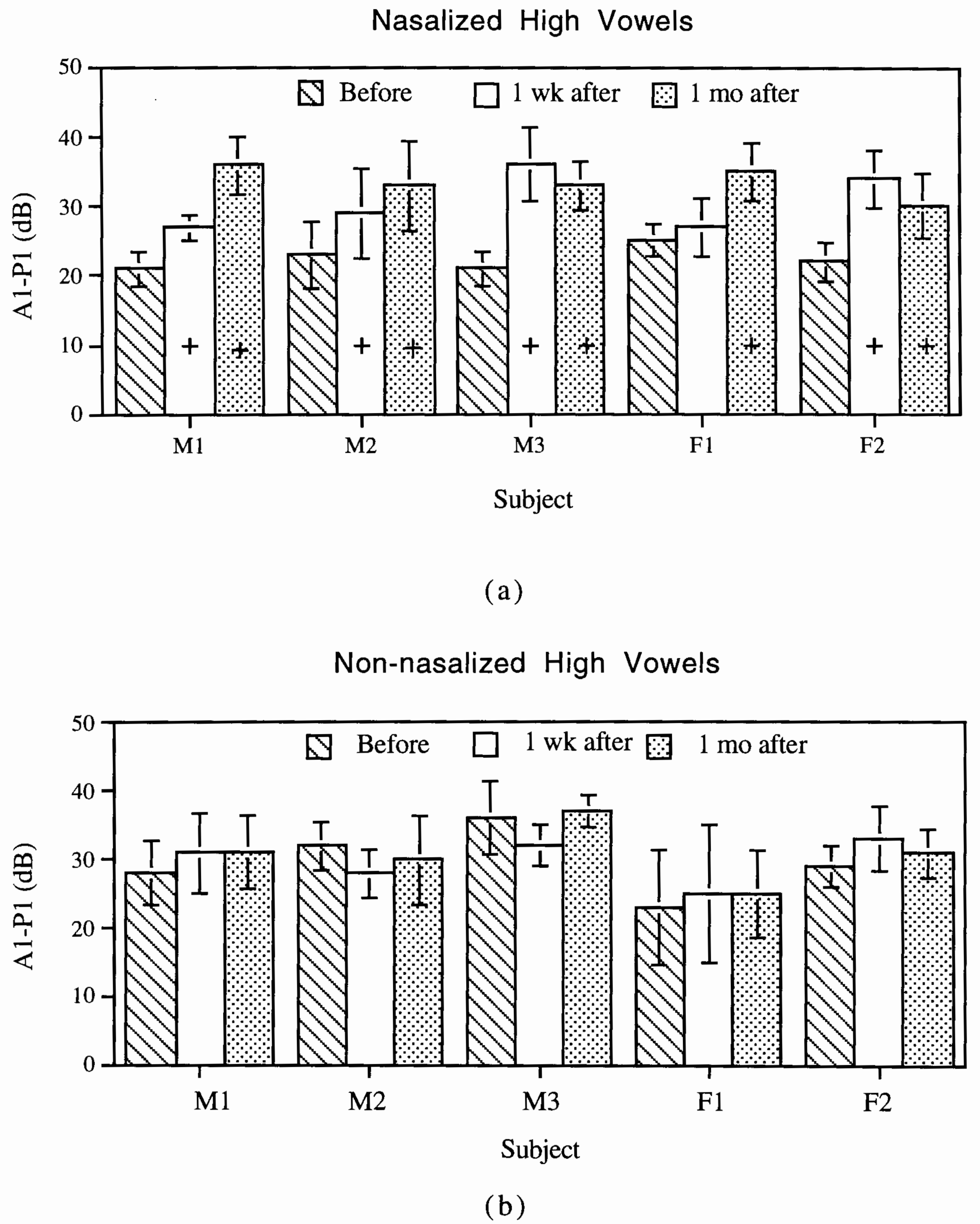


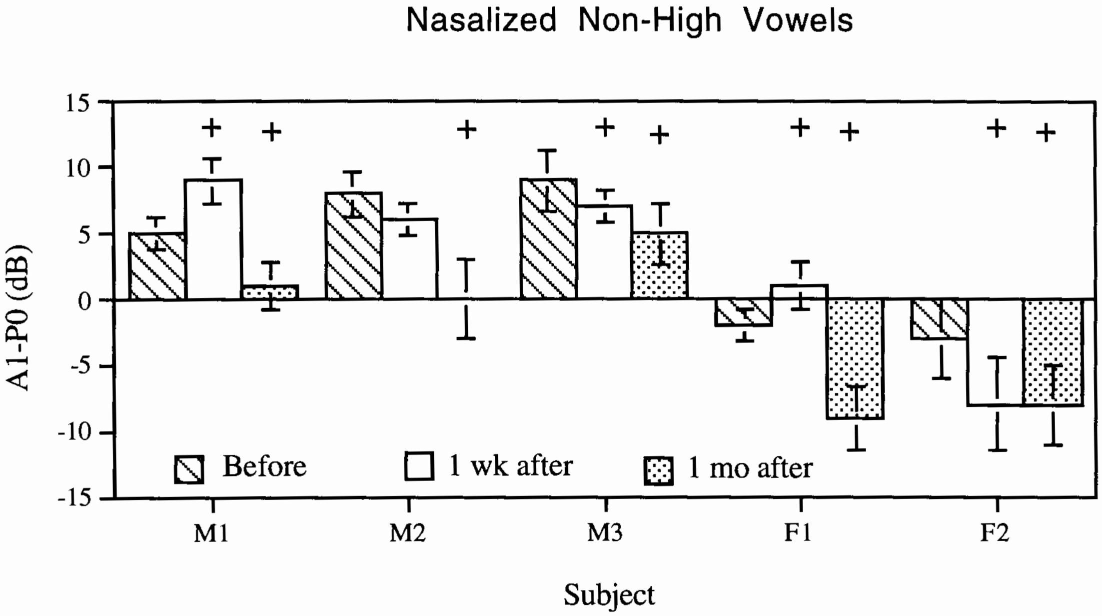
Figure 5.13: (a) A1-P1 was averaged for nasalized high vowels /i/ and /u/ across repetitions and locations within the vowel. (b) A1-P1 was also examined for non-nasalized /i/ and /u/ for comparison. The symbol '+' indicates $p < 0.01$ between A1-P1 from before the surgery and after the surgery.

7 dB for females. Figure 5.14b shows the results from corresponding non-nasalized vowels /æ/, /ɑ/, /ɛ/, and /ʌ/. The symbol '*' indicates $p < 0.05$ between A1-P0 from before the surgery and after the surgery. Only two speakers showed a statistically significant decrease in A1-P0 more than one month after the surgery, with a decrease of 2 dB to 4 dB for males and an increase of 1 dB to 0 dB for females.

In comparing the average values of A1-P0, before the surgery all of the speakers had an average in the nasal vowel lower than that in the corresponding non-nasal vowel. Furthermore, the averages from after the surgery in the nasal vowels were even lower than those in the non-nasal vowels. This trend indicates that before the surgery, the spectra of the nasal vowels had lower A1 and/or a higher P0 due to coupling to the nose and possibly greater open quotient of the glottal waveform. After the surgery, the spectra may have a even lower A1 and/or a higher P0. The average A1-P0 was greater for the male than for the female speakers, probably because females have more breathy voicing. With A1-P0 even lower after the surgery in the nasal vowels, they are expected to be perceived as more nasal after the surgery.

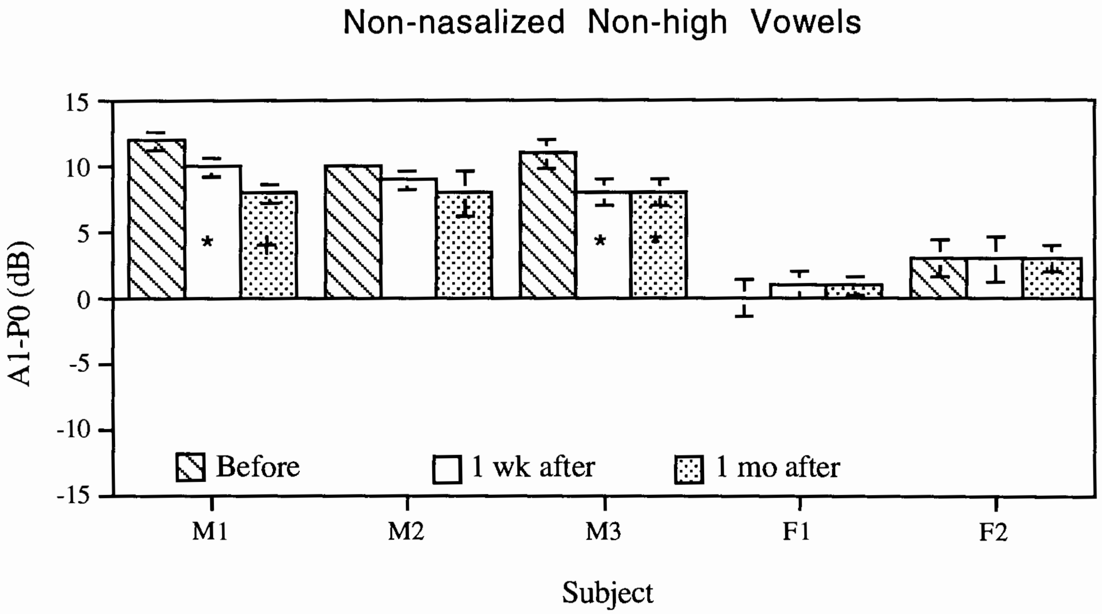
In general, A1-P1 increased after the surgery in nasal high vowels. This finding may be due to several factors: the removal of bony tissue, e.g. middle turbinates, and polyps, would decrease the total surface area and also eliminate constrictions. As the total surface area decreases and the constriction area is enlarged after surgery and healing, there would be smaller loss. Therefore, the first formant bandwidth is narrowed while its amplitude, A1, is raised. The peak amplitude P1 may be lowered by the zero in the vicinity of the pole due to greater coupling to the maxillary or sphenoid sinus from the surgery. According to Dang and Honda (1995), the nasal zero due to the maxillary sinus cavity resonance can be around 900 to 950 Hz while it can be around 1050 Hz due to the sphenoid sinus cavity resonance, depending on the speaker. The removal of tissues from the nasal cavity would enlarge the nasal passage volume where according to the perturbation theory (Stevens, in preparation), would shift the nasal pole to a lower frequency so that measuring P1 in its original region would yield a lower value.

On the other hand, A1-P0 generally decreased after the surgery in nasal non-high vowels. For all of the patients, the surgery enlarged the maxillary ostium (see Sec. 5.2.1), which can introduce a pole-zero pair in the lower frequency range (see Sec. 2.2.2). The cross-sectional area of the ostium was enlarged, giving a higher Q and a smaller bandwidth to the nasal pole. This reduced bandwidth



Nasalized Non-High Vowels

(a)



Non-nasalized Non-high Vowels

(b)

Figure 5.14: (a) A1-P0 was averaged for nasalized non-high vowels /æ/, /a/, /ɛ/, and /ʌ/ across repetitions and locations within the vowel. (b) A1-P0 was also examined for non-nasal /æ/, /a/, /ɛ/, and /ʌ/ for comparison. The symbol '+' indicates $p < 0.01$ and the symbol '' indicates $p < 0.05$ between A1-P0 from before the surgery and after the surgery.*

would increase the amplitude of the peak below the first formant, P0, and also lower A1 since the nasal zero is expected to be at a higher frequency than the nasal pole. Smaller losses, however, would raise A1.

Because of the opposing effects on A1 due to the nasal zero and decreased losses, P1 and P0 play more significant roles in A1-P1 and A1-P0 than A1. The decrease in P1 would cause A1-P1 to be larger after the surgery. Since an increase of A1-P1 is indeed observed in the high vowels /i/ and /u/, it is expected that nasal high vowels would sound less nasal after the surgery. The increase in P0 would cause A1-P0 to be smaller after the surgery. This change in A1-P0 was observed for non-high vowels, /æ/, /ɑ/, /ɛ/, and /ʌ/, especially for the low vowels. This finding suggests that nasal low vowels would sound more nasal after the surgery.

5.3 Perceptual Experiments

Patient self-evaluation of voice change after the surgery was obtained by using the questionnaire in Fig. 5.15. It gave some information on the perception of the patients' speech after the surgery. Table 5.5 summarizes the results for four patients. (M2 did no perceive any changes in his speech.) Most of the patients did not notice the change in their voice until others pointed it out to them, especially over the phone. The male speakers reported that their speech was deeper after the surgery. The changes were more noticeable in the morning, although they were not specific to particular sounds except for nasal consonants according to M1. Three speakers said that their speech was less nasal than before while one speaker said that it was more nasal; two said that their speech was

Table 5.5: Results from the patient questionnaire on the effect of surgery on the perception of speech. No noticeable change was detected in the speech of M2 according to the questionnaire.

Subj.	noticed by	change	time of day	most different	specific sounds	self help	intelligibility
M1	others	deeper < nasal	morning	3rd week	nasals	no	same
M3	phone	deeper < nasal	-	1st week	all similar	no	more
F1	phone	< nasal	morning	1st week	all similar	no	more
F2	self	more nasal	-	2nd week	all similar	clearing throat	same

Compared to immediately prior to the surgery, has the sound of your voice/speech changed after your surgery? (if so, complete the rest of the form; if not, you can stop here.)

How did you first notice a change in your voice/speech? (please check one)

- __ your own impression
- __ someone else pointed it out (please check one)
    - __ more noticeable when talking over the phone
    - __ more noticeable when talking in person
- __ other: ________________

How would you describe the change? (please check one)

- __ deeper
- __ clearer
- __ more nasal
- __ higher
- __ foggier
- __ less nasal
- __ others: __________________

Did or does your voice vary at different times during the day? If so, when did it vary the most?

- __ morning
- __ afternoon
- __ night
- __ other: ____________________

Did your voice vary during your recovery period (approximately the first month after surgery)? If so, when did it vary the most?

- __ the first week after the surgery
- __ the second week after the surgery
- __ the third week after the surgery

Did your voice sound more different when you said certain sounds? If so, which sounds were more different than prior to the surgery?

- __ words with nasal consonants (e.g. man)
- __ words without nasal consonants
- __ words spoken at a slow rate
- __ words spoken softly
- __ words spoken at a fast rate
- __ words spoken loudly
- __ others: ___________________

Have you tried to do something to compensate for the change? If so, how? Did it help?

Do you think you sound more intelligible, less intelligible, or equally intelligible since your surgery? What do you think influences your speech intelligibility?

Figure 5.15: Post-surgery speech assessment questionnaire answered by patients who have gone through endonasal sinus surgery more than a month after the surgery for self-evaluation of voice change after the surgery.

more intelligible. To further determine if the differences observed in the acoustic spectra are detected perceptually, isolated nasal vowels were presented to listeners for judgment of nasality.

5.3.1 Stimuli

The recordings of the five patients were used in the perceptual experiments. Nasalized vowels /i/ and /æ/ were excised from digitized waveforms of C_nVC_n (where C_n is a nasal consonant). A high vowel /i/ and a low vowel /æ/ were chosen since A1-P1 increased the most in high vowels while A1-P0 decreased the most in low vowels after the surgery according to acoustic analysis. Two tokens of isolated vowels with similar durations were chosen from those spoken before the surgery (b1 and b2) and two tokens were chosen from those spoken more than one month after the surgery (a1 and a2) for both vowels of all subjects. The tokens of the same vowel type between sessions were paired in all possible combinations, e.g. (b1, a1), (b1, a2), (b2, a1), and (b2, a2).

5.3.2 Procedure

There was 0.4 s silence between the stimuli within the pair. Each pair was presented once with 2.5 s of silence between consecutive pairs. The listeners judged /i/ and /æ/ separately. They finished judging the stimuli of a given speaker before moving on to the test stimuli of the next speaker. Five repetitions of each combination in one order and five repetitions of the combinations in the reversed order were randomized and recorded onto Maxell XLII cassettes using a Yamaha model 1000 cassette recorder and a custom software program for creating randomized tests of stimuli. The stimuli were presented through binaural headphones in a sound-treated room.

The listeners were tested individually. Each was asked to compare the two members of the pair to see which is more nasal and write the response '1' indicating the first member is more nasal or '2' indicating the second member is more nasal on a blank for the test item. If the listener judged the test stimulus to have no detectable difference in nasality, a random guess was recorded. Before each test, instructions were given orally and were also written on the response sheet. Also, the listeners were specifically told that there is no 'correct' answer and that there was no attempt to create an even distribution of answers on the test scale. To accustom the listeners to the task, the first ten pairs of each test were used solely for practice and not for scoring.

5.3.3 Listeners

Two trained phoneticians participated in the perceptual experiments. The inter-listener correlation was determined by counting the number of times, χ, a pair received the same expected judgment. /i/ from before the surgery was expected to be more nasal and /æ/ from after the surgery was expected to be more nasal. If χ was 4 or 5, the judge was considered to be clear about her judgment for that particular token. If it was 2 or 3, the judge was considered to be equivocal in her judgment, and 0 or 1 means the judge missed the task for that given token. The χ 's for one judge were correlated with those of the other judge. The inter-listener correlation r was 0.80.

5.3.4 Results and Discussion

5.3.4.1 High Vowel

The percentage of times the nasal vowel /i/ in a given pair was perceived to be more nasal than the other member of the pair is shown in Table 5.6 for each speaker. The left column of each pair

Table 5.6: The percentage of times a given stimulus was perceived to be more nasal for the isolated nasal /i/ spoken by patients who have gone through endonasal sinus surgery. Two stimuli from before the surgery and two stimuli from more than one month after the surgery were compared.

	inter-session comparison								average	
Subj	b1	a1	b1	a2	b2	a1	b2	a2	b	a
M1	**85**	15	**80**	20	**90**	10	**95**	5	**88**	12
M2	**75**	25	**55**	45	**100**	0	**70**	30	**75**	25
M3	**100**	0	**100**	0	**85**	15	**95**	5	**95**	5
F1	**70**	30	50	50	**65**	35	**95**	5	**70**	30
F2	**85**	15	**75**	25	**70**	30	**85**	15	**79**	21

shows the percentage for vowels from before the surgery (b1 or b2) to be perceived as more nasal and the right column of each pair shows the percentage for the vowel from after the surgery (a1 or a2) to be perceived as more nasal. The higher percentage within the pair is indicated by bold-faced numbers. For all of the test pairs, except one, tokens from before the surgery were judged to be more nasal greater percentages of the time than tokens from after the surgery. The last two columns list the average percentage over all pairs for tokens from before the surgery, 'b', and from after the surgery 'a'. This result shows that, on the average, nasalized /i/ was perceived to be less nasal after the surgery for all speakers. The maximum difference between the average percentages 'b' and 'a' was 90% and the minimum difference was 40% with an average difference of 63%. This agrees with the results from Sec. 5.2.3.b, which show A1-P1 increasing significantly for high vowels after the surgery. Since A1-P1 and nasality are inversely related, it is expected that nasalized /i/ would be perceived to be less nasal after the surgery.

5.3.4.2 Low Vowel

The percentage of times the nasal vowel /æ/ in a given pair was perceived to be more nasal than the other member of the pair is shown in Table 5.7. Again, the left column of each pair shows the

Table 5.7: The percentage of times a given stimulus was perceived to be more nasal for the isolated nasal /æ/ spoken by patients who have gone through endonasal sinus surgery. Two stimuli from before the surgery and two stimuli from more than one month after the surgery were compared.

	inter-session comparison								average	
Subj	b1	a1	b1	a2	b2	a1	b2	a2	b	a
M1	25	**75**	10	**90**	15	**85**	30	**70**	20	**80**
M2	**90**	10	**90**	10	**95**	5	**100**	0	**94**	6
M3	35	**65**	20	**80**	40	**60**	40	**60**	34	**66**
F1	30	**70**	45	**55**	0	**100**	5	**95**	20	**80**
F2	5	**95**	25	**75**	30	**70**	**65**	35	31	**69**

percentage for vowels from before the surgery (b1 or b2) and the right column of each pair shows the percentage for vowels from after the surgery (a1 or a2) that was perceived to be more nasal. The higher percentage within the pair is indicated by bold-faced numbers. For all of the speakers other than M2, all except one token from after the surgery were judged to be more nasal greater percentages of the time than tokens from before the surgery. The last two columns list the average percentage over all pairs for tokens from before the surgery, 'b', and from after the surgery 'a'. These data show that, on the average for four speakers, /æ/ was perceived to be more nasal after the surgery. The maximum difference between the average percentages 'b' and 'a' for the four speakers was 60% and the minimum difference was 32% with an average difference of 48%. This agrees with the results from Sec. 5.2.3.2 showing A1-P0 decreasing significantly for low vowels after the surgery. Since A1-P0 and nasality are inversely related, it is expected that /æ/ would be perceived to be more nasal after the surgery. Based on the percentages in Table 5.6 and Table 5.7, the surgery seems to have greater and more consistent effect on nasality for /i/ than for /æ/.

An attempt was made to determine the source of the discrepancy between the results of /æ/ for M2 and the other speakers. From Fig. 5.8a, A1-P0 in nasalized /æ/ decreased after the surgery, as for the other speakers. One would expect that nasalized /æ/ would be perceived as more nasal after the surgery. Figure 5.16a shows a spectrum of /æ/ more than a month after the surgery for M2. P0 seems more prominent and distinct than A1. Also, because it is at a higher frequency than F_{P0} of other male speakers, as indicated by Table 5.4, it may be mistaken as the first formant. Since the peak at F_{P0} is quite sharp, the first formant may be perceived as more prominent and therefore, the vowel may be perceived as less nasal. Fig. 5.16b shows the spectrum of /æ/ for M1 with P0 equal to A1 so that both peaks may contribute to the overall formant bandwidth, causing this vowel to be perceived as more nasal. Both the amplitude and frequency of the nasal pole, the nasal zero and the first formant could play a role in determining the perceived nasality. From the anatomical data, P0 would be relatively more prominent for M2 compared to the other speakers if the sphenoid sinus introduced the peak at F_{P0} after polyps were removed from its ostium. Dang and Honda (1995), by using sound pressure gradient in the nasal tract and sound-pressure at the nostril, showed that the nasal zero due to the sphenoid sinus can be as low as 310 Hz with the pole at a lower frequency. Another possible explanation is that M2 spoke at a faster rate so that the isolated vowels were much shorter in duration than those of the other speakers, causing the judgments to be difficult. According to both listeners, M2's vowels were the most difficult to judge, and consequently the reliability of their judgments could be influenced. In addition, from the patient questionnaire, M2 was the only one who did not notice a change in his own speech, indicating that the difference noticed by the judges may not reflect the overall effect of the surgery on his speech.

5.4 Summary

In the study of patients who have gone through endonasal sinus surgery, the effect of surgical alterations of the nasal anatomy on the acoustic characteristics of nasal consonants and nasalized vowels were examined and related to the results of perceptual experiments. In general, the surgery introduced to the nasal consonants less loss, resulting in a more prominent A1. With regard to anatomical changes, bony growth resection decreased the total surface area and removed the narrow constrictions in the nasal passage, therefore, reducing losses. In addition, the surgery lowered the amplitude of the peak around 1 kHz, $P1_n$. This modification may be due to a zero from coupling to the sphenoid sinus and a greater nasal cavity volume. $A1\text{-}P1_n$ was found to increase significantly after the surgery. For the males, the average $A1\text{-}P1_n$ increased between 11 dB and 23 dB; for females, it increased between 9 dB and 11 dB.

For nasal vowels, the surgery reduced the peak amplitude P1 around 900 Hz, possibly due to coupling to the maxillary or sphenoid sinus and greater nasal cavity volume. There was a significant increase in A1-P1 more than one month after the surgery for nasal high vowels /i/ and /u/, which did not occur for the non-nasalized vowels. A1-P1 was also examined for /ɛ/, and /ʌ/ recorded before and after the surgery. Although A1-P1 was found to be a measure of nasalization for the mid vowels as well, the surgery did not change it consistently for the different speakers. The average A1-P1 measured in the high vowels increased between 10 dB to 15 dB for males and 10 dB to 12 dB for females after the surgery. The results of perceptual experiment with nasalized /i/ were consistent with this finding: for all speakers, the vowel from after the surgery was perceived to be less nasal than that from before the surgery. The maximum difference between the average percentage of time the

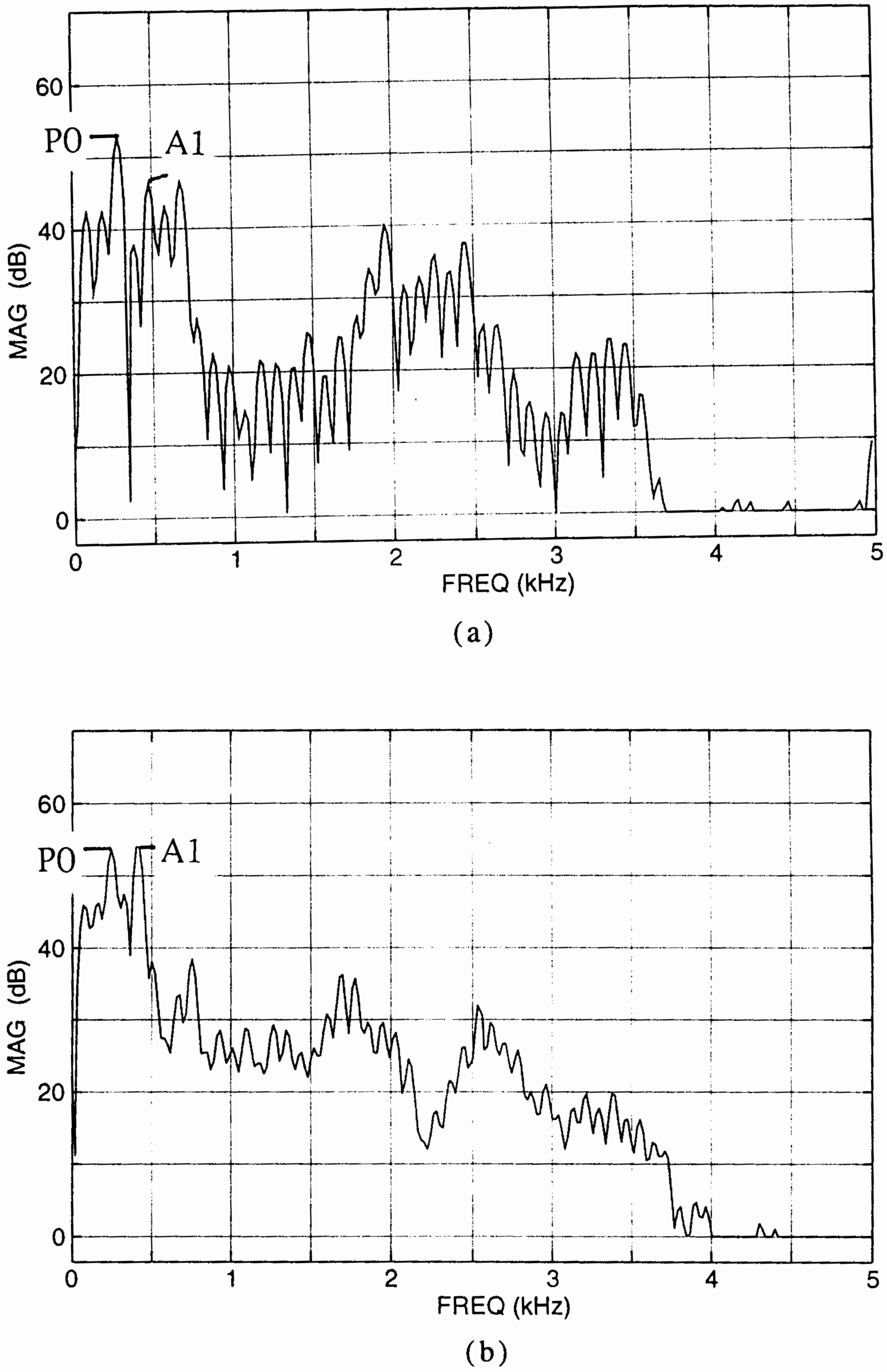


Figure 5.16: (a) Spectrum of /æ/ more than a month after the surgery for speaker M2 with amplitudes A1 and P0 labeled. (b) For comparison, the spectrum of /æ/ for speaker M1 recorded more than a month after the surgery is shown.

vowel was perceived as more nasal from before the surgery and after the surgery was 90% and the minimum difference was 40% with an average difference of 63%.

On the other hand, the surgery introduced a greater peak amplitude P0 around 250 Hz, possibly due to greater coupling to the maxillary or sphenoid sinus. There was a significant decrease in A1-P0 more than one month after the surgery for nasal low vowels /æ/ and /ɑ/, which did not occur as consistently for the non-nasalized vowels. From spectral analysis, A1-P0 was also found to be a measure of nasalization for mid vowels /ɛ/, and /ʌ/. A1-P0 showed more consistent decreasing than changes in A1-P1 in nasalized /ɛ/ and /ʌ/ after surgery. The decrease in the average A1-P0 for the non-high vowels was between 4 dB and 6 dB for males and 6 dB to 7 dB for females. The results of the perceptual experiments with nasalized /æ/ were consistent with this observation: for four of the five speakers, the vowel from after the surgery was perceived to be more nasal than that from before the surgery. The maximum difference between the average percentage of time the vowel was perceived as more nasal from before the surgery and after the surgery for the four speakers was 60% and the minimum difference was 32% with an average difference of 48%. The results show that A1-P1 and A1-P0 have the potential to be used clinically in accessing the effects of surgery on speech.

Chapter 6

Sources of Variability in Measures of Nasalization

The influence of three factors other than nasalization on measurements of A1-P1 and A1-P0 were investigated. (1) Since P1 and P0 are dependent on the vowel type, the parameters may need to be adjusted according to the frequencies of the first and second formants. (2) Excessive breathiness, another prevalent speech abnormality that contributes to reduced intelligibility and naturalness, may occur together with vowel nasalization, as in the speech of profoundly hearing-impaired speakers (Calvert, 1961). (3) The condition of a normal speaker's nasal cavities varies due to nasal cycles of spontaneous reciprocating nasal congestion, allergies, colds, moisture, etc. This and other time-varying conditions could also influence the stability of the parameters. This chapter examines the significance of the effect by the three factors on A1-P1 and A1-P0 and suggests some alternatives to make the measurements more robust in detecting nasalization.

6.1 Vowel-Type Adjustment

The effect of vowel type on the nasal peak amplitude P1 and P0 was observed and adjusted. The amplitude of the extra peak P1 is expected to be dependent on the vowel type. Since P1 rides on the skirts of the first and second formants, it is dependent on the formant frequencies and bandwidths. For a given degree of nasal cavity coupling, if F1 and F2 move closer to each other, P1 is boosted up, resulting in a decrease of A1-P1. Also, F1 is expected to influence P0 due to their proximity in frequency: as F1 lowers, P0 is boosted, decreasing A1-P0. The effect of the second formant on A1 is expected to be less significant since the formants are farther away in frequency. The parameters A1-P1 and A1-P0 should be adjusted for vowel types by correcting for the effects of the formants on the extra peaks. This adjustment is a function of the frequencies and bandwidths of the formants and the frequency of the extra peak.

6.1.1 Method

The general form of the vocal-tract transfer function during a nasalized vowel consists of poles similar to the formants of an oral vowel and nasal poles with corresponding nasal zeroes. The effect of each formant on the transfer function (omitting the effects of pole-zero pairs) can be expressed as

$$T_1(s) = \frac{(s_1 s_1^*)}{(s-s_1)(s-s_1^*)} \qquad 6.1$$

where s_1 is the complex frequency of the formant (the asterisk represents complex conjugate). If the frequency of the extra peak is F_{P1}, the frequency of the first formant is F1 and the bandwidth of the first formant is B1, the effect of the first formant component at F_{P1} is:

$$T1(F_{P1}) = \frac{(0.5B1)^2 + F1^2}{[((0.5B1)^2 + (F_{P1}-F1)^2)\cdot((0.5B1)^2 + (F1+F_{P1})^2)]^{1/2}}. \qquad 6.2$$

The effect of the second formant component at F_{P1} is:

$$T2(F_{P1}) = \frac{(0.5B2)^2 + F2^2}{[((0.5B2)^2 + (F2-F_{P1})^2)\cdot((0.5B2)^2 + (F2+F_{P1})^2)]^{1/2}}. \qquad 6.3$$

where F2 is the frequency of the second formant and B2 its bandwidth. The effects of higher formants at F_{P1} are small so they are not considered. By replacing P1 in dB by (P1-T1-T2) in dB, the parameter A1-P1 may be adjusted.

Approximations can be made if the bandwidths of the formants are much less than the formant frequencies. T1 can be approximated by

$$T1_{approx} = \frac{F1^2}{(F_{P1}-F1)\cdot(F1+F_{P1})}. \qquad 6.4$$

and T2 can be approximated by

$$T2_{approx} = \frac{F2^2}{(F2-F_{P1})\cdot(F2+F_{P1})}. \qquad 6.5$$

An identical technique can be used to determine the effect of the formants on P0 and the effect of F2 on A1.

6.1.2 Results and Discussion

6.1.2.1 English Vowels

Although A1-P1 and average nasality judgments of the synthetic stimuli in Sec. 4.1.4 are highly, negatively correlated for some vowels, the overall correlation coefficient across all vowels is only -0.19. The correlation coefficient of the average nasality judgment and a vowel-type-adjusted A1-P1 is -0.79, which is much more highly correlated than without adjustment. With approximation by using Eqs. 6.4 and 6.5, the correlation coefficient of the judgment and adjusted A1-P1 was -0.76. Therefore, it is sufficient to use approximated values to adjust for the effect of the formants on P1. The frequency of the second formant had little effect on A1 since no significant change of the correlation coefficient was found when the adjustments were made.

If the corrected A1-P1 is a measure of nasalization, then when there is no nasalization of the vowel, this value ideally should be the same, regardless of vowel type. For the non-nasal vowel, P1 was measured by using the harmonic with the maximum amplitude in the vicinity of F_{P1} in the corresponding nasalized vowel. In the synthesized vowels of Sec. 4.1.4, the amplitude at 950 Hz was used. The values of adjusted A1-P1 using the approximation method and unadjusted A1-P1 are shown in Fig. 6.1. The unadjusted A1-P1 was the average of values taken every 40 ms throughout the vowel. The adjusted A1-P1 was the average of A1-P1 after taking $T1_{approx}$ and $T2_{approx}$ into consideration. The horizontal line indicates the mean of the adjusted A1-P1 across vowels. The adjusted A1-P1 has a mean of 18.9 dB with a standard deviation (SD) of 2.1 dB; the unadjusted A1-P1 has a mean of 19.6 dB with a SD of 12.0 dB. Also, without adjustment, A1-P1 for the non-nasal reference differed as much as 34.6 dB (between "bot" and "beat"). The maximum difference was greatly reduced to 6.7 dB (between "bet" and "bait") by adjusting A1-P1. The correction was not perfect for the synthetic vowel since the frequency and the amplitude of the formants and the nasal peak can only be approximated by those of the harmonic

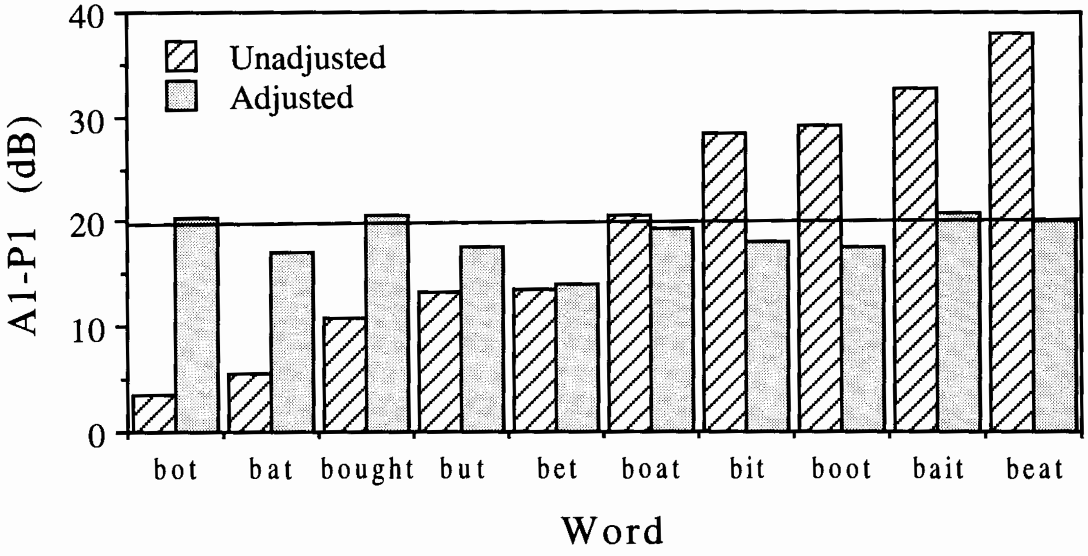


Figure 6.1: The values of unadjusted (solid bars) and adjusted (striped bars) A1-P1 were measured in the non-nasal synthesized reference vowels. The adjusted A1-P1 was the average of A1-P1 after taking $T1_{approx}$ and $T2_{approx}$ into consideration.

closest to them. Also, factor other than nasalization may influence the first formant amplitude since there was a difference in B1 by as much as a factor of two for the non-nasal vowels, as shown in Table 4.1.

The adjustment technique using the frequency of the formants and the nasal peak was applied to the speech of English and French speakers mentioned in Chap. 3. Figures 6.2a-b and Figs. 6.3a-b show the results for the English vowels across speakers, repetitions and locations within the vowel. The symbol '+' indicates $p < 0.01$ between the values of the non-nasal vowels and the nasal vowels indicating that with adjustment, nasal vowels can still be distinguished from non-nasal vowels. Figure 6.2a and Fig. 6.2b compare A1-P1 without adjustment and with adjustment, respectively, for both non-nasal and nasal vowels. The horizontal lines in Fig. 6.2b indicate the mean of the adjusted A1-P1 for non-nasalized and nasalized vowels. The means, standard deviations, and ranges of A1-P1 are listed in Table 6.1. With adjustment, the mean was lowered by 4.5 dB and 1.8 dB for the non-nasalized and nasalized vowels, receptively. The standard deviation (SD) was reduced by 9.0 dB and 4.3 dB after adjustment for the non-nasalized vowels and the nasalized vowels, respectively. The range for the non-nasal vowels was larger than that of the nasal vowels before correction but it became smaller after correction.

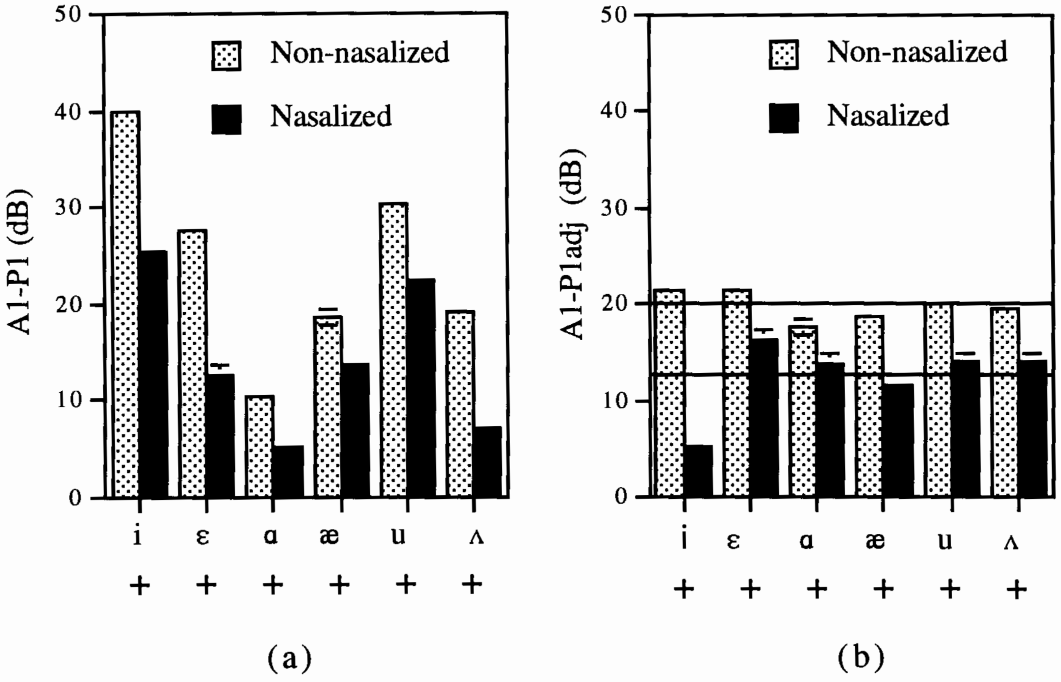


Figure 6.2: (a) The measured A1-P1 for the English vowels averaged across speakers, repetitions and locations within the vowel are shown for non-nasalized vowels and nasalized vowels. (b) The adjustment technique was applied to the speech of English speakers. The symbol '+' indicates p < 0.01 between the values of the non-nasal vowels and the nasal vowels.

Table 6.1: The effect of adjusting for vowel types on A1-P1 is shown by examining the means, standard deviations (SD) and ranges for the non-nasalized and nasalized English vowels.

A1-P1 (dB)						
Vowel Nasalization	Unadjusted			Adjusted		
	mean	SD	range	mean	SD	range
Non-nasal	24.3	10.5	29.8	19.8	1.5	4.0
Nasal	14.3	8.2	20.3	12.5	3.9	9.0

Figure 6.3a and Fig. 6.3b compare A1-P0 before and after applying the correction technique for the non-nasal and nasal vowels. The adjustment took into account the effect of the first formant on P0. Table 6.2 shows the means, standard deviations, and ranges of A1-P0. With adjustment, the mean was raised by 1 dB and 1.9 dB for both non-nasal and nasal vowels, respectively. The standard deviation remained about the same. The range was greater for the non-nasalized vowels than for the nasalized vowels both with

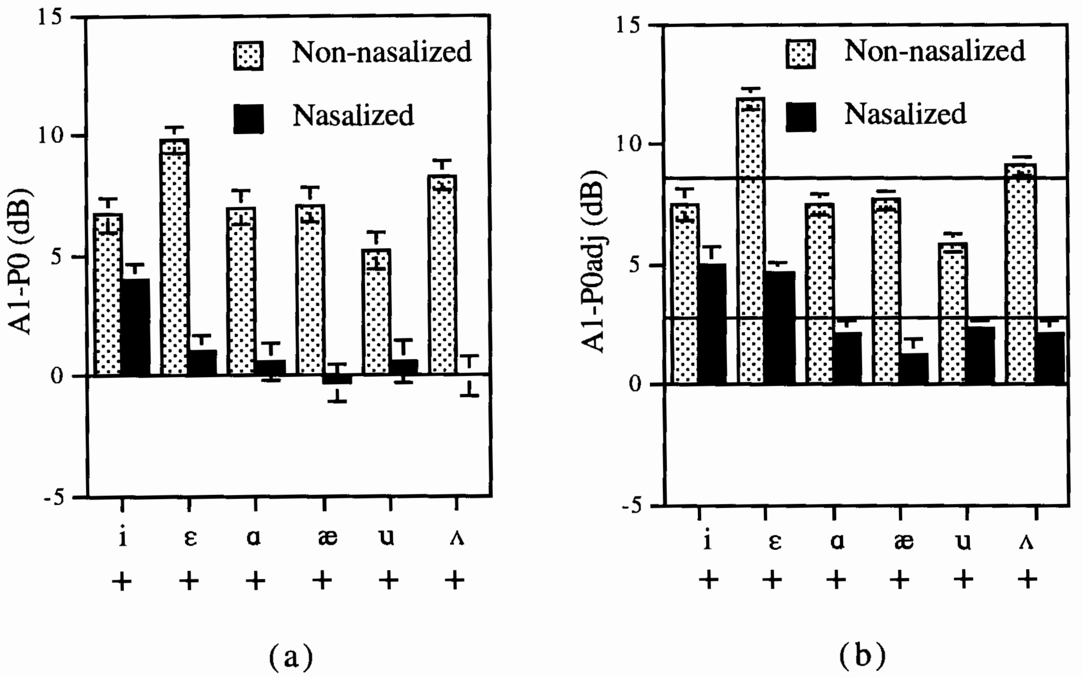


Figure 6.3: (a) The measured A1-P0 for the English vowels averaged across speakers, repetitions and locations within the vowel is shown for non-nasalized vowels and nasalized vowels. (b) The adjustment technique using the frequency of the formants and the nasal peak was applied to the speech of English speakers. The symbol '+' indicates p < 0.01 between the values of the non-nasal vowels and the nasal vowels.

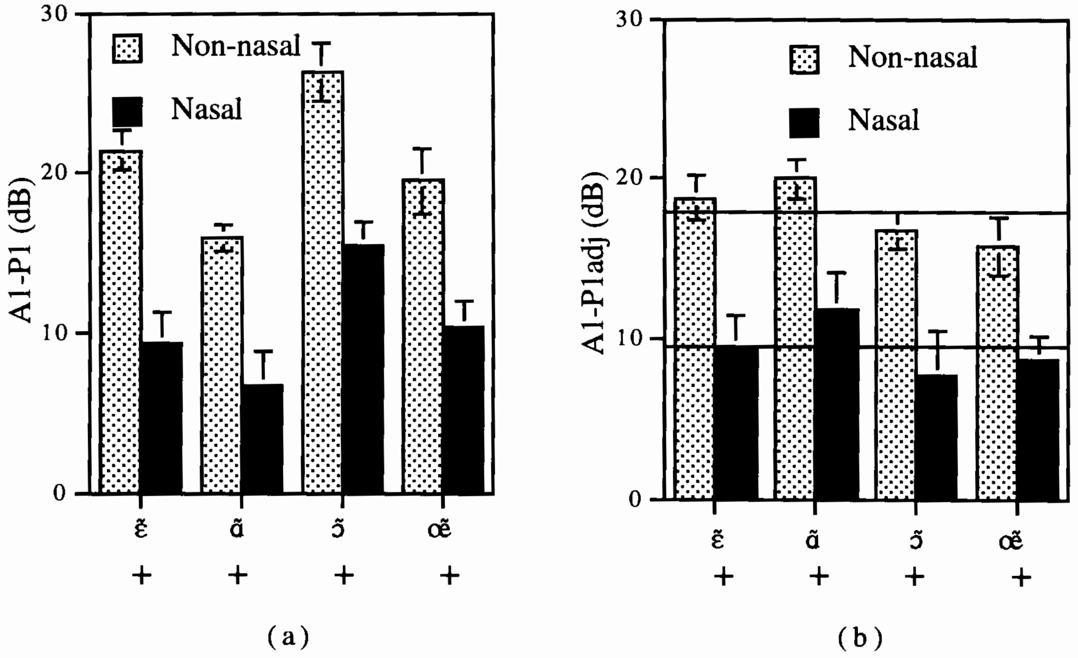
Table 6.2: The effect of adjusting for vowel types on A1-P0 is shown by examining the means, standard deviations (SD) and ranges for the non-nasalized and nasalized English vowels.

A1-P0 (dB)						
Vowel	Unadjusted			Adjusted		
Nasalization	mean	SD	range	mean	SD	range
Non-nasal	7.4	1.6	9	8.4	2.0	6
Nasal	1.0	1.6	6	2.9	1.5	4

and without adjustment. The adjustment had a much smaller effect on A1-P0 than on A1-P1. This result indicates that A1-P0 does not require adjustment for vowel types since the correction is usually small.

6.1.1.2 French Vowels

Figures 6.4a-b and Figs. 6.5a-b show the results for the French vowels across speakers and repetitions. Figure 6.4a and Fig. 6.4b compare A1-P1 before adjustment and after adjustment, respectively, for the non-nasal and nasal portions of the vowels. The horizontal lines in Fig. 6.4b indicate the mean of A1-P1 for the vowels. The means, standard deviations, and ranges of A1-P1 are listed in Table 6.3. The mean was lowered by 2.8 dB and 0.9 dB after adjustment for the non-nasal and nasal portion of the vowels. The standard deviation was lowered with adjustment for both the non-nasal and nasal portions. The range without adjustment was greater for the non-nasal portion than the nasal portion but with adjustment it was greater for the nasal portion than the non-nasal portion.



(a) (b)

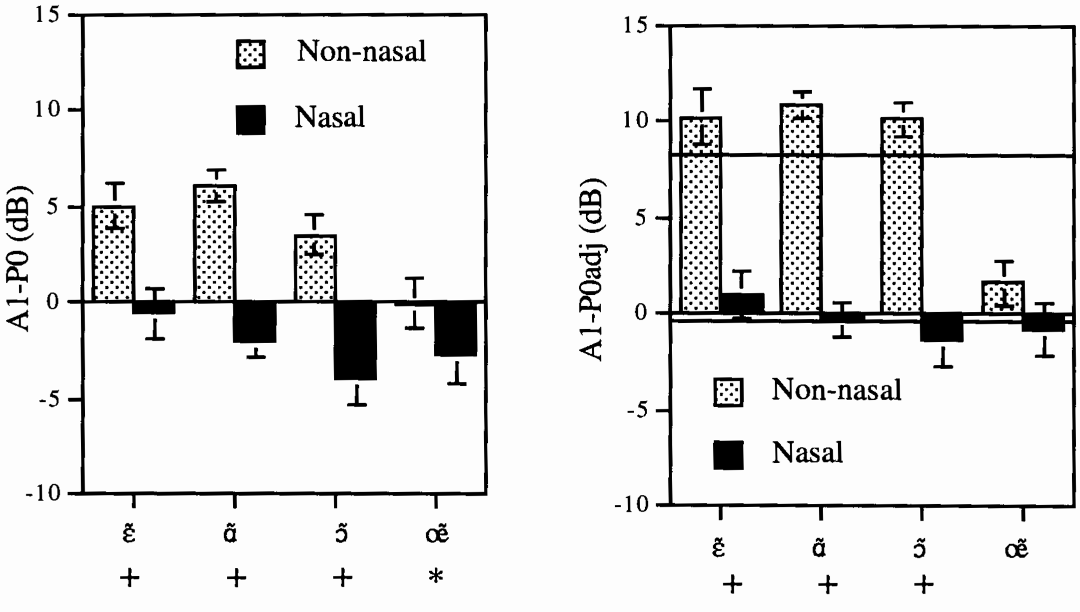
Figure 6.4: (a) The measured A1-P1 for the French nasal vowels averaged across speakers and repetitions is shown for non-nasal portion and nasal portion. (b) The adjustment technique was applied. The symbol '+' indicates p < 0.01 between non-nasal and nasal portions.

Table 6.3: The effect of adjusting for vowel types on A1-P1 is shown by examining the means, standard deviations (SD) and ranges for the non-nasalized and nasalized French vowels.

A1-P1 (dB)						
Nasal Vowel Portion	Unadjusted			Adjusted		
	mean	SD	range	mean	SD	range
Non-nasal	20.8	4.3	10.3	18.0	1.8	4.0
Nasal	10.5	3.7	8.8	9.6	1.8	4.2

Figure 6.5a and Fig. 6.5b compare the results for A1-P0 without adjustment and with adjustment, respectively, for the non-nasal and nasal portion of the French vowels. The means, standard deviations, and ranges of A1-P0 are listed in Table 6.4. The mean was greater after adjustment for both the non-nasal and nasal portions of the vowel. The standard deviation did not reduce much with adjustment, in fact, it even increased for the non-nasal portion. The range of A1-P0 for the non-nasal portion was greater than the nasal portion both before and after correction. The French vowels showed that correction for vowel type is unnecessary for A1-P0. The smaller range of A1-P0 for the nasal portion of the vowels than for the non-nasal portion might be due to the nasal peak being more prominent in the nasal vowels. It was therefore measured more accurately while for the non-nasal vowels, the nasal peak was more randomly assigned, depending on whichever harmonic has a greater amplitude.

There were several causes for the differences between the English and French results. The English vowels consisted of high and low vowels while the French vowels were non-high and non-low vowels. This would cause the range of the English vowels to be



(a) (b)

Figure 6.5: (a) The measured A1-P0 for the French nasal vowels averaged across speakers and repetitions is shown for non-nasal portion and nasal portion. (b) The adjustment technique was applied. '+' indicates $p<0.01$, and '' indicates $p<0.05$ between non-nasal and nasal portions.*

Table 6.4: The effect of adjusting for vowel types on A1-P0 is shown by examining the means, standard deviations (SD) and range for the non-nasalized and nasalized French vowels.

A1-P0 (dB)						
Nasal Vowel Portion	Unadjusted			Adjusted		
	mean	SD	range	mean	SD	range
Non-nasal	3.6	2.6	6.1	8.2	4.4	9.4
Nasal	-2.3	1.5	3.5	-0.3	1.0	2.2

greater than the French vowels, which is reflected in the ranges for the unadjusted nasal and non-nasal vowels, as shown in Tables 6.1 through 6.4. The fact that nasalization of the English vowels was due to coarticulation while the nasalization of the French vowels was due to intentional nasalization may introduce variations in the velopharyngeal opening size which also affect the ranges of A1-P1 and A1-P0 for the nasal vowels. Furthermore, the formants influencing the nasal peaks and the effect of breathiness may affect the ranges. Therefore, there was no clear difference in the ranges of the adjusted acoustic correlates between English and French vowels.

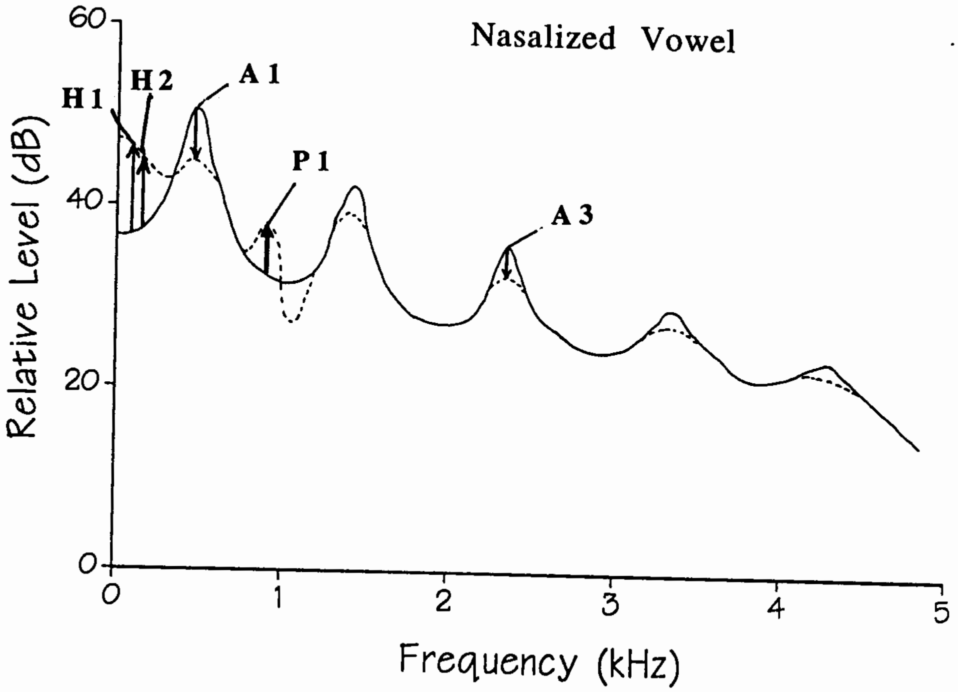
6.2 Nasality and Breathiness

In addition to inadvertent nasalization, another prevalent speech abnormality is excessive breathiness. Both speech disorders often occur together, particularly in the speech of the deaf (Calvert, 1961 and Brehm, 1922). Breathy vowels are formed with a larger amount of air passing through the glottis due to a wider opening between the arytenoids (Fischer-Jorgensen, 1967) or a more gradual closing of the vocal folds than in vowels with a modal phonation. Without an abrupt closure, the glottal volume velocity waveform tends to be more sinusoidal with no discontinuity in its slope, so that the tilt, which is the downward slope of the spectrum, is greater, especially at the middle and high frequencies. In fact, the third formant amplitude A3 was lowered by 15 dB when the tilt of 15 dB was introduced via synthesis (Hanson, 1995). Furthermore, a glottal waveform with a larger open quotient leads to a reduction in higher-frequency harmonics relative to the first harmonic amplitude H1.

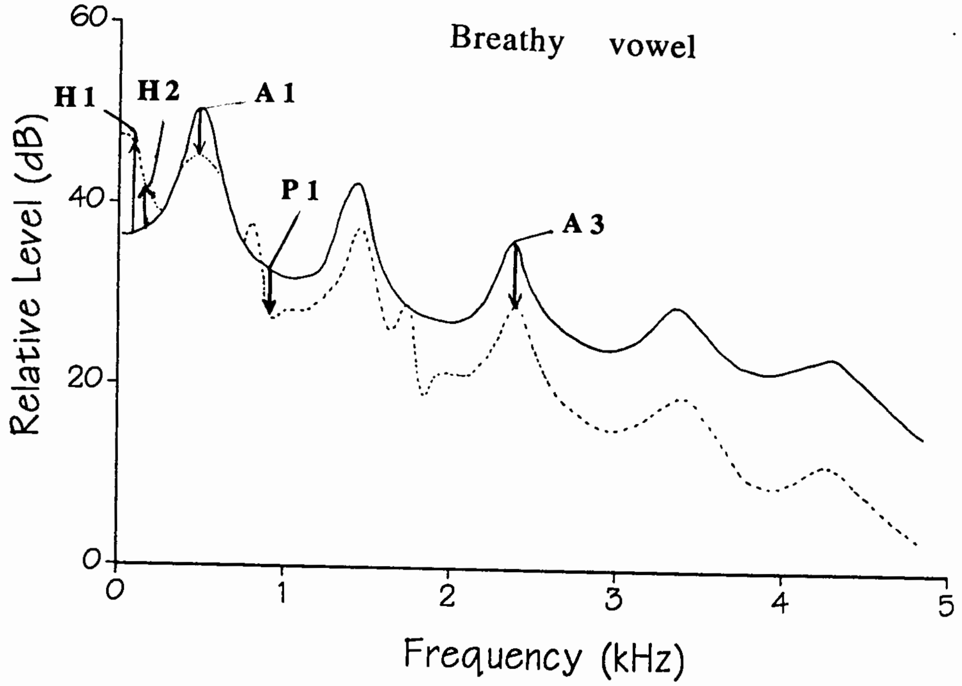
Breathiness involves acoustic coupling to the subglottal system. The partially open glottis can be modeled approximately as an acoustic mass in series with a resistance. The resistance causes energy loss in the frequency range of the first formant and therefore a bandwidth increase. Another attribute of tracheal coupling is additional pole-zero pairs. Theoretically, the second subglottal resonance around 1400-1800 Hz is often the most prominent since losses affect the first resonance and the greater acoustic mass of the glottis reduces the prominence of the higher resonances (Stevens, in preparation). The first resonance can occur around 700-800 Hz, which is close to the frequency of the nasal peak with amplitude P1.

Figure 6.6a and Fig. 6.6b show the possible effect of nasalization and breathiness, respectively, on a non-breathy and non-nasal reference vowel spectrum. The reference spectrum is shown in solid line. In Fig. 6.6a, nasalization increases the amplitude of the first and the second harmonic, H1 and H2, respectively, as a consequence of the nasal pole due to the sinuses. Intentionally large OQ to enhance nasality would also raise H1. A1 decreases due to a wider bandwidth from losses in the nasal cavity. P1 is expected to increase with nasalization since it is the amplitude of the nasal peak above the first formant. The amplitude of the third formant, A3, would decrease slightly due to the widening of the formant bandwidth. As shown in Fig. 6.6b, the effects of breathiness on the spectra are similar to those of nasalization. Due to a larger open quotient, H1 would be increased relative to H2. A1 and A3 are lowered as for nasalization since there are losses due to coupling to the tracheal region as well as an increased spectral tilt. The amplitude of the nasal peak, P1, could turn into a valley in a breathy vowel since it is in the region of the first tracheal zero.

According to Klatt and Klatt (1990), the acoustic measure H1-H2 is correlated with the OQ and perceived breathiness. By synthesis (Hanson, 1995), varying the OQ from 30 to 70 % causes H1-H2 to change about 10 dB without affecting the spectrum at high frequencies. Hanson also suggested that H1-A3, which may exceed 9 dB for most speakers, can be used as a measure for spectral tilt. For a breathy utterance, nasality measurements A1-P1 and A1-P0 should take into account the lowering of the higher harmonics, the relative raising of the first harmonic amplitude, and the extra bandwidth increase of the first formant. Even though the first tracheal resonance is generally less prominent than the second tracheal resonance, its effect on the nasality measurements should still be considered.



(a)



(b)

Figure 6.6: Comparing the spectrum of a non-breathy and non-nasal reference vowel, shown in solid line, to (a) spectrum of nasalized vowel and (b) spectrum of breathy vowel. The effect on the first and second harmonics, H1 and H2, respectively, the first and third formant amplitudes, A1 and A3, respectively, and the peak around 1 kHz, P1, are shown.

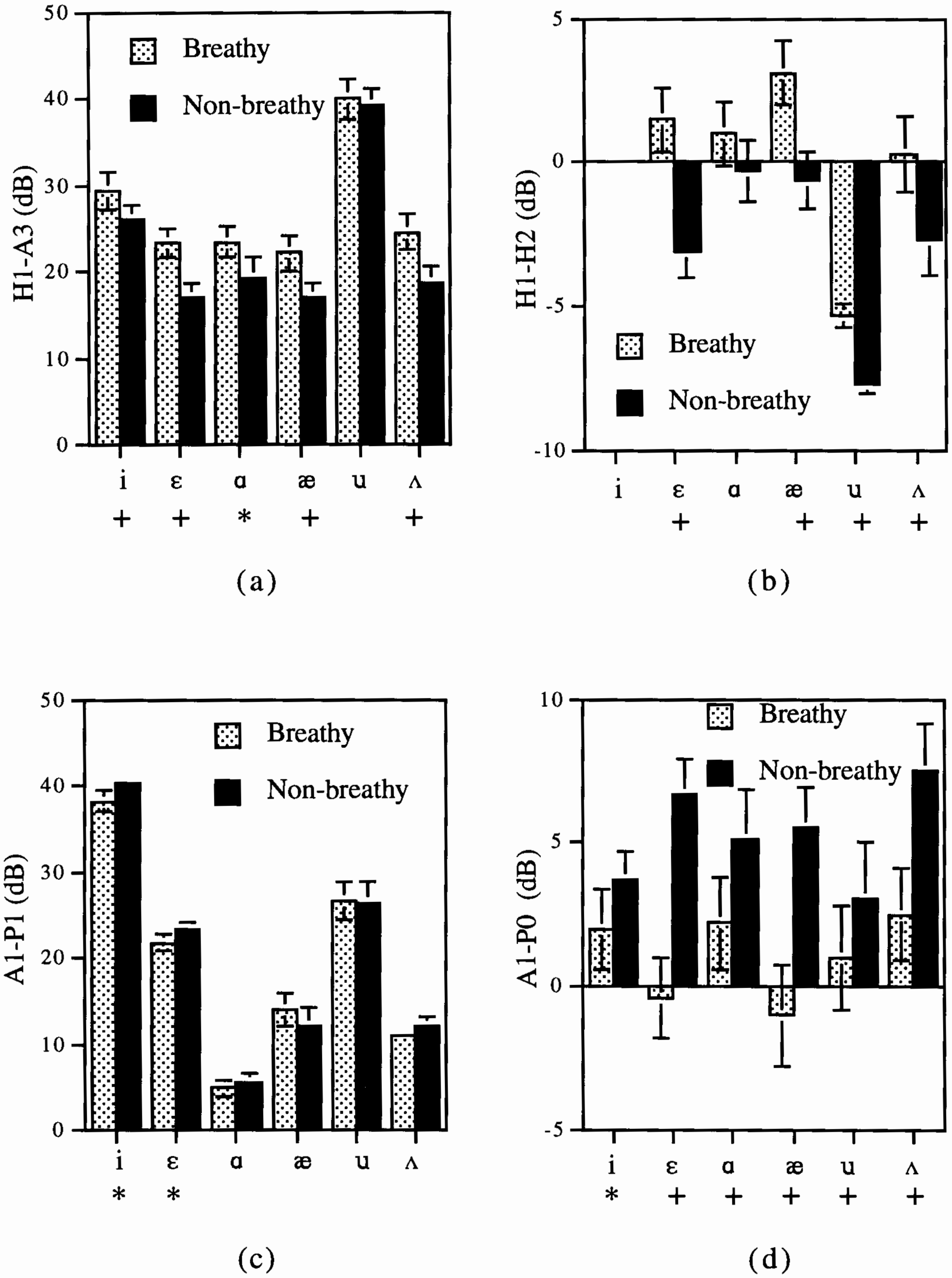
6.2.1 Method

To compare the effect of breathiness and nasalization quantitatively, seven native-English speakers were recorded saying /i/, /u/, /ɛ/, /ʌ/, /æ/, or /ɑ/ following /h/, /m/, or /b/ in the carrier phrase, "think ________ clearly". Each token was repeated three times. The same data acquisition procedure was followed as for the English and French speakers discussed in Sec. 3.1. Since the portion of the vowel closer to /h/ is more breathy, acoustic parameters of breathiness, H1-A3 and H1-H2, were measured at the beginning and the middle of the vowel to determine if there were any significant differences due to breathiness. Nasality correlates A1-P1 and A1-P0 were also measured to determine if there is any significant effect due to breathiness. The vowel following a stop consonant, e.g., bud, was considered as the non-nasal non-breathy reference. It was compared with the breathy vowel following /h/, e.g. hut, and with nasal vowel following a nasal consonant, e.g., mum by measuring A1-P1, A1-P0, H1-A3, and H1-H2 at the beginning of the vowel onset.

6.2.2 Results and Discussion

Figures 6.7a-d compare the acoustic parameters measured in the breathy (beginning) and non-breathy (middle) portions of the vowels following /h/. H1-A3 should be larger for breathy vowels since H1 is relatively higher due to the larger OQ and A3 is lower with tracheal losses and tilt. Figure 6.7a shows the result for H1-A3 measured in the breathy and non-breathy portions of the vowels. For all of the vowels, the average H1-A3 was larger for the breathy portion compared to the non-breathy portion, with five of the vowels showing statistically significant differences. Figure 6.7b shows values for H1-H2, which should be larger for breathy vowels since H2 decreases due to the larger OQ. H1-H2 was not measured in /i/ since the first formant was as low as the second harmonic. For the other vowels, H1-H2 was larger for the breathy portion of the vowel compared to the non-breathy portion, with four vowels showing statistically significant differences.

A1-P1 and A1-P0 were also measured in the beginning and middle of the vowels following /h/ to determine if these acoustic correlates of nasality are influenced by breathiness, as shown in Fig. 6.7c and Fig. 6.7d, respectively. Differences in A1-P1 were found to be statistically significant only for two of the vowels. For /i/, the



(a) (b)

(c) (d)

Figure 6.7: Comparing the acoustic parameters (a) H1-A3, (b) H1-H2, (c) A1-P1, and (d) A1-P0 measured in the breathy and non-breathy portions of the vowels. Measurements were made in the beginning and middle of the vowels following /h/ to determine if these acoustic correlates are influenced by breathiness. The symbol '+' indicates $p < 0.01$ and '' indicates $p < 0.05$ between the values of the breathy portion and the non-breathy portion of the vowels.*

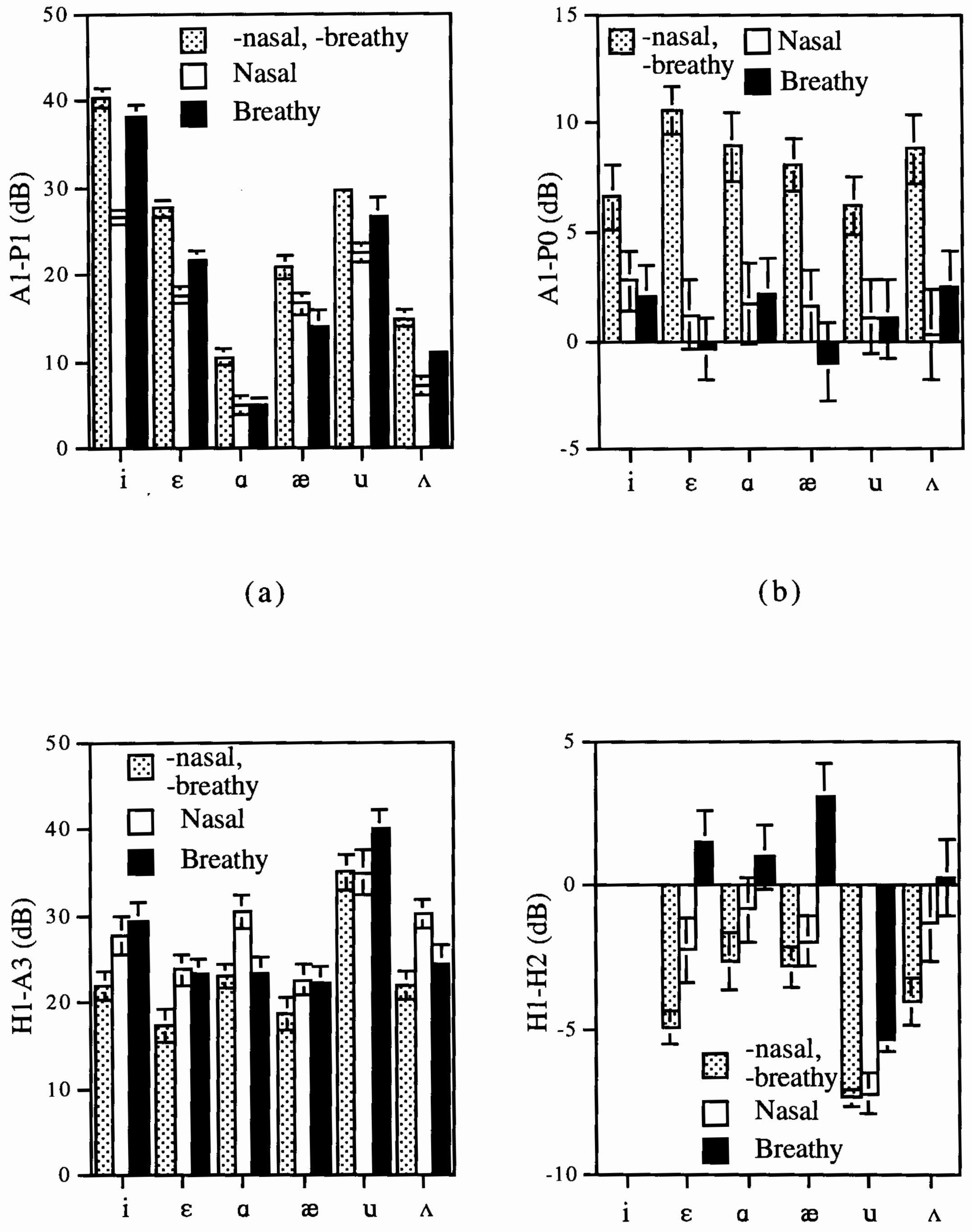
average was only 2 dB lower for the breathy portion than for the non-breathy portion of the vowel. With breathiness, both A1 and P1 are lowered, while with nasalization A1 is lowered but P1 is raised. This result indicates that A1-P1 is a good acoustic correlate to detect nasality, even when breathiness is present. A1-P0, on the other hand, shows a significant difference in breathy and non-breathy portions for all of the vowels. In breathy vowels, P0 is most likely to be H1 since H2 is relatively lower. With A1 lowered due to losses as well as tilt, A1-P0 is much smaller for breathy than non-breathy vowels. For nasal vowels, A1 is lowered due to losses and P0 is raised due to the nasal pole so that A1-P0 is also smaller. Thus A1-P0 is influenced by both nasalization and breathiness.

Figures 6.8a-d further examine the acoustic parameters in nasal vowels, breathy vowels, and non-nasal non-breathy references. Figure 6.8a shows that both breathiness and nasalization cause A1-P1 to decrease but with nasalization making a much greater effect, especially for the non-low vowels. The result was subjected to a repeated-measures analysis of variance on 3 vowel characteristics (non-nasal non-breathy, nasal, and breathy) x 6 vowel types (/i/, /u/, /ɛ/, /ʌ/, /æ/, and /ɑ/) and 7 speakers. There was a highly significant effect for vowel characteristics, [F(2,12) = 35.62, $p < 0.0001$] and a highly significant effect for vowel types [F(5,30) = 66.96, $p < 0.0001$].

Figure 6.8b shows that breathiness and nasalization reduce A1-P0 by about the same amount for all of the vowels. The result was subjected to a repeated-measures analysis of variance on 3 vowel characteristics (non-nasal-non-breathy, nasal, and breathy) x 6 vowel types (/i/, /u/, /ɛ/, /ʌ/, /æ/, and /ɑ/) and 7 speakers. There was a highly significant effect for vowel characteristics, [F(2,12) = 15.66, $p < 0.0005$] but a non-significant effect for vowel types [F(5,30) = 0.22, $p < 0.95$].

Figure 6.8c shows that in general breathiness and nasalization increase H1-A3 with varying effect on different vowel types. The result was subjected to a repeated-measures analysis of variance on 3 vowel characteristics (non-nasal-non-breathy, nasal, and breathy) x 6 vowel types (/i/, /u/, /ɛ/, /ʌ/, /æ/, and /ɑ/) and 7 speakers. There was a significant effect for vowel characteristics, [F(2,12) = 4.72, $p < 0.03$] and a highly significant effect for vowel types [F(5,30) = 27.51, $p < 0.0001$].

Figure 6.8d shows that for all of the vowels, H1-H2 increased for more nasal and more breathy vowels with the latter having the greater change. The result was subjected to a repeated-measures



(a)

(b)

(c)

(d)

Figure 6.8: Comparing the acoustic parameters (a) A1-P1, (b) A1-P0, (c) H1-A3, and (d) H1-H2 measured in the non-nasal, non-breathy (dotted bars), nasal (white bars), and breathy (black bars) portions of the vowels. Measurements were made in the beginning of the vowels following /b/, /m/, and /h/.

analysis of variance on 3 vowel characteristics (non-nasal-non-breathy, nasal, and breathy) x 4 vowel types (/ɛ/, /ʌ/, /æ/, and /ɑ/) and 7 speakers. There was a highly significant effect for vowel characteristics, [F(2,12) = 17.79, $p < 0.0003$] but not a significant effect for vowel types [F(3,18) = 1.40, $p < 0.28$].

The statistics indicate that all four parameters, A1-P1, A1-P0, H1-A3, and H1-H2, can be used to distinguish nasality and/or breathiness from the non-nasal non-breathy reference. However, it does not mean that they can be used to distinguish nasality from breathiness, as illustrated by Fig. 6.8b for A1-P0. Also, A1-P1 was statistically dependent on the vowel types, with nasalization having a greater effect than breathiness in the non-low vowels; H1-A3 was also dependent on the vowel types but no obvious trend can be observed.

To determine if any of the parameters is useful in distinguishing nasality from breathiness, a contrast by using means comparison was done between the nasal and breathy vowels. A1-P1 and H1-H2 showed statistically significant differences with $p < 0.003$ and $p < 0.004$, respectively, but A1-P0 and H1-A3 did not show significant differences with $p < 0.78$ and $p < 0.52$, respectively. This result suggests that A1-P1 can be used to detect nasality even when breathiness is present. In order to use A1-P0 to detect nasality when breathiness is suspected, H1-H2 may be used to distinguish the two vowel characteristics. More study needs to be done to better understand the role of these parameters in measuring nasality. The use of EGG to measure the glottal waveform may be helpful in determining the degree of breathiness.

6.3 Speaker and Time Variability

The complex anatomy of the nose varies greatly among speakers, not just in people who have had nasal surgery. From MRI studies, Dang *et al.* (1994) observed that the area function of the nasal tract varies widely from one subject to another. In addition, the volume of the paranasal sinuses varies more among speakers than that of the main passages. Furthermore, the velar and pharyngeal wall movement may differ from one speaker to another.

Even for a given speaker, the condition of the nasal cavities varies throughout the day, and from day to day. The speaking rate and velar and pharyngeal wall movement may change for different recordings, or even within the same recording. In addition, there is the “nasal cycle” which is the cycle of spontaneous reciprocating nasal congestion and decongestion. The duration of the cycle varies from 2-7 hours and occurs in the majority of the population (Maran and Lund, 1990). It occurs with any posture and persists with topical anesthesia, occluded nostril, or mouth breathing; it could be abolished temporarily with vasoactive substances, exercise and hyperventilation (McLean, 1984). Other factors that may affect the condition of the nasal cavity are age, exercise, respiration, nasal reflexes such as sneezing, inflammation such as allergic rhinitis, skin and air temperature, and emotions.

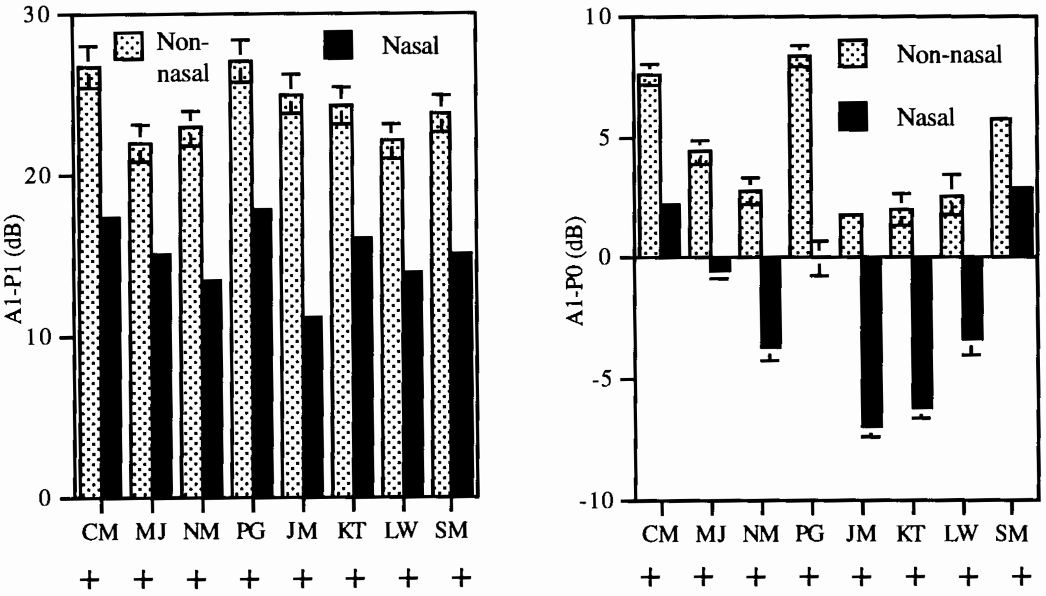
These differences may affect the acoustics of nasal sounds, and in particular may influence the measures A1-P1 and A1-P0. It is important to determine the range of the correlates for normal speakers so that changes due to speaker and/or time may be distinguished from changes due to incorrect velopharyngeal control or abnormal anatomical changes.

6.3.1 Method

To determine the effect of speaker variability on the acoustic correlates, results from the eight native-English speakers mentioned in Sec. 3.3 were examined in greater detail. To determine the effect of time variability, two additional recording sessions with the same procedure as that in Sec. 3.3.1 were conducted for three of the eight English speakers on different days. Both nasalized vowels adjacent to nasal consonants and non-nasalized vowels adjacent to stop consonants were digitized using the technique in Sec. 3.1. Only A1-P1 for the non-low vowels and A1-P0 for the non-high vowels were analyzed. The stability of A1-P1 and A1-P0 for both non-nasal and nasal vowels was determined and compared.

6.3.2 Results and Discussion

Figures 6.9a and Fig. 6.9b show the average A1-P1 and A1-P0, respectively, across vowel types and repetitions in a given recording session for the eight English speakers. The mean, standard deviation (SD), and range across speakers are shown in Table 6.5. The mean was lower for the nasalized vowels than the non-nasalized vowels. The SD's of A1-P0 was 3.5 dB and 1.8 dB larger than the SD's of A1-P1 for the non-nasalized and nasalized vowels, respectively. The range for nasal vowels was similar to that of the non-nasal vowels for A1-P1, which suggests that variations among speakers are due to differences other than nasalization, such as voicing, subglottal



(a) (b)

Figure 6.9: The average (a) A1-P1 across non-low vowels and (b) A1-P0 across non-high vowels and repetitions in a given recording session for the eight English speakers are examined for speaker variability. The non-nasal vowels are represented by dotted bars and the nasal vowels are represented by solid bars. The symbol '+' indicates p < 0.01 between the values of non-nasal and nasal vowels.

Table 6.5: Speaker variation on A1-P1 and A1-P0 is shown with the means, standard deviations (SD) and ranges for the non-nasalized and nasalized English vowels.

Vowel Type	A1-P1 (dB)			A1-P0 (dB)		
	mean	SD	range	mean	SD	range
Non-nasal	24.3	2.0	5.0	4.4	5.5	6.5
Nasal	15.0	2.2	6.9	-1.1	4.0	10.0

coupling, vowel intensity, and formant shifts. The SD for A1-P0 was larger for the non-nasal vowels, suggesting that variations among speakers are also due to differences other than nasalization, such as voicing, subglottal coupling, vowel intensity, and formant shifts.

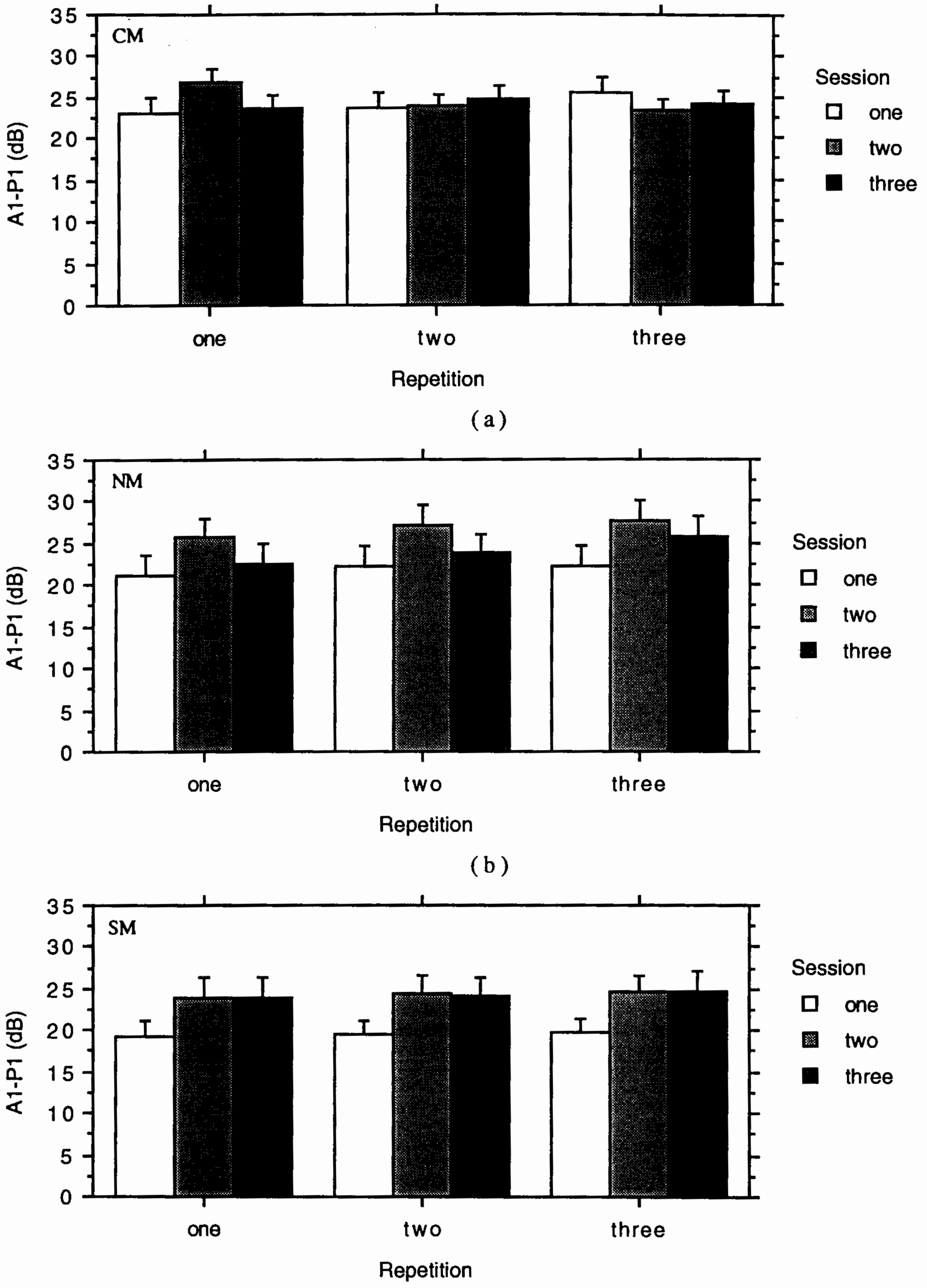
Speaker variability was further examined according to vowel type. Since A1-P1 is best used for the non-low vowels and A1-P0 is best used for the non-high vowels (Figures 3.3a-b to Figs. 3.8a-b), only the results for those values are used in discussing speaker variability. Table 6.6 lists the ranges among speakers for the non-

Table 6.6: Range of average A1-P1 and A1-P0 taken from the beginning, middle, and end of non-nasal and nasal vowels among eight English speakers

Vowel: Correlate	Non-nasal Vowel	Nasal Vowel
iː A1-P1	10 dB	7 dB
uː A1-P1	11 dB	14 dB
ɛː A1-P1	9 dB	10 dB
ʌː A1-P1	8 dB	11 dB
ɛː A1-P0	8 dB	13 dB
ʌː A1-P0	7 dB	12 dB
ɑː A1-P0	11 dB	12 dB
æː A1-P0	9 dB	14 dB

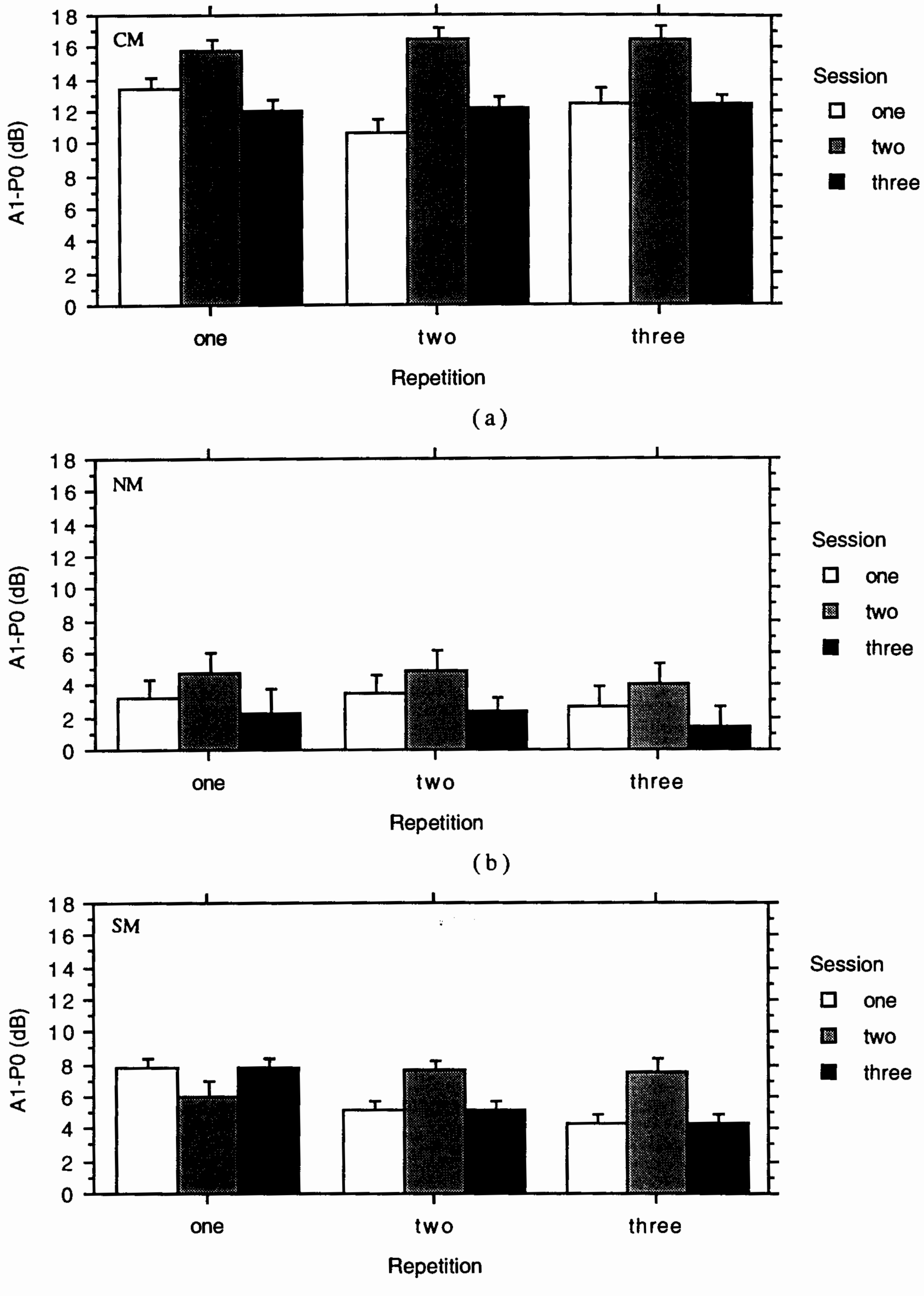
nasal and nasal vowels. The range of A1-P1 for /u/ was greater than that of the other vowels since the frequency of the second formant was closer to the frequency of P1 causing the measurement of A1-P1 to be less accurate. While for /i/, the range was the smallest since its P1 was the easiest to measure. The first formant frequency and the frequency of P0 were far apart for all of the non-high vowels, allowing more accurate measurements of P0 so that the range of A1-P0 did not vary much for the nasal vowels. The difference in the range for the non-nasal vowels may be due to measurement errors rather than speakers being less consistent for one vowel type than another. Therefore, speaker variability according to vowel types was found to be reduced when the nasal peak is farther away from the formants so that more accurate measurements of the acoustic correlates could be made.

The effect of time is examined in Figure 6.10a-c and Figs. 6.11a-c, which show plots of A1-P1 for non-low vowels and A1-P0 for non-high vowels, respectively, for speakers CM, NM, and SM, for each of the three recording sessions averaged across three repetitions and vowel types. The standard errors are shown as error bars in the plots. There was no trend for the values of A1-P1 from one repetition to another for a given session (bars with the same color), as shown in Fig. 6.10a-c for any of the three speakers. In most cases there was only a slight difference among the repetitions. For a given repetition, going from one session to another, there was a trend for NM and SM, as shown in Figs. 6.10b and 6.10c. The average



(c)

Figure 6.10: Average across non-low vowels of A1-P1 of (a) speaker CM, (b) speaker NM, and (c) speaker SM with three repetitions for recording session one (white bars), two (gray bars) and three (black bars) across all of the vowel types.



(a)

(b)

(c)

Figure 6.11: Average A1-P0 across non-high vowels of (a) speaker CM, (b) speaker NM, and (c) speaker SM with three repetitions for recording session one (white bars), two (gray bars) and three (black bars) across all of the vowel types.

value of A1-P1 was greater for session two than for the other sessions for NM. For SM, sessions two and three had a greater average than session one. The values were subjected to a repeated-measures analysis of variance on 3 time sessions x 3 repetitions and 3 speakers. There was not a significant effect for either time sessions, [F(2, 4) = 4.25, $p < 0.1$], or repetitions [F(2,4) = 1.59, $p < 0.1$]. This means that the values of A1-P1 from repetitions within sessions or across sessions are not affected significantly by factors that may vary over time.

Fig. 6.11a-c show values of A1-P0 for CM, NM, and SM, respectively. For a given session (bars with the same color), there was no fixed trend for each of the repetitions for CM and SM. For NM, the third repetition had a lower average in all of the sessions, although the values from the second repetition were about the same as those of the first repetition. For a given repetition, going from one session to another, the second session had higher average values of A1-P0 than the other two sessions for all of the repetitions and speakers except for repetition one of SM. Also, for NM, the values for the third session were lower than the other two sessions. The values were subjected to a repeated-measures analysis of variance on 3 time sessions x 3 repetitions and 3 speakers. There was not a significant effect for repetitions [F(2,4) = 1.59, $p < 0.1$] but a significant effect for time sessions, [F(2, 4) = 7.16, $p < 0.05$]. This means that the values of A1-P0 from repetitions within sessions do not vary much but when different sessions are considered, there is a significant difference in A1-P0, possibly due to variability in the glottal source.

The variation of A1-P0 was examined further by comparing the "range" and "threshold" for the non-nasal vowels and their corresponding nasal vowels. Figure 6.12 is a schematic diagram used to define "range" and "threshold." A1-P0 was first averaged across the three repetitions for a given session, a given vowel type, and a given location in the vowel for nasalized vowels as well as for non-nasalized vowels. The $\text{range}_{\text{-nasal}}$ is the maximum difference among the average values of A1-P0 for the non-nasal vowels (solid bars) and the $\text{range}_{\text{nasal}}$ is the maximum difference for the nasal vowels (stripped bars). The threshold is the average of $\Delta 1$, $\Delta 2$, and $\Delta 3$, which are the differences between average A1-P0 measured in the non-nasal vowel and in the nasal vowel within a session. It approximates the amount that A1-P0 needs to be lowered for the vowel to be classified as nasalized. Figures 6.13a and Fig. 6.13b plot the threshold versus the range for the non-nasal vowels and the nasal vowels, respectively. Each point in the plot represents

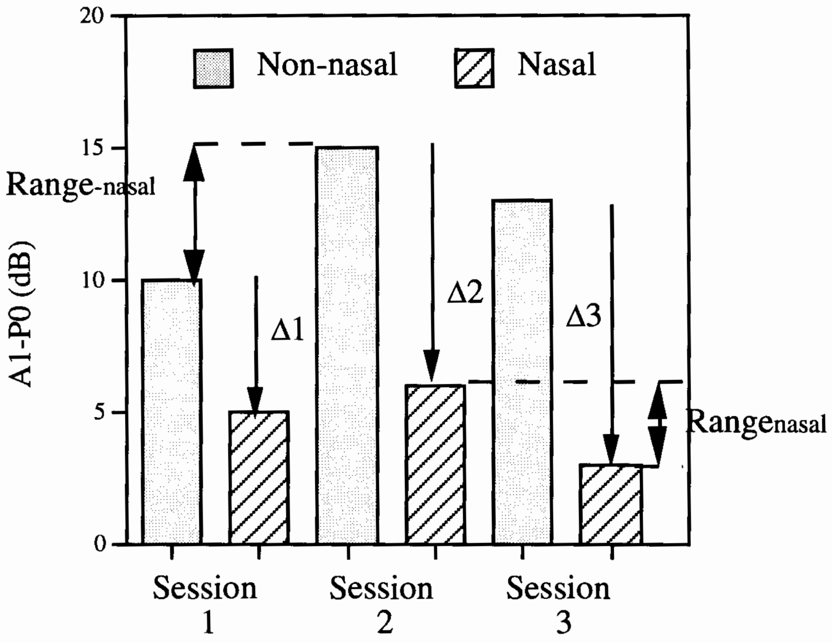


Figure 6.12: A schematic diagram used to define "range" and "threshold" for A1-P0. The range is the maximum difference among the non-nasal vowels (dotted bars) or among the nasal vowels (stripped bars). The threshold is the average of the differences, Δ1, Δ2, and Δ3, measured within each session.

averaged values across three repetitions at a location in a particular vowel for three sessions. If the range is larger than the threshold (below the diagonal), the effect of time variation on A1-P0 may cause a non-nasal vowel to be mistaken as a nasal vowel. The ratio of the number of points that fall below the diagonal to the total number of points is 0.49 for the non-nasal vowels and 0.27 for the nasal vowels. For most of the data for the nasal vowels, the range was smaller than the threshold. The larger ratio for the non-nasal vowels suggests that the variation may be due to alterations of factors other than nasalization.

6.4 Summary

Some of the factors that may influence the acoustic correlates A1-P1 and A1-P0 were examined to determine the limitation of these measures as well as to develop procedures for measuring the robustness of the measures for assessing nasality.

Vowel-type adjustment was found to be necessary for A1-P1 but not for A1-P0. In order to correct for the effect of vowel type,

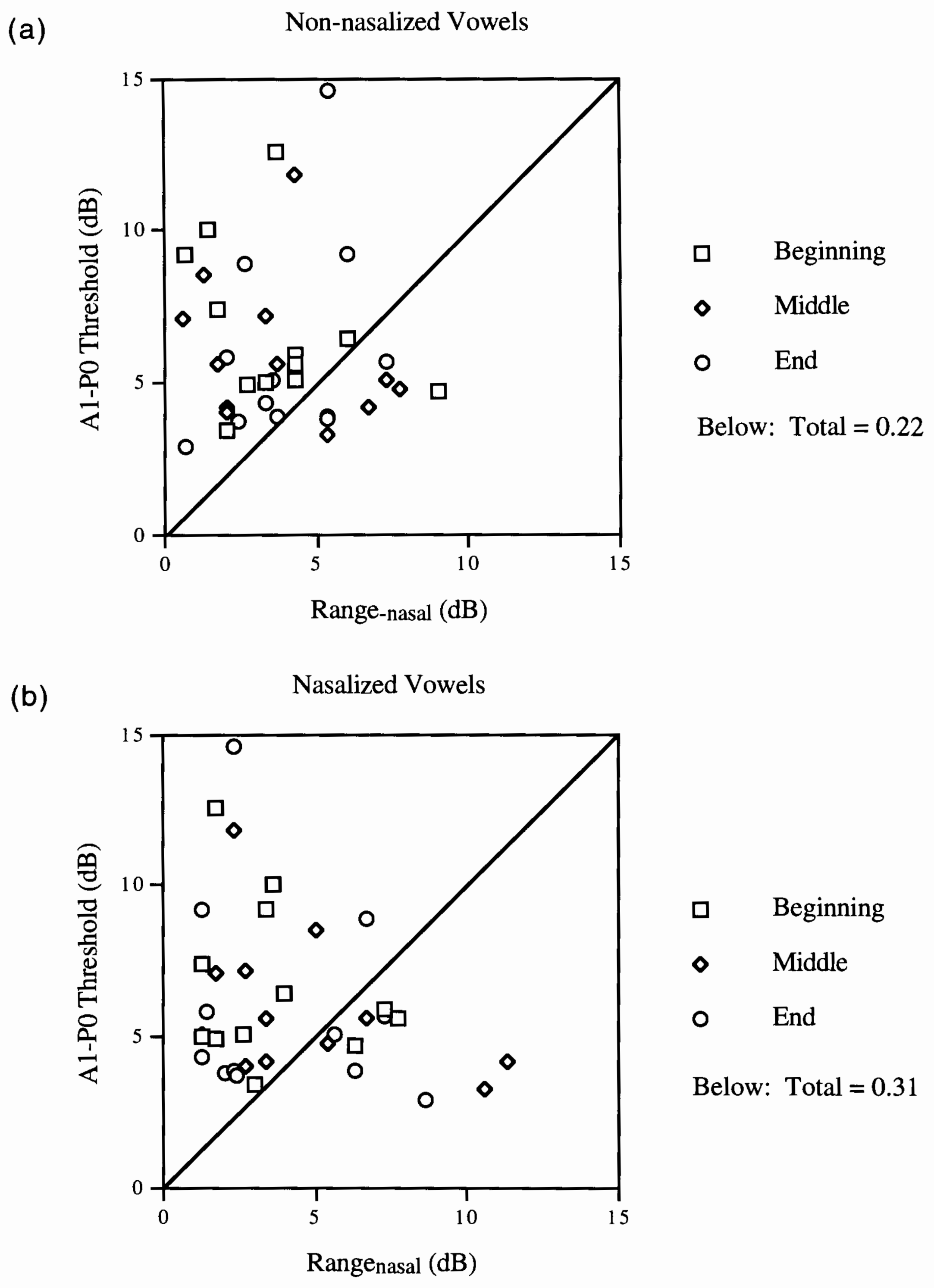


Figure 6.13: The threshold versus the range of A1-P0 for (a) the non-nasal vowels and (b) the nasal vowels. The ranges and thresholds were measured at the beginning (squares), middle (diamonds), and end (circles) of the vowel. If the range is larger than the threshold, the effect of time variation on A1-P0 may cause a non-nasal vowel to be mistaken as a nasal vowel.

the influence of the first two formants on P1 was taken into account. With adjustment, the range of A1-P1 in the synthetic speech for the non-nasal references was reduced by 28 dB. In the actual speech by English speakers, the standard deviation of A1-P1 was reduced by 9.0 dB and 4.3 dB in the non-nasal and nasal vowels, respectively. In the speech by French speakers, the SD of A1-P1 was reduced by 2.5 dB and 1.9 dB in the non-nasal and nasal vowels, respectively.

Since breathiness may occur with nasality, the relevant acoustic correlates, A1-P1, A1-P0, H1-A3 and H1-H2, were measured in the non-nasal and nasal vowels of English speakers. All of the parameters could be used to distinguish nasal and breathy examples from the non-nasal non-breathy references. A1-P0 and H1-H2 were independent of vowel types; while A1-P1 was a better measure for non-low vowels and A1-H3 was also vowel dependent. A1-P1 and H1-H2 showed statistically significant differences between nasal and breathy vowels but A1-P0 and H1-A3 did not show significant differences. These results suggest that A1-P1 can be used to detect nasality even when breathiness is present. However, A1-P0 cannot be used to distinguish nasality and breathiness. It may be possible to use H1-H2 or EGG to detect the presence of breathiness to help interpret the results for A1-P0.

The differences in anatomy, velum and pharyngeal wall movement, and nasal cycle were expected to affect the acoustics of nasal sounds. It was determined that the SD for A1-P1 among the speakers did not vary much while the SD for A1-P0 was larger for the non-nasal vowels than the nasal vowels, possibly due to variations in voicing characteristics. In comparing the ranges for different vowel types, speaker variability was found to be reduced when the nasal peak is farther away from the formants so that more accurate measurements of the acoustic correlates could be made without the influences of peaks close to each other in frequency. A1-P1 from repetitions within sessions or across sessions also did not vary much. A1-P0 showed little variation from repetitions within sessions but had much greater variations when different sessions were considered. Comparison of the threshold and range of non-nasal and nasal vowels showed that the changes in A1-P0 may be due to differences in the glottal waveform over time rather than in the anatomy and condition of the nasal cavity.

Chapter 7

Conclusion

7.1 Summary of Results

According to acoustic theory, the prominence of the extra peak P1 due to the pole-zero pair above the first formant, the lowering of the first-formant amplitude A1, and the prominence of the extra peak P0 due to the pole-zero pair at lower frequencies and/or a larger open quotient are significant acoustic cues for nasalization of vowels. A more prominent P1 is predicted for a larger velopharyngeal opening according to analysis of nasal configurations based on susceptance curves. A larger nasal coupling corresponds to a shift of the pole-zero pair to higher frequencies, with a greater effect on the zero, causing a greater peak amplitude at the frequency of the pole. The modification to the spectrum at the extra peak may be around 12 dB. A wider first formant is predicted from the loss of sound energy within the nasal cavity, which has a relatively large ratio of surface area to cross-sectional area compared to the oral cavity. The theory predicts that coupling to the nasal cavity may lower A1 by as much as 11 dB. A more prominent extra peak P0 is also predicted due to greater coupling to the maxillary or sphenoid sinus which has larger cavity volume compared to the other sinuses. The pole-zero pair may raise this extra peak by 2.4 dB. If the speaker “intentionally” increases the open quotient of the glottal waveform and if P0 is at the first harmonic, the amplitude may be expected to increase further by 5 dB.

From spectral analysis, the average across speakers and vowels for the frequency of the extra peak above the first formant, F_{P1}, was 966 Hz for English speakers and 936 Hz for French speakers. The average frequency across speakers and vowels for the extra peak at lower frequency, F_{P0}, was 216 Hz for English speakers and 237 Hz for French speakers. The difference between A1, which is inversely related to the bandwidth, and P1, which is directly related to the amount of separation between the nasal pole and zero, was proposed as a measure of the degree of vowel nasalization since A1-P1 involves both relevant cues. It is especially useful for non-low vowels which have a lower first formant frequency that may influence the measurement of P0. Similarly, A1-P0 was used as a measure of degree of nasalization for non-high vowels, which have higher first formant frequency that may obscure the measurement of P1. For English speakers, the difference between the average A1-P1 in the non-nasal vowels and that of the nasal vowels, (i.e., vowels between two nasal consonants), ranged from 10 dB to 15 dB, while the difference of A1-P0 ranged from 6 dB to 8 dB. For the French speakers, the difference between the average A1-P1 in the non-nasal vowels and that of the nasal vowels ranged from 10 dB to 12 dB while the difference of A1-P0 ranged from 7 dB to 9 dB. These values, including the maximum differences, may be explained by theoretical calculations. The correlates, however, were speaker-dependent, possibly due to differences in the spectrum of the glottal excitation, in the velopharyngeal opening, and in anatomy.

A perceptual experiment using synthesized versions of utterances (of the form bVt from a normal-hearing speaker) was carried out to further assess the validity of A1-P1 and A1-P0 as measures of nasality. For the experiment on A1-P1, as B1 widened, the perception of nasality increased, especially when B1 started out to be small and when FNZ was much closer to FNP. There was a ceiling effect of varying B1 and FNZ for stimuli that already had a wide B1. The vowel type x B1 x FNZ interaction was not significant, and consequently, one can analyze the effect of B1 and FNZ on vowel type separately. The low vowel group had slightly higher nasality perception than the other vowel groups at narrow B1, but as B1 increased nasality judgments were lower relative to the high- and mid-vowel groups. Also, the low-vowel group had lower nasality judgments than the high- and mid-vowel groups for a given FNZ. For the study on A1-P0, increasing the OQ to 60 or 70 had little effect on nasality. However, as the OQ was increased further to 80 or 90, it was enough to change nasality perception significantly. The effect was especially noticeable for the low vowels. For the individual non-low vowels, the correlation between A1-P1 and the nasality judgments was more significant than for the low vowels while for the individual non-high vowels, the correlation between A1-P0 and the nasality judgments was more significant than for the high vowels.

In the study of patients who have gone through endonasal sinus surgery, the effects of surgical alterations of the nasal anatomy on the acoustic characteristics of nasal consonants and nasal vowels were examined and related to the results of perceptual experiments. In general, the surgery produced narrower first formant bandwidths in the nasal consonants, resulting in a more prominent A1. This increased prominence is probably due to bony growth resection, which decreases the total surface area, and removal of narrow constrictions in the nasal passage. In addition, the surgery lowered the amplitude of the nasal peak at 1 kHz ($P1_n$), presumably due to greater coupling to the sphenoid sinus and greater nasal cavity volume. $A1\text{-}P1_n$ was found to increase significantly after the surgery. For nasalized vowels, the surgery resulted in a smaller peak amplitude around 900 Hz (P1). This reduced prominence may also be due to coupling to the sphenoid sinus and greater nasal cavity volume. A1-P1 increased significantly one month after the surgery for non-low vowels. These acoustic data are consistent with the results of a perceptual experiment with the vowel /i/. For all speakers, the vowel recorded after the surgery was perceived to be less nasal than that before the surgery. On the other hand, the surgery resulted in a greater peak amplitude around 300 Hz (P0), which may be due to greater coupling to the maxillary sinus or possibly the sphenoid sinus. A1-P0 decreased significantly one month after the surgery for non-high vowels, consistent with the results of a perceptual experiment with /æ/. For four of the five speakers, the vowel recorded after the surgery was perceived to be more nasal than that before the surgery.

A closer examination of the acoustic correlates A1-P1 and A1-P0 was done by addressing the influence of vowel-type adjustment, breathiness, and inter- and intra- speaker variability. Vowel-type adjustment by approximating the influence of the first two formants on P1 using their frequencies was necessary for A1-P1. With adjustment, A1-P1 in the synthetic speech for the non-nasal references had a standard deviation (SD) of 2 dB compared to the SD of 12 dB without adjustment. In the speech by English speakers, the SD was reduced by 9 dB and 4 dB in the non-nasal and nasal vowels, respectively. In the speech by French speakers, the SD was reduced by 3 dB and 2 dB in the non-nasal and nasal vowels, respectively. The acoustic correlates, A1-P1, A1-P0, H1-A3 and H1-H2, relevant to either nasality or breathiness, were measured in the non-nasal and nasal vowels of English speakers. A1-P1 and H1-H2 showed statistically significant differences between nasal and breathy vowels but A1-P0 and H1-A3 did not show significant differences. Therefore, A1-P1 may be used to detect nasality even when breathiness is present. On the other hand, when A1-P0 suggests nasality, H1-H2 and other parameters are needed to detect the presence of breathiness when it is also suspected. In the study of variability with English speakers, it was determined that A1-P1 did not vary much among speakers or with repetitions within session or across sessions. A1-P0 did vary among speakers, more so for non-nasal vowels, possibly due to voicing characteristics, vowel intensity, subglottal coupling, or formant shifts. In addition, A1-P0 from different sessions varied significantly, possibly due to differences other than the condition of the nasal cavity over time.

7.2 Directions for Future Research

Studies of languages with minimal pairs of nasal and non-nasal vowels would allow comparisons with the results of the English speakers. Alternativley, non-nasal French vowels in stop and nasal contexts should be studied. Such comparisons would determine if the speakers of languages with nasal-non-nasal distinction nasalize their vowels more than those speakers of languages without nasal-non-nasal distinctions, as suggested for French (Delattre, 1965; Mrayati, 1975).

Acoustic analysis and perceptual experiments also need to be done for those populations with high frequencies of inadvertent nasalization. It may be better to use non-naive listeners in those studies since their judgments could be more reliable. The acoustic correlates A1-P1 and A1-P0 may be useful in clinical applications which require quantification of nasality for patients with velopharyngeal insufficiency or cleft palate. The measures can also be used with deaf speakers as well as speakers with neurogenic and other disorders. Furthermore, patients who have gone through endonasal sinus surgery should be studied more to develop a more quantitative model of speech production that incorporates the size of the constrictions in the nasal passages, the volume change of the nasal cavity, and the dimensions of the paranasal sinuses. More careful study of acoustic-articulatory relations with better data on anatomy for individual speakers is needed.

Another task for future research is to automate the measurement of the acoustic correlates. This requires consistent detection of A1, P1, and P0. Measuring A1 involves locating the first formant using a formant tracker; measuring P1 requires locating an extra peak around 900 Hz to 1 kHz; and measuring P0 involves determining the low frequency harmonic that has the maximum amplitude. P1 should be fairly clear for non-low vowels and P0 should be fairly clear for non-high vowels. An initial pilot study could be done to test the consistency of this technique. For example, 10% of the data used in this study could be analyzed manually by another researcher following the same instructions. If the results were consistent with this study, automation should be implemented. The algorithm may be modified by assigning weights to the two acoustic correlates, A1-P1 and A1-P0, and combining them in a way to optimize the correlation between the acoustic correlate and the nasality judgments.

Bibliography

Abramson, A. S., and Cooper, F. S. (**1963**). “Slow motion x-ray pictures with stretched speech as a research tool,” J. Acoust. Soc. Am. **35**, 1888-1889.

Agur, A.M.R. (**1991**). *Grant’s Atlas of Anatomy, ninth edition.* (Williams & Wilkins, Baltimore).

Båvegård, M., Fant, G., Gauffin, J., and Liljencrants, J. (**1993**). “Vocal tract sweeptone data and model simulations of vowels, laterals and nasals,” STL-QPSR **4**, 43-76.

Baer, T., Gore, J. C., Gracco, L. C. and Nye, P. W. (**1991**). “Analysis of vocal tract shape and dimensions using magnetic resonance imaging: vowels,” J. Acoust. Soc. Am., **90**, 799-828.

BjÖrk, L., Nylén, B., Møller, A., and Fant, G. (**1961**). “Velopharyngeal function in connected speech,” Speech Trans. Lab. Q. Prog. Stat. Rep.**1**, Royal Institute of Technology, Stockholm, 13-14.

Bjuggren, G. and G. Fant (**1964**). "The nasal cavity structures," Speech Trans. Lab. Q. Prog. Stat. Rep. **4**, Royal Institute of Technology, Stockholm, 5-7.

Bolger, W.E. (**1991**). "Paranasal sinus bony anatomic variations and mucosal abnormalities: CT analysis for endoscopic sinus surgery," Laryngoscope, **101**, 56-64.

Brehm, F. E. (**1922**). "Speech correction," Am. Ann. Deaf. **67**, 361-370.

Broadbent, Sr., B. H., Broadbent, Jr. B. H., and Golden, W. H. (**1975**). *Bolton Standards of Dentofacial Developmental Growth.* (C. V. Mosby, St. Louis).

Calvert, D. R. (**1961**). *Some Acoustic Characteristics of the Speech of Profoundly Deaf Individuals.* Ph. D. dissertation, Stanford University.

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

Chakeres, D.W. (**1985**). "Computed tomography of the ethmoid sinuses," Otolaryngol Clin North Am. **18**, 29-42.

Dalston, R. M., Warren, D. W., and Dalston, E. T. (**1991**). "A preliminary investigation concerning the use of nasometry in identifying patients with hyponasality and/or nasal airway impairment," J. Sp. Hear. Res. **34**, 11-18.

Dang, J., Honda, K., and Suzuki, H. (**1994**). “Morphological and acoustical analysis of the nasal and the paranasal cavities,” J. Acoust. Soc. Am. **96**, 2088-2100.

Dang, J. and Honda, K. (**1995**). "An investigation of the acoustic characteristics of the paranasal cavities," Proceeding of the XIIIth International Congress of Phonetic Sciences **1**, 342-345.

Delattre, P. (**1954**). Les attributs acoustiques de la nasalite vocalique et consonantique. Studia Linguist. **8**, 103-109.

Delattre, P. (**1965**). “La nasalité vocalique en français et en anglais,” The French Review **39**, 92-109.

Delgutte, B. (**1980**). “Representation of speech-like sounds in the discharge patterns of auditory nerve fibers,” J. Acoust. Soc. Am. **68**, 843-857.

Draf, W. (**1983**). *Endoscopy of the Paranasal Sinuses.* (Springer-Verlag, New York).

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

Fant, G. (**1980**). “The relations between area functions and the acoustic signals,” Phonetica **37**, 55-86.

Fischer-Jorgensen, E. (**1967**). “Phonetic analysis of breathy (murmured) vowels in Gujarati,” Indian Linguistics **28**, 71-139.

Flanagan, J. L. (**1972**). *Speech Analysis Synthesis and Perception.* (Springer-Verlag, New York).

Friedman, W.H., Katsantonis, G.P., Rosenblum, B.N., *et al.* (**1986**) "Sphenoethmoidectomy: the case for ethmoid marsupialization," Laryngoscope, **96**, 473-479.

Fujimura, O. (**1960**). “Spectra of nasalized vowels,” Q. Prog Rep. **58**, Massachusetts Institute of Technology, 214-218.

Fujimura, O. and J. Lindqvist (**1964**). “The sinewave response of the nasal tract,” Speech Trans. Lab. Q. Prog. Stat. Rep. **1**, Royal Institute of Technology, Stockholm, 5-10.

Fujimura, O. and J. Lindqvist (**1971**). “Sweep-tone measurements of vocal-tract characteristics,” J. Acoust. Soc. Am. **49**, 541-558.

Fujimura, O., Kiritani, S., and Ishida, H. (**1973**). “Computed controlled radiography for observation of movements of articulatory and other human organs,” Comput. Biol. Med. **3**, 371-384.

Hanson, H. M. (**1995**). Glottal Characteristics of Female Speakers. Ph. D. Dissertation, Harvard University.

Hardt, J.M. and Montgomery, W.W. (**1976**). "Osteoplastic frontal sinusostomy: an analysis of 250 operations," Ann Otol Rhinolo Laryngol, **85**, 523-532.

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

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

Horii, Y. (**1980**). "An accelerometric approach to nasality measurement: a preliminary report," Cleft Palate J., **1**, 254-261.

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

House, A. S. and Stevens, K. N. (**1958**). "Estimation of formant bandwidths from measurements of transient response of the vocal tract," J. Sp. Hear. Res. **1**, 309-315.

Hudgins, C. V. and F. C. Numbers (**1942**). An investigation of the intelligibility of the deaf. Genet. Psychol. Monogr. **25**, 289-392.

Huffman, M. (**1990**). "The role of F1 amplitude in producing nasal percepts," J. Acoust. Soc. Am. Suppl. 1 **88**, S54.

Johansson, C., Sundberg, J., Wilbrand, H., and Ytterbergh, C. (**1983**). "From sagittal distance to area: a study of transverse, cross-sectional area in the pharynx by means of computer tomography," R. Inst. Technol. STL-QPSR **4/1983**, 39-49.

Katsantonis, G.P., Friedman, W.H., Sivore, M.C. (**1990**). "The role of computed tomography in revision sinus surgery," Laryngoscope **100**, 811-816.

Kennedy, D.W. (**1985**). "Functional endoscopic sinus surgery: technique," Arch. Otolaryngol **111**, 643-649.

Kennedy, D.W. (**1992**). "Prognostic factors, outcomes and staging in ethmoid sinus surgery," Laryngoscope **102**, 1-18.

Kennedy, D.W. , Zinreich, S.J., Rosenbaum, A.E., et al. (**1985**). "Functional endoscopic sinus surgery. Theory and diagnostic evaluation," Arch Otolaryngol **111**, 576-582.

Kiritani, S., Tateno, Y., Iinuma, T., and Sawashima, M. (**1977**). "Computed tomography of the vocal tract," in *Dynamic Aspects of Speech Production*, edited by M. Sawashima and F. Cooper (Univ. of Tokyo Press, Tokyo), pp. 203-206.

Kirschner, J.A., Yanagisawa, E., and Crelin, E.S. (**1961**). "Surgical anatomy of the ethmoidal arteries," Arch Otol **74**, 382-386.

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

Klatt, D. H. (**1984**). "The new MIT speech vax computer facility," Speech Commuication Group Working Papers IV, Research Laboratory of Electronics, MIT, Cambridge, MA, 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. and Beddor, P. (**1991**). "Coarticulation and the perception of nasality" Proceeding of the XIIth International Congress of Phonetic Sciences **4**, 38-41.

Larsen, P.L. and Tos, M. (**1991**). "Origin of nasal polyps," Laryngoscope **101**, 305-312.

Lawson, W. (**1991**). "The intranasal ethmoidectomy: an experience with 1,077 procedures," Laryngoscope **101**, 367-371, .

Lazar, R.H., Younis, R.T., and Long T.E. (**1993**). "Functional endonasal sinus surgery in adults and children," Laryngoscope **103**, 1-5.

Levine, H.L. (**1990**). "Functional endoscopic sinus surgery: evaluation, surgery, and follow-up of 250 patients," Laryngoscope **100**, 79-84.

Lindqvist, J. and Sundberg, J. (**1972**). "Acoustic properties of the nasal tract," Speech Trans. Lab. Q. Prog. Stat. Rep. **1**, Royal Institute of Technology, Stockholm, 13-17.

Lindqvist-Gauffin, J. and Sundberg, J. (**1976**). "Acoustic properties of the nasal tract," Phonetica **33**, 161-168.

Lothrop, H.A. (**1903**). "The anatomy of the inferior ethmoidal turbinate bone with particular reference to cell formation: surgical importance of such ethmoid cells," Ann Surg **38**, 233-255.

Loury, M.C. (**1993**). “Endoscopic frontal recess and frontal sinus ostium dissection,” Laryngoscope **103**, 455-458, .

McClumpha, S. L. (**1966**). *Cineflurographic investigation of velopharyngeal function in selected deaf speakers.* Master's thesis, Univ. of Florida.

McIntosh, C. (**1937**). *An Auditory Study of Nasality.* Master's thesis, Univ.of Iowa.

McLean, J. A. (**1984**). “Nasal rhinomanometry and experimental nasal challenges,” in *Rhinitis* , edited by G. A. Settipane (The New England and Regional Allergy Proceedings, Providence), p. 63.

Maeda, S. (**1982a**). "The role of the sinus cavities in the production of nasal vowels," Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, Paris, 911-914.

Maeda, S. (**1982b**). "Acoustic cues of vowel nasalization: a simulation study," J. Acoust. Soc. Am. Suppl. 1 **72**, S102.

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

Malme, C. (**1959**). Detectability of small irregularities in a broadband noise spectrum. Quarterly Progress Report **52**: 139-142. Cambridge: Research Laboratory of Electronics, Massachusetts Institute of Technology.

Maran, A. G. and Lund, V. J. (**1990**). *Clinical Rhinology.* (Thieme Medical Publishers, New York), pp. 35-36.

Masuda, S. (**1992**). “Role of maxillary sinus as a resonant cavity,” Nippon Jibiinkoka Gakkai Kaiho **95**, 71-80.

May, M., Sobol, S.M., and Korzec, K. (**1990**). “The location of the maxillary os and its importance to the endoscopic sinus surgeon,” Laryngoscope **100**,1037-1042, .

Messerklinger, W. (**1987**). *Endoscopy of the nose.* (Urban and Schwarzenberg, Baltimore/Munich), 6-18.

Metson, R. (**1992**). “Endoscopic treatment of frontal sinusitis,” Laryngoscope **102**, 712-716.

Moll, K. L. (**1962**). "Velopharyngeal closure on vowels," J. Sp. Hear. Res. **5**, 30-37.

Moll, K. L. and R. G. Daniloff (**1970**). “Investigation of the timing of velar movements during speech,” J. Acoust. Soc. Am. **50**, 678-684.

Moor, C. A. (**1992**). “The correspondence of vocal tract resonance with volumes obtained from magnetic resonance images,” J. Speech and Hearing Res., **35**,1009-1023.

Morse, Ph. M. (1948). *Vibration and Sound.* (McGraw-Hill, New York).

Mrayati, M. (**1975**). “Etude des voyelles nasales francaises,” Bulletin de l’Institut de Phonétique de Grenoble **IV**, 1-26.

Naumann, H. (**1965**). “Pathologische anatomie der chronische rhinitis und sinusitis,” in Proceedings VIII International Congress of Oto-rhinolaryngology. Amsterdam, Excerpta Medica, p. 80.

Ohnishi, T., Tachibana, T., Kaneko, Y., and Esaki, S. (**1993**). “High-risk areas in endoscopic sinus surgery and prevention of complications,” Laryngoscope **103**, 1181-1185.

Parsons, D.S., Phillips, S.E. (**1993**). “Functional endoscopic surgery in children: a retrospective analysis of results,” Laryngoscope **103**, 899-903.

Perkell, J. S. (**1969**). *Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study.* (MIT, Cambridge, MA).

Quigley, L. F., Shiere, F. R., Webster, R. C., and Cobb, C. M. (**1964**). "Measuring palatopharyngeal competence with the nasal anemometer," Cleft Palate J. **1**, 304-313.

Redenbaugh, M. A. and Reich, A. R. (**1985**). "Correspondence between an accelerometric nasal/voice amplitude ratio and listeners' direct magnitude estimations of hypernasality," J. of Sp and Hear. Res. **28**, 273-281.

Rice, D.H. (**1993**). *Endoscopic paranasal sinus surgery, 2nd ed.* (Raven Press, New York).

Rontal, E., Rontal, M., and Guilford, F.T. (**1979**). "Surgical anatomy of the orbit," Ann Otol **88**, 382-386.

Rosenthal, R. and Rosnow, R. (**1991**). *Essentials of behavioral research: methods and data analysis.* (McGraw-Hill, New York).

Sachs, L. (**1984**). *Applied statistics., 2nd edition.* (Springer-Verlag, New York).

Schaefer, S.D. and Close, L.G. (**1990**). "Endoscopic management of frontal sinus disease," Laryngoscope **100**, 155-160.

Schaefer, S.D., Manning, S, Close L.G. (**1989**). "Endoscopic paranasal sinus surgery: indications and considerations," Laryngoscope **99**, 1-5.

Schaitkin, B., May, M., Shapiro, A., Fucci, M., and Mester, S.J. (**1993**). "Endoscopic sinus surgery: 4-Year follow-up on the first 100 patients," Laryngoscope **103**, 1117-1120.

Schonle, P. W., Grabe, K., Wenig, P., Hohne, J., Schrader, J., and Conrad, B. (**1987**). "Electromagnetic articulography: use of alternating magnetic fields for tracking movements of multiple points inside and outside the vocal tract," Brain Lang. **31**, 26-35.

Shelton, R., Knox, A., Arndt, W., and Elbert, M. (**1967**). "The relationship between nasality score values and oral and nasal sound pressure level," J. Sp. Hear. Res. **10**, 542-548.

Shelton, R. , Knox, A. , Elbert, M., and, Johnson, T. (**1970**). Palate awareness and nonspeech voluntary palate movement. In J. F. Bosma (Ed.) *A Second Symposium on Oral Sensation and Perception.* (Charles C Thomas, Springfield, IL).

Sogg, A. (**1989**). “Long-term results of ethmoid surgery,” Ann Otol Rhinol Laryngol **98**, 699-701.

Som, P.M. (**1985**). “CT of the paranasal sinuses,” Neuroradiology **27**,189-201.

Spriestersbach, D. C. (**1955**). “Assessing nasal quality in cleft palate speech of children,” J. Sp. Hear. Dis. **20**, 266-270.

Staelin, D.H., Morgenthaler, and Kong, J.A. (**1994**). *Electromagnetic waves.* (Prentice Hall, Englewood Cliffs, NJ).

Stammberger, H. (**1986**). “Endoscopic sinus surgery-- concepts in treatment of recurring rhinosinusitis. part II. Surgical technique,” Otolaryngol Head Neck Surg **94**, 147-156.

Stammberger, H. and Wolf, G. (**1988**). “Headaches and sinus disease: the endoscopic approach,” Ann Otol Rhinol Larygol **97**, 3-21.

Stevens, K. N. (**in preparation**). Book on acoustic phonetics.

Stevens, K. N., Andrade, A., and M. Ceu Viana. (**1988**). Perception of vowel nasalization in VC contexts: a cross-language study. Presented at the 116th Meeting of the Acoustic Society of America, Miami, Florida.

Stevens, K. N., Fant, G., and S. Hawkins (**1987**). *Some acoustical and perceptual correlates of nasal vowels. In R. Channon and L. Shockey (Eds.), In Honor of Ilse Lehiste.* (Foris, Dordrecht, Holland).

Stevens, K. N., Kalikow, D. N., and Willemain, T. R. (**1975**). "A miniature accelerometer for detecting glottal waveforms and nasalization," J. Sp. Hear. Res. **18**, 594-599.

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

Takeuchi, S., Kasuya, H. and Kido, K. (**1977**). “A study on the effects of nasal and paranasal cavities on the spectra of nasal sounds,” J. Acoust. Soc. Jpn. **33**, 163-172.

Van Alyea, O.E. (**1939**). "Ethmoid labyrinth: anatomic study, with consideration of the clinical significance of its structural characteristics," Arch Otolaryngol **29**, 881-902.

Vining, E.M., Yanagisawa, K., Yanagisawa, E. (**1993**). "The importance of preoperative nasal endoscopy in patients with sinonasal disease," Laryngoscope **103**, 512-519.

Warren, D. W. (**1979**). "Perci: a method for rating palatal efficiency," Cleft Palate Journal **16**, 279-285.

Weiss, A. (**1954**). *Oral and nasal sound pressure levels as related to judged severity of nasality.* Lafayette, IN: unpublished Ph.D. dissertation, Purdue University.

Wigand, M.E. (**1981**). "Transnasal ethmoidectomy under endoscopical control," Rhinology **19**, 7-10.

Wright, J. (**1980**). The behavior of nasalized vowels in the perceptual vowel space. Report of the Phonology Laboratory **5** (University of California, Berkeley): 127-163.

Zinreich, S.J., Kennedy, D.W., Rosenbaum, A.E., *et al.* (**1987**). "Paranasal sinuses: CT imaging requirements for endoscopic surgery," Radiology **163**, 769-775.

Appendix

To determine the bandwidth increase due to coupling to the nasal cavity, losses need to be considered. This can be done at low frequencies by introducing to the circuit in Fig. 2.4 a resistive element R_1. To simplify the calculation, losses due to heat conduction and yielding walls are represented by R_1 in parallel with the inductance of the nasal cavity as shown in Fig. A.1.

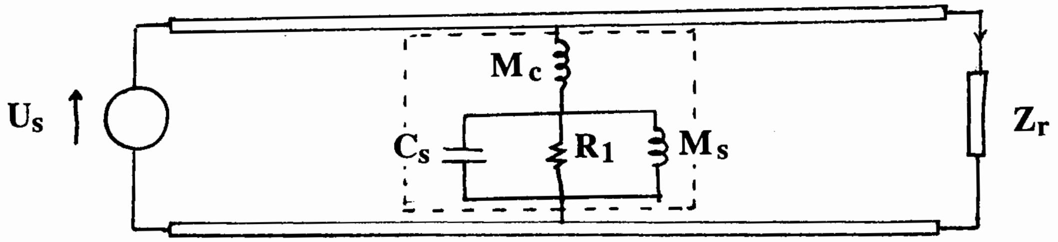


Figure A.1: A lumped element circuit is used to represent the nasal branch at low frequencies with a resistive element in parallel with the inductance of the nasal cavity representing losses due to heat conduction and yielding walls.

The admittance Y looking into the nose is that of the velopharyngeal port and the nasal cavity Y_s, as schematized by circuit elements in Fig. A.1. Let

$$Y = G + jB, \qquad \text{A.1}$$

$$Y_s = R_1 + jB_s, \qquad \text{A.2}$$

and

$$Z_c = jX_c, \qquad \text{A.3}$$

where $X_c = \omega M_c$. Then,

$$Y = \frac{1}{jX_c + \dfrac{1}{R_1 + jB_s}}$$

$$= \frac{R_1 + jB_s}{(1 - B_sX_c) + jX_cR_1} \qquad \text{A.4}$$

so that

$$Y = \frac{R_1(1 - B_sX_c) + B_sX_cR_1}{(1 - B_sX_c)^2 + (X_cR_1)^2} + j\frac{B_s(1 - B_sX_c) - X_cR_1{}^2}{(1 - B_sX_c)^2 + (X_cR_1)^2}. \qquad \text{A.5}$$

From Eqs. A.1 and A.5,

$$G = \frac{G_l}{(1-B_sX_c)^2 + (X_cG_l)^2} \qquad \text{A.6}$$

with $G_l = R_l$ and

$$X_c = \omega M_c = 2\Pi f \frac{\rho L_{vp}}{A_{vp}} \qquad \text{A.7}$$

where ρ is the density of air, L_{vp} and A_{vp} are the velopharyngeal opening length and cross sectional area, respectively. Also,

$$B_s = \omega C_s - \frac{1}{\omega M_s}. \qquad \text{A.8}$$

Since the natural frequency of the closed nasal cavity is:

$$f_n = \frac{1}{2\Pi\sqrt{M_sC_s}}, \qquad \text{A.9}$$

from Eqs. A.8,

$$B_s = \frac{2\Pi f}{(2\Pi f_n)^2 \frac{\rho L_n}{A_n}} - \frac{1}{2\Pi f \frac{\rho L_n}{A_n}} \qquad \text{A.10}$$

where L_{vp} and A_{vp} are the nasal cavity length and cross-sectional area, respectively. Since the bandwidth of the closed nasal cavity is:

$$\Delta f_n = \frac{R_l}{2\Pi C_s}, \qquad \text{A.11}$$

then

$$R_l = 2\Pi C_s \Delta f_n. \qquad \text{A.12}$$

By solving C_s from A.9 and substituting into A.12,

$$R_l = 2\Pi \Delta f_n \frac{1}{(2\Pi f_n)^2 \frac{\rho L_n}{A_n}}. \qquad \text{A.13}$$

The bandwidth of the nasal vowel is given approximately by:

$$\Delta f = \frac{G}{2\Pi C_v \ell_v} \tag{A.14}$$

C_v is the compliance/length of the vocal tract and ℓ_v is the length of the vocal tract, where

$$C_v = \frac{A_v}{\rho c^2} \tag{A.15}$$

with A_v as the volume per length of the vocal tract and c as the speed of sound. Equation A.14 is a reasonable approximation for a vocal tract with uniform cross-sectional area, and assuming that acoustic coupling to the nasal cavity does not significantly modify the distribution of sound pressure in the vocal tract for the uniform non-nasal configuration. G can be found by substituting Eqs. A.7, A.10, and A.13 into Eq. A.6. The values assigned to the parameters in these equations are listed in Table A.1. The value for L_{vp} was used by Stevens (in progress) in his calculations of nasal pole and zero frequencies. The values for L_n and A_n were obtained from MRI measurements (Dang *et al.*, 1994). The values for f_n and Δf_n were determined from sweep tone measurements of the nasal cavity (Lindqvist and Sundberg, 1972). The values for A_v and ℓ_v are based on the average of male speakers. From Eq. A.14, the bandwidth increase due to the nasal cavity coupling can be calculated for the various values of the velopharyngeal opening and

Table A.1: The values assigned to the parameters used to calculate X_C, G_l, and B_s in order to determine the amount by which the amplitude of the first formant, A1, is lowered for different velopharyngeal opening sizes and first formant frequencies.

Parameter	Value
ρ	0.00114 $\mathrm{gm/cm^3}$
L_{vp}	2 cm
L_n	11.6 cm
A_n	2 $\mathrm{cm^2}$
f_n	500 Hz
Δf_n	160 Hz
A_v	3 $\mathrm{cm^2}$
ℓ_v	17 cm
c	35400 cm/sec

first formant frequency. (The approximation of the uniform tube for F1 other than 500 Hz would introduce error to the calculations.) If one assumes that the bandwidth of the non-nasal vowel is 60 Hz, the amount by which the amplitude of the first formant is lowered, ΔA1, is shown in Fig. 2.7.

Loss due to viscous friction is represented by R_o in series with the inductance of the nasal cavity as shown in Fig. A.2. Eq. A.6 still holds if

$$G_l = \frac{R_o}{R_o^2 + \omega^2 M_s^2} \qquad \text{A.15}$$

and

$$B_s = \omega\frac{C_s R_o^2 + \omega^2 M_s^2 C_s - M_s}{R_o^2 + \omega^2 M_1^2} . \qquad \text{A.16}$$

which leads to

$$B_s = \frac{C_s R_o^2}{\omega M_s^2} + \omega C_s - \frac{1}{\omega M_s} \qquad \text{A.17}$$

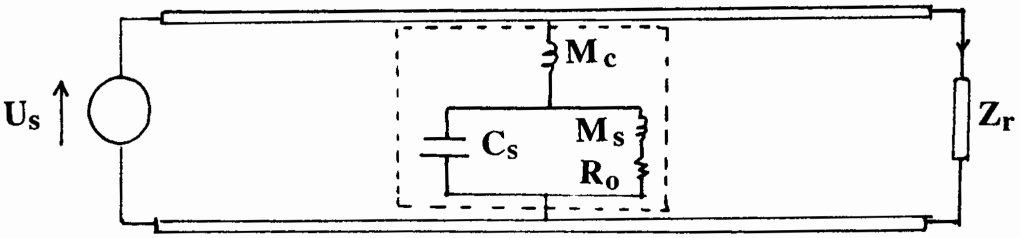


Figure A.2: A lump element circuit is used to represent the nasal branch at low frequencies with resistive element in series with the inductance of the nasal cavity representing loss due to viscous friction.

With substitution for M_s, Eq. A.15 becomes

$$G_l = \frac{R_o}{R_o^2 + (2\Pi f\frac{\rho L_n}{A_n})^2} . \qquad \text{A.18}$$

With substitution for M_s and C_s, Eq. A.17 becomes

$$B_l = \frac{R_o^2}{f_n^2 (2\Pi\frac{\rho L_n}{A_n})^3 f} + \frac{f}{f_n^2\, 2\Pi\frac{\rho L_n}{A_n}} - \frac{1}{2\Pi\frac{\rho L_n}{A_n}\, f} \qquad \text{A.19}$$

Since

$$\Delta f_n = \frac{R_o}{2\Pi M_s} \qquad \text{A.20}$$

then

$$R_o = 2\Pi M_s \Delta f_n = 2\Pi \Delta f_n \frac{\rho L_n}{A_n} . \qquad \text{A.21}$$

Using the values in Table A.1, R_o is calculated to be 6.65 Ω. By substituting the values in Table A.1 in Eqs. A.7, A.18 and A.19, X_s, G_1, and B_s can be calculated for the various values of A_{vp} (in cm^2) and the frequency of the first formant, F_1. The calculation for G is based on Eq. A.6 and the values of X_c, B_s, and G_1. From Eq. A.14, the bandwidth increase due to the nasal cavity coupling can be calculated for the various values of velopharyngeal opening and first formant frequency. If one assumes that the bandwidth of the non-nasal vowel is 60 Hz, the amount by which the amplitude of the first formant, A1 is lowered is shown in Fig. 2.8.