Simulation and analysis of nasalized vowels based on magnetic resonance imaging data[a)]

Tarun Pruthi[b)] and Carol Y. Espy-Wilson
Speech Communication Laboratory, Institute of Systems Research and Department of Electrical and Computer Engineering, University of Maryland, College Park, Maryland 20742

Brad H. Story
Speech Acoustics Laboratory, Department of Speech, Language and Hearing Sciences, University of Arizona, Tucson, Arizona 85721

(Received 29 July 2005; revised 10 January 2007; accepted 13 March 2007)

In this study, vocal tract area functions for one American English speaker, recorded using magnetic resonance imaging, were used to simulate and analyze the acoustics of vowel nasalization. Computer vocal tract models and susceptance plots were used to study the three most important sources of acoustic variability involved in the production of nasalized vowels: velar coupling area, asymmetry of nasal passages, and the sinus cavities. Analysis of the susceptance plots of the pharyngeal and oral cavities, $-(B_p+B_o)$, and the nasal cavity, B_n, helped in understanding the movement of poles and zeros with varying coupling areas. Simulations using two nasal passages clearly showed the introduction of extra pole-zero pairs due to the asymmetry between the passages. Simulations with the inclusion of maxillary and sphenoidal sinuses showed that each sinus can potentially introduce one pole-zero pair in the spectrum. Further, the right maxillary sinus introduced a pole-zero pair at the lowest frequency. The effective frequencies of these poles and zeros due to the sinuses in the sum of the oral and nasal cavity outputs changes with a change in the configuration of the oral cavity, which may happen due to a change in the coupling area, or in the vowel being articulated. [DOI: 10.1121/1.2722220]

PACS number(s): 43.72.Ar, 43.70.Bk, 43.70.Fq [DOS] Pages: 3858–3873

I. INTRODUCTION

Vowel nasalization has been studied extensively over the last several decades. Although researchers have been successful in finding several acoustical (Bognar and Fujisaki, 1986; Dickson, 1962; Fant, 1960; Fujimura and Lindqvist, 1971; Hattori *et al.*, 1958; Hawkins and Stevens, 1985; House and Stevens, 1956; Maeda, 1982b, c; Stevens *et al.*, 1987) and perceptual (Beddor, 1993; Bognar and Fujisaki, 1986; Hattori *et al.*, 1958; Hawkins and Stevens, 1985; House and Stevens, 1956; Maeda, 1982c) correlates of nasality, automatically extractable acoustic parameters (APs) that work well in a speaker-independent manner still remain elusive. A study to understand the salient features of nasalization, and the sources of acoustic variability in nasalized vowels is, therefore, not just desirable, but in fact needed to find knowledge-based APs to detect vowel nasalization.

Vowel nasalization is not an easy feature to study because the exact acoustic characteristics of nasalization vary not only with the speaker (i.e., with changes in the exact anatomical structure of the nasal cavity), but also with the particular sound upon which it is superimposed (i.e., vowel identity), and with the degree of nasal coupling (Fant, 1960, page 149). Further, even though the articulatory maneuver required to introduce nasalization, a falling velum, is simple, the acoustic consequences of this coupling are very complex because of the complicated structure of the nasal cavity. The human nasal cavity consists of a movable fold called the velum which controls the coupling between the vocal tract and the nasal tract, two asymmetrical nasal passages which end at the two nostrils, and several paranasal cavities (also called sinuses). This combination of degree of velum lowering, asymmetry in the left and right nasal passages, and side branches due to sinuses is the major source of acoustic variability observed in the spectra of nasalized vowels. Several researchers in the past have tried to understand the spectral effects of each of these components in isolation.

Fujimura and Lindqvist (1971) obtained continuous frequency functions of the transfer characteristics of several vowels, stops, nasal consonants and nasalized vowels by exciting the vocal tract by an external sweep-tone signal. They observed the appearance of an extra nasal pole-zero pair around 400 Hz and damping of F2 in the spectra of the nasalized vowels /u/, /o/ and /ɑ/. They also proposed several rules to predict the movement of the poles and zeros with changes in velar coupling area for a lossless system based on the plots of susceptance (imaginary part of the admittance) of the pharyngeal, oral and nasal cavities looking into the respective cavities from the coupling location. Maeda (1993) simulated the spectra for French nasal vowels, and reiterated

[a)]Portions of this work have been presented in Pruthi, T., and Espy-Wilson, C.Y. (2005), "Simulating and understanding the effects of velar coupling area on nasalized vowel spectra," J. Acoust. Soc. Am. **118**(3), p. 2024 (ASA Meeting abstract), and Pruthi, T., and Espy-Wilson, C.Y. (2006), "An MRI based study of the acoustic effects of sinus cavities and its application to speaker recognition," in Proceedings of Interspeech, pp. 2110–2113.

[b)]Author to whom correspondence should be addressed. Electronic address: tpruthi@umd.edu

the rules for pole-zero movements suggested by Fujimura and Lindqvist (1971) for varying coupling areas. He proposed four different kinds of spectral modifications corresponding to the vowels /i/, /u/, /ɛ/, and /ɔ/. Feng and Castelli (1996) viewed nasalized vowels as a dynamic trend from an oral configuration towards an /ŋ/-like configuration. The latter configuration corresponds to the case of just the pharyngonasal tract when the velum falls to such an extent that it blocks off the oral cavity completely. The first two resonance frequencies were used to characterize this pharyngonasal configuration. Simulations of the 11 French nasal vowels with configurations varying between oral and pharyngonasal were presented to illustrate pole-zero evolutions with varying coupling areas. Feng and Castelli (1996) also suggested that the main effect of maxillary sinus was to add complexity to the low-frequency nasal spectrum. However, they did not deal with the effects of the sinuses in detail.

Dang *et al.* (1994) recorded the nasal cavities of six male speakers and one female speaker using magnetic resonance imaging (MRI). They observed a large asymmetry between the left and right nasal passages of the speakers and suggested that this asymmetry leads to the introduction of a pole-zero pair in the spectrum because of the side branch effect. Acoustic modeling of the nasal cavity as two passages confirmed the appearance of the extra pole-zero pair between 2 and 2.5 KHz. Further, simulation of nasal consonant spectra with the inclusion of the maxillary and sphenoidal sinuses suggested that it was necessary to include the paranasal cavities to model the acoustic characteristics of the nasal cavity adequately.

Lindqvist-Gauffin and Sundberg (1976) indicated that a simple two-tract model for the nasal cavity is insufficient to explain the complications seen in the low frequency spectrum of nasals and nasalized vowels obtained by the sweep tone experiments of Fujimura and Lindqvist (1971). They suggested that the observed acoustic effects could, however, be explained if additional shunting cavities (i.e., sinuses) were included in the modeling of the nasal cavity. They also suggested that asymmetries in the two nasal passages, or the sinuses, would introduce additional pole-zero pairs in the spectrum. Maeda (1982b) has shown that the low resonance characteristic of nasalized vowels is due to the sinus cavities. He also found that the low resonance was important to make the nasalized vowels sound more natural. However, he suggested that further corroborations based on acoustic and anatomical data were required to confirm the effects. Dang and Honda (1996) estimated the zero frequencies and the locations of the sinus openings by directly measuring the transmission characteristics of the nasal tract of three subjects. Their results indicated that each of the sphenoidal, maxillary and frontal sinuses introduce their own zeros into the transmission characteristics of the nasal tract. In two out of the three cases considered, the maxillary sinuses accounted for the lowest zeros in the spectrum.

These studies, however, still leave several questions unanswered. While the study by Fujimura and Lindqvist (1971) gave a theoretical basis for predicting the movement of poles and zeros as a function of the coupling area using susceptance plots, they assumed a very simplistic but nonrealistic model for changes in the coupling area. Therefore, this model might be slightly inaccurate in predicting pole/zero movements in a real setting and a more realistic model should be pursued. Further, even though past literature has clearly illustrated that each sinus can introduce its own zero in the spectrum, it still does not tell us how these poles/zeros would move with changes in coupling area or vowel. Moreover, only the studies by Dang *et al.* (1994) and Dang and Honda (1996) were based on real anatomical data collected by MRI.

MRI has become a standard for volumetric imaging of the vocal tract during sustained production of speech sounds (Alwan *et al.*, 1997; Baer *et al.*, 1991; Dang *et al.*, 1994; Matsumura, 1992; Moore, 1992; Narayanan *et al.*, 1995, 1997). Story *et al.* (1996) used MRI to create an inventory of speaker-specific, three-dimensional (3D), vocal tract air space shapes for 12 vowels, three nasals and three plosives. They also imaged the speaker's nasal tract along with his left and right maxillary sinuses and sphenoidal sinuses. The area functions for the nasal cavity were not included in Story *et al.* (1996), but were made available in Story (1995). Hardly any attempts have ever been made to analyze nasalized vowels using MRI data. In this work, therefore, we will use the MRI based area functions recorded by Story (1995) and Story *et al.* (1996) to simulate the spectral effects of vowel nasalization and analyze them in detail.

II. AREA FUNCTIONS

A. Vocal tract

Vocal tract area functions measured with MRI were reported by Story *et al.* (1996) for one adult male speaker (the first author of that study). The area function collection included 12 vowels, three nasal consonants, and three stop consonants. For the present study, area functions for the /i/ and /ɑ/ vowels reported in Story *et al.* (1996) served as the two vocal tract shapes for simulation of vowels. They are shown in the middle (/i/) and bottom (/ɑ/) panels of Fig. 1. Also indicated on these plots are the approximate locations of the coupling of the main vocal tract to the nasal tract under the condition of nasalization. These locations are specified to be 8 cm from the glottis in both cases, as was reported for the nasal consonants of Story *et al.* (1996). The coupling location serves to divide the vocal tract into pharyngeal cavity and oral cavity. The region from glottis to the coupling location is called the pharyngeal cavity, and the region from the coupling location to the lips is called the oral cavity. It is acknowledged that these coupling points may be fairly rough approximations of their actual locations during nasalized vowel production. Furthermore, it is also possible, and perhaps likely, that the vocal tract shape for each of these vowels would be somewhat modified from those shown in Fig. 1 under nasalized conditions.

B. Nasal tract

MRI-based area functions of the nasal system were obtained at the same time as the vowels and consonants. The availability of area functions of the vocal and nasal tracts,

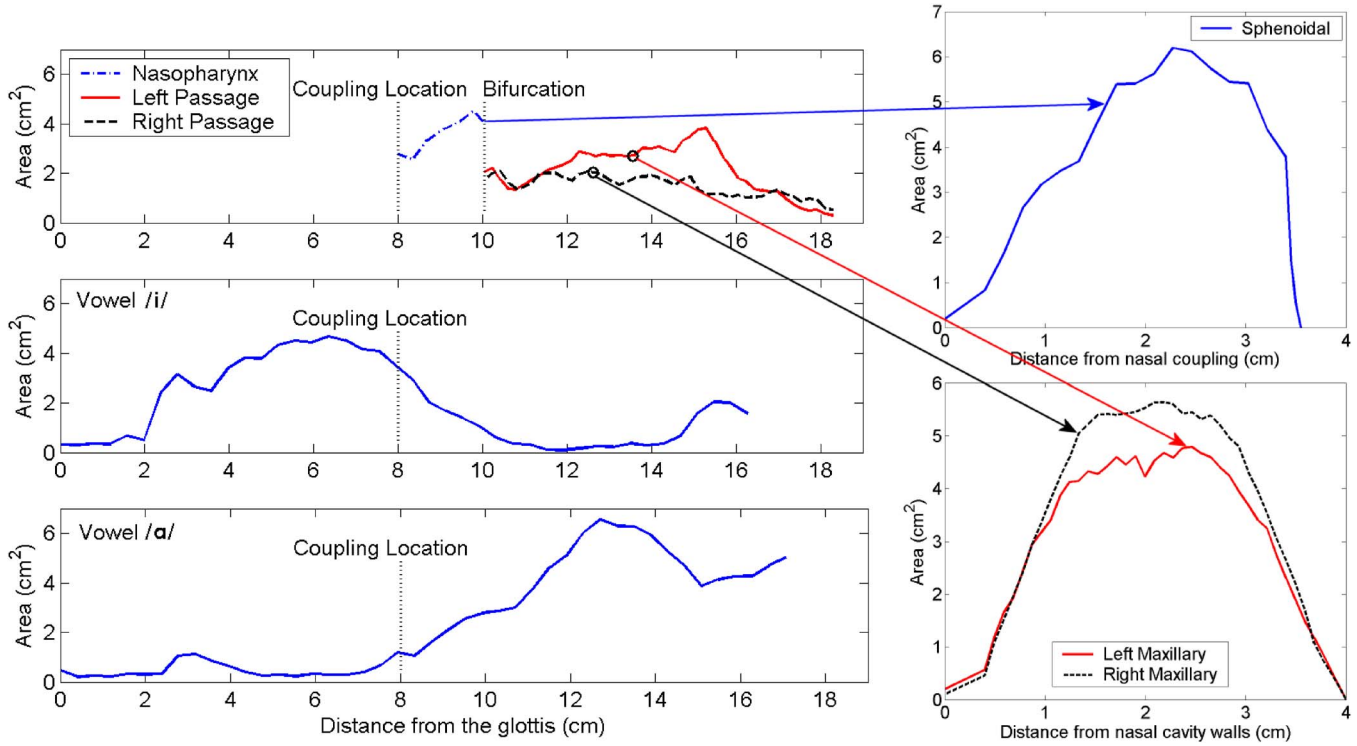

FIG. 1. (Color online) Areas for the vocal tract (oral cavity and pharyngeal cavity), nasal tract, maxillary sinuses and sphenoidal sinus.

both from the same speaker, made this data set particularly attractive for simulating nasalized and nonnasalized vowels.

The MR images of the nasal system used in the present study were acquired in June 1994 at the University of Iowa Hospitals and Clinics. A General Electric Signa 1.5 T scanner was used with the data acquisition mode set to Fast Spin Echo and the scanning parameters set to TE=13 ms, TR =4000 ms, ETL=16 ms, and NEX=2. During normal respiration (no specific speech task), an image set was acquired as a contiguous series of 3-mm-thick coronal image slices extending from the nares (nostrils) to approximately the posterior pharyngeal wall. The field of view for each slice was 26 cm×26 cm which, with a pixel matrix of 256×256, provided a pixel dimension of 1.016 mm/pixel. Two image sets of the speaker's nasal system were obtained. The first captured the *natural* state of the speaker's nasal tract at the specific point in time at which the images were collected. The second image set was obtained after the speaker's nasal system had been treated with oxymetazoline nasal spray (e.g., *Afrin*) to reduce any nasal tissue swelling that may have caused slight congestion. Although the first image set may represent a more natural speaking condition, the second set presumably represented the most open condition of the nasal system and would also be considered the most repeatable for making subsequent audio recordings of the speaker (i.e., because the nasal spray can be reapplied at any time to bring the speaker's nasal system to its most open state). Thus, all subsequent analysis and modeling will be concerned only with the image set based on this *decongested* condition.

The analysis of the image sets was performed with VIDA (Volumetric Image Display and Analysis) which is a general image display and analysis package (Hoffman *et al.*, 1992). Each image set was first subjected to a cubic voxel interpolation so that the spatial resolution was the same in all dimensions (1.015 mm). The airspace within each consecutive coronal slice was then determined with a seeded-region growing algorithm. Two views of a 3D reconstruction of the nasal airspace in the *decongested* condition are shown in Fig. 2. These indicate the natural separation of the nasal system into several distinct passages and cavities. In Fig. 2(a), the inferior portion of the nasopharynx (labeled A) can be seen as a tubular structure that acoustically and aerodynamically couples the main vocal tract to the nasal system when the velum is in a lowered position (as it was during the image collection). As the nasopharynx tube extends superiorly it bifurcates into the left and right nasal passages, the inferior portion of which can be seen in Fig. 2(b) (E_L and E_R). The openings of these passages into the nasopharynx are called the posterior nares. These passages extend anteriorly to their termination at the anterior nares/nostrils (D in Fig. 2(a) and D_L and D_R in Fig. 2(b)). Located at nearly the same point as the bifurcation of the nasopharynx is the connection of the ostia leading into the Sphenoid Sinus (SS) cavity; the cavity can be seen in the upper left part of Fig. 2(a) (labeled B) but the ostia are not visible. Connected to the left and right nasal passages are the Maxillary Sinus (MS) cavities. These can be seen in Fig. 2(b) (C_L and C_R) as the wing-like structures extending laterally from both sides of the nasal system. No visible connection from any part of the nasal system to either the Frontal Sinus (FS) or the Ethmoid Sinus (ES) cavities could be determined. Hence, they were excluded from the image analysis.

Cross-sectional areas of the nasopharynx tube were measured with the same iterative bisection algorithm used for measurement of the vocal tract area functions (Story *et al.*,

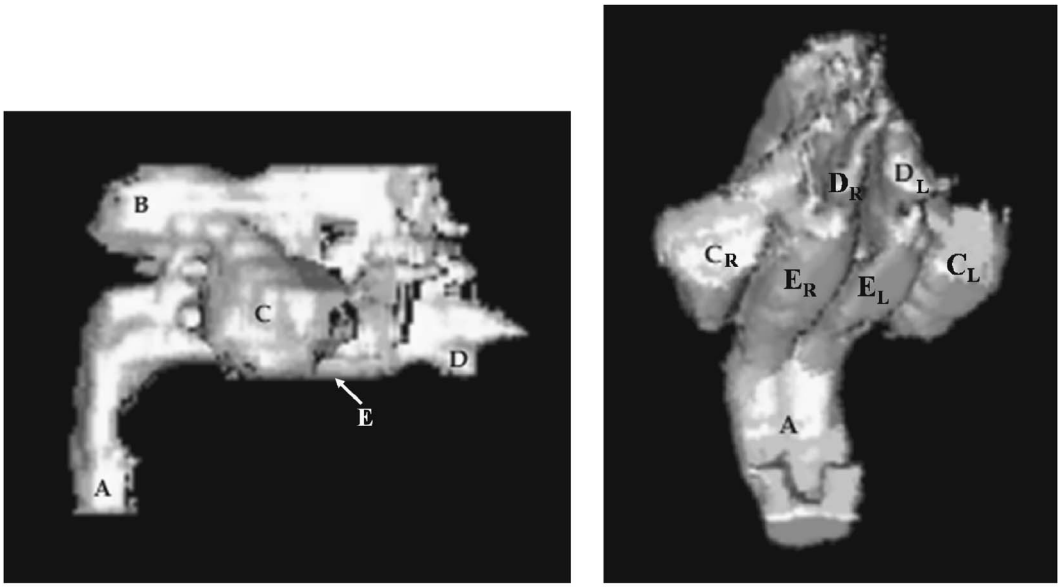

FIG. 2. Two views of the 3D reconstruction of the nasal cavity from the MRI data used in this study. The labels refer to specific regions within the nasal tract. (a) Sagittal projection showing the speaker's right side where *A* is the inferior portion of the nasopharynx, *B* is the sphenoid sinus cavity, *C* is the maxillary sinus, and *D* is the nostril. (b) Nasal tract rotated to show the left and right sides from an inferior view where *A* is again the inferior portion of the nasopharynx, C_L and C_R are the left and right maxillary sinuses, D_L and D_R are the left and right nostrils, and E_L and E_R are the left and right nasal passages, respectively.

1996, page 542). The measured area variation can be seen in the upper panel of Fig. 1 as the section that extends from 8–10 cm from the glottis. Although this is a short (2 cm) section, the iterative bisection technique was needed to traverse the curved portion of the nasopharynx. The 10 cm point in this plot signifies the location at which the nasopharynx bifurcates into the left and right nasal passages. The cross-sectional areas along the length of these passages were measured slice by slice with a region-of-interest (ROI) algorithm, and the resulting area functions are again shown in the upper panel of Fig. 1. Both begin at 10 cm from the glottis, and, due to the bifurcation, it is noted that the cross-sectional areas of the left and right passages at this point sum to nearly the final cross-sectional area of the nasapharynx; their sum is not exactly the same because the measurements of the left and right nasal passages were made at a slightly more anterior location than the final section of the nasopharynx. Both the left and right passages extend to a location shown as 18.25 cm from the glottis. Between approximately 12–17 cm from the glottis, the right nasal passage maintains cross-sectional areas that are about 1–1.5 cm^2 less than those of the left passage, thus producing a left versus right asymmetry. Because these were measured from the *decongested* condition, it is unlikely that the asymmetry is due to congestion. A more likely cause is the speaker's deviated nasal septum that was revealed, in the raw images, to constrict the right passage.

The cross-sectional areas of the SS and the right and left MS were also measured with the ROI algorithm. Anatomically, the sphenoid bone contains two sinus cavities separated by an irregular midline septum (Zemlin, 1998, pages 220, 224). Ostia provide the coupling of these two cavities to the left and right nasal passages. Based on the images collected, however, the separation of the sphenoid into two separate cavities was not visible. Hence, the airspace within the SS was analyzed as if it were one cavity. Consequently, the measurements of the two ostia were summed to provide one coupling port. The variation in cross-sectional area of the sphenoid cavity is shown in the upper right plot of Fig. 1. The *x* axis represents the distance from the point of coupling to the left and right nasal passages. The cross-sectional areas of both maxillary cavities were measured along the extent from their respective anterior to posterior walls, and are shown in the lower right plot of Fig. 1. They are nearly identical in shape, though the left cavity is slightly smaller than the right. The coupling of these cavities to the left and right passages occurs at the locations indicated with open circles in the upper panel of Fig. 1.

III. METHOD

In this study, VTAR (Zhang and Espy-Wilson, 2004), a computer vocal tract model, was used to simulate the spectra for nasalized vowels with successive addition of complexity to the nasal cavity to highlight the effects of each addition. Given the description of area functions in Sec. II, the complete structure of the model of the vocal tract and the nasal tract used in this study is shown schematically in Fig. 3(c). Section IV A analyzes the acoustic changes due to the introduction of coupling between the vocal tract and the nasal tract, and due to changes in the coupling area. Hence, in this section a simplified model of the vocal tract and nasal tract (shown schematically in Fig. 3(a)) is considered. Section IV B analyzes the effects of asymmetry between the left and right nasal passages, and therefore, the model shown in Fig. 3(b) adds the complexity due to nasal bifurcation in the model considered in this section. Section IV C examines the effects of MS and SS on the acoustic spectrum. Hence, the

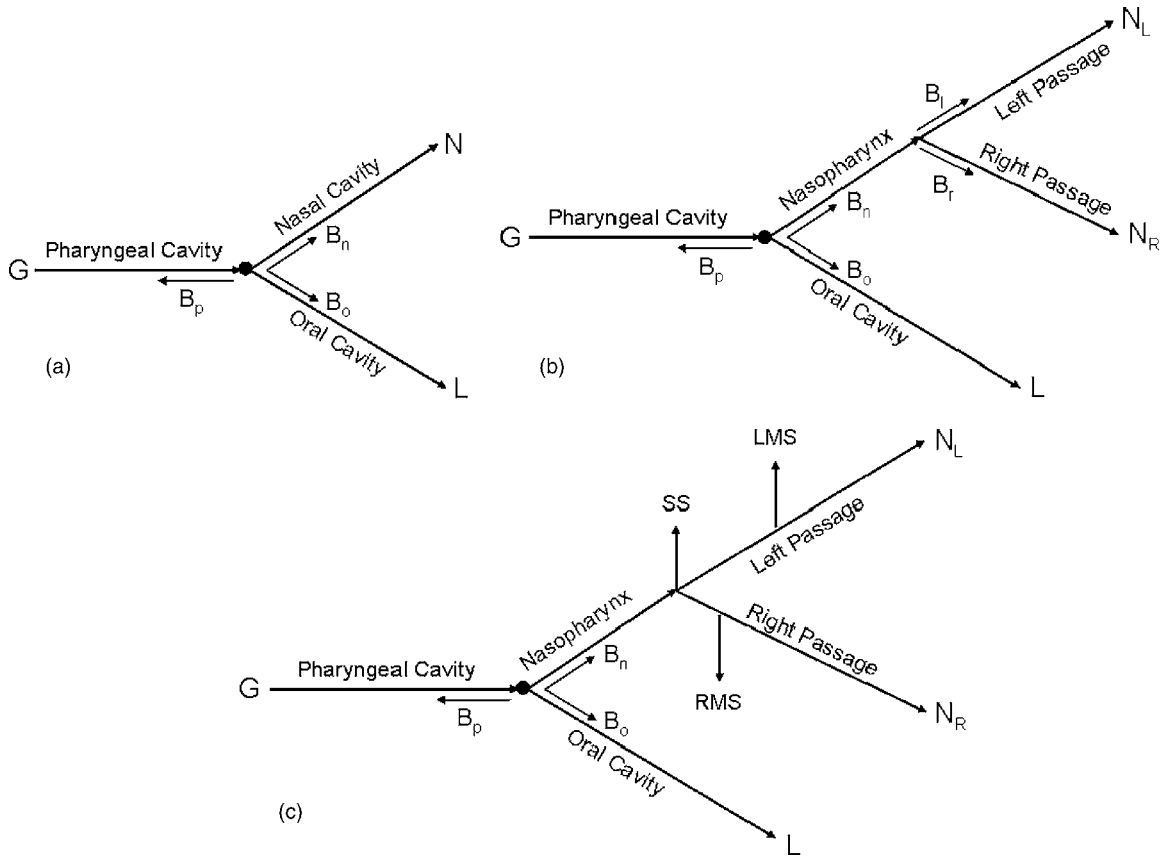

FIG. 3. Structure of the vocal tract model used in this study. (a) Simplified structure used in Sec. IV A, (b) simplified structure used in Sec. IV B, (c) complete structure. G= glottis, L= lips, N= nostrils, N_L= left nostril, N_R= right nostril, RMS= right maxillary sinus, LMS= left maxillary sinus, SS= sphenoidal sinus, B_p= susceptance of the pharyngeal cavity, B_o= susceptance of the oral cavity, B_n= susceptance of the nasal cavity, B_l= susceptance of the left nasal passage, and B_r= susceptance of the right nasal passage. The black dot marks the coupling location.

model shown in Fig. 3(c) is used for simulations in this section.

The nasal cavity data shown in Fig. 1 were combined with the oral cavity data for vowels /i/ and /ɑ/ to obtain the area functions for the nasalized vowels /i/ and /ɑ/. It is assumed that this combination gives an approximate model for nasalized vowels. Two different methods to couple the vocal tract with the nasal tract were considered in this study:

- **Trapdoor coupling method:** The area of the first section of the nasopharynx (of length 0.34 cm) was set to the desired coupling area and no other changes were made to either the areas of the nasopharynx or the areas of the oral cavity. This method of coupling approximates the model used by Fujimura and Lindqvist (1971) where the coupling port is essentially treated as a trap door with variable opening and no effect on the shape of the vocal tract and nasal tract.
- **Distributed coupling method:** The area for the first section of the nasopharynx was set to the desired coupling area and the areas of the rest of the sections of the nasopharynx were linearly interpolated to get a smooth variation in areas (i.e., the coupling was distributed across several sections). The difference between the areas of the sections of the nasopharynx with the given coupling area and the areas of the sections of the nasopharynx with no coupling (0.0 cm^2) was subtracted from the corresponding sections of the oral cavity to model the effect of reduction in the areas of the oral cavity because of the falling velum. This procedure is also illustrated in the flow chart in Fig. 4(a). Figure 4(b) shows an example of the adjusted/new areas of the nasopharynx and the corresponding sections of the oral cavity calculated by this procedure when the coupling area is increased from 0.0 to 1.0 cm^2. Maeda (1982b) and Feng and Castelli (1996) used a similar procedure to model the reduction in oral cavity areas. According to Maeda (1982b), this reduction in the oral cavity area is very important to produce natural sounding nasalized vowels.

In Sec. IV A, both the methods are used for introducing coupling. The coupling areas are varied between 0.0 cm^2 and a maximum value which is limited by the vocal tract area at the coupling location. In the case where the coupling area is equal to the maximum value, the oral cavity is completely blocked off by the velum and sound is output only from the nasal cavity. This maximum value of the coupling area will, henceforth, be referred to as the **maximum coupling area**. Even though this pharyngonasal configuration is interesting in an asymptotic sense (Feng and Castelli, 1996), it should be noted that it is unnatural or nonphysiological in the sense that it would never really happen. A close look at Fig. 1 reveals that although /i/ is more closed than /ɑ/ in the oral

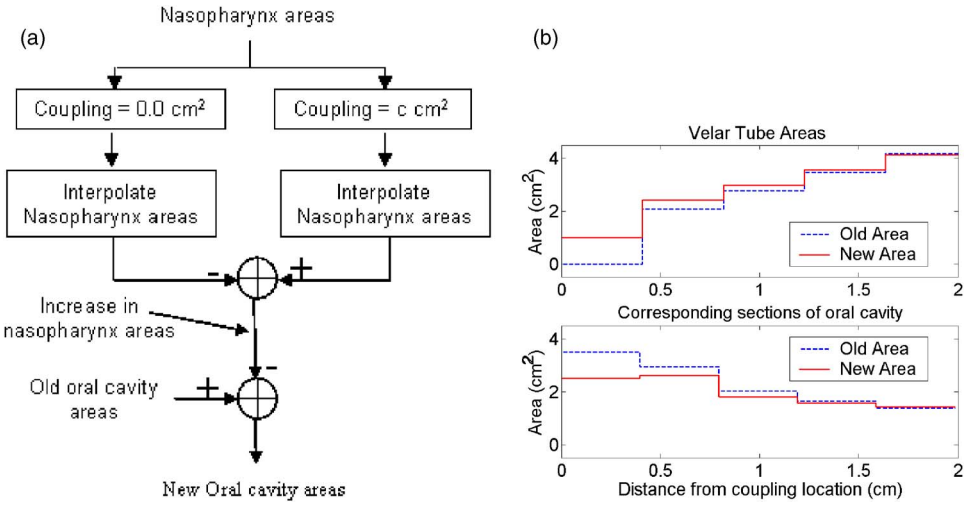

FIG. 4. (Color online) (a) Flow chart of the procedure to get the area functions for the oral and nasal cavity for increase in coupling area, (b) An example to illustrate the changes in nasopharynx areas and areas of corresponding sections of the oral cavity when the coupling area is changed from 0.0 to 1.0 cm^2.

cavity, it is much more open than /ɑ/ at the coupling location. Hence, the possible range of coupling areas is much larger for /i/ than for /ɑ/. Simulations discussed in all other sections of this paper use only the distributed coupling method.

Losses in the vocal tract and nasal tract were not included in the simulations in this work in order to clearly show the effects of each change in terms of poles and zeros. The actual effects of additional poles and zeros introduced into the spectrum due to nasalization might be small because of these losses.

The following convention has been followed throughout the paper: All peaks and dips due to the vocal tract are always referred to as formants and antiformants. All peaks and dips due to the nasal cavity, asymmetrical passages, or the sinuses are referred to as poles and zeros and pole-zero pairs. A set of peaks or dips, some of which are due to the vocal tract and some due to the nasal tract, is also referred to as poles and zeros. This convention has been used partly because in many of the cases, extra peaks and dips due to sinuses and asymmetry may only appear as small ripples in the spectrum because of losses, and because of their proximity to each other. Therefore, it might not be fair to refer to each small ripple as a formant. It must be noted, however, that a peak (dip) in the spectrum is due to a pair of complex conjugate poles (zeros), even though in this paper the pair of complex conjugate poles (zeros) is simply referred to as "pole (zero)." Further, note that the method used to decide whether a peak or a dip is due to either the vocal tract or the nasal tract is described below in Sec. IV A.

IV. VOCAL TRACT MODELING SIMULATIONS

In the simulations below, the effects of the following will be analyzed in detail: (1) Degree of coupling between the nasal cavity and the rest of the vocal tract, (2) asymmetry between the two parallel passages in the nasal cavity, and (3) the Maxillary and Sphenoidal sinuses.

A. Effect of coupling between oral and nasal cavities

Figures 5(a), 5(b), 6(a) and 6(b) show the transfer functions, as calculated by VTAR (see Appendix A for a description of the procedure used to calculate the transfer functions), for the simulated vowels /i/ and /ɑ/ for several different coupling areas. Figure 5 corresponds to the trapdoor coupling method, and Fig. 6 corresponds to the distributed coupling method. The curve for the coupling area of 0.0 cm^2 corresponds to the transfer function of the pharyngeal and oral cavities (from the glottis to the lips) in the absence of any nasal coupling. The curve for the maximum coupling area, as defined in Sec. III, corresponds to the transfer function from the glottis to the nostrils when the oral cavity is completely blocked off by the velum. Note that for the trapdoor coupling method, only the output from the nose is considered for the case of maximum coupling area, even though the oral cavity does not get blocked in this case. Further, note that the transfer functions for the maximum coupling area for the vowels /i/ and /ɑ/ do not match because of differences in the area function of the pharyngeal cavity even though the nasal cavity is approximately the same. The curves for the other coupling areas correspond to the combined output from the lips and the nostrils.

Figures 5(c) and 5(d) show the susceptance plots, as calculated by VTAR (see Appendix A for a description of the procedure used to calculate the susceptance plots), for the combined pharyngeal and oral cavities, $-(B_p+B_o)$, along with the nasal cavity, B_n, for different coupling areas. These, susceptances are calculated by looking into the particular cavity from the coupling location (as illustrated in Fig. 3(a)). As seen in the figures, the susceptance curves have singularities at the frequencies where the corresponding impedance is equal to zero. In Figs. 5(c) and 5(d), B_n and $-(B_p+B_o)$ are plotted for all the coupling areas for which the transfer functions are plotted in Figs. 5(a) and 5(b). With an increase in coupling area, plots for B_n move to the right, while the plot for $-(B_p+B_o)$ does not change since there is no change in the oral and pharyngeal cavity areas. Plots of B_n correspond to areas which vary between the least nonzero coupling area and the maximum coupling area (for, e.g., 0.1–3.51 cm^2 for /i/), since the nasal cavity is completely cut off for 0.0 cm^2 coupling area. Figure 6 gives the same information as Fig. 5 except that Fig. 6 corresponds to the distributed coupling method as described in Sec. III. Thus, in Figs. 6(c) and 6(d), in addition to the movement of the plots of $B_n$to the right, the plots for $-(B_p+B_o)$ move to the left with an increase in the

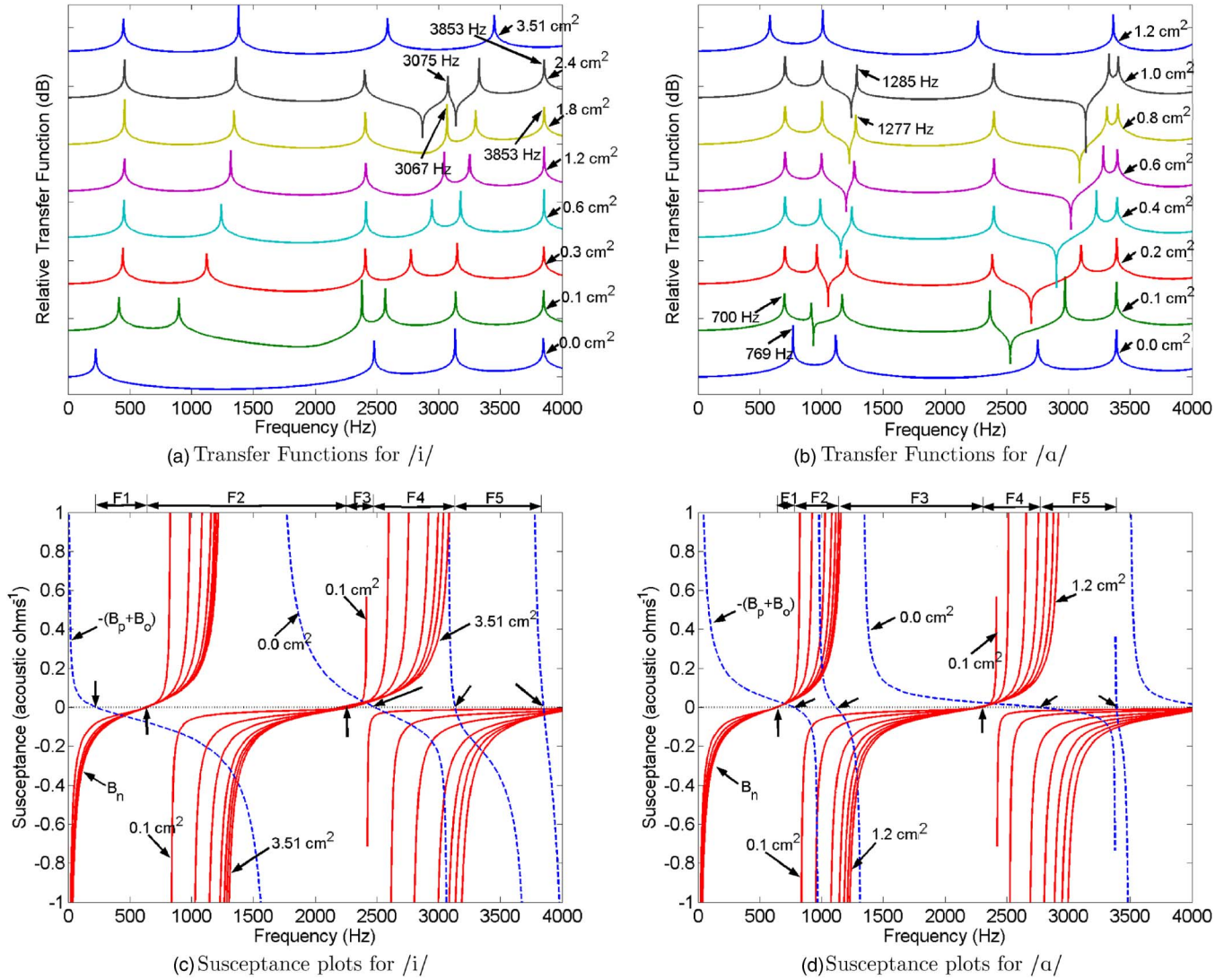

FIG. 5. (Color online) Plots of the transfer functions and susceptances for /i/ and /ɑ/ for the trapdoor coupling method as discussed in Sec. III. (a,b) Transfer functions for different coupling areas, (c,d) plots of susceptances $-(B_p+B_o)$ (dashed) and B_n (solid) for different coupling areas. The arrows above the zero susceptance line mark the frequencies where $B_p+B_o=0$, and the arrows below the zero susceptance line mark the frequencies where $B_n=0$. The markers above the (c) and (d) figures highlight the frequencies between which the different poles can move.

coupling area. The plots for $-(B_p+B_o)$ correspond to areas which vary between 0.0 cm^2 and the second highest coupling area (for, e.g., 0.0–2.4 cm^2 for /i/), since the oral cavity is completely cut off for the maximum coupling area. In Figs. 5(c), 5(d), 6(c) and 6(d) the arrows above the zero susceptance line mark the frequencies where $B_p+B_o=0$. These frequencies are the formant frequencies for the non-nasalized vowels. The arrows below the zero susceptance line mark the frequencies where $B_n=0$. These frequencies are the pole frequencies of the uncoupled nasal cavity. The poles of the combined output from the lips and the nostrils appear at frequencies where the curves for B_n and $-(B_p+B_o)$ intersect (i.e., frequencies where $B_p+B_o+B_n=0$). Note that the frequencies of the poles in Figs. 5(a), 5(b), 6(a) and 6(b) correspond exactly to the frequencies at which the curves for $-(B_p+B_o)$ and B_n for the corresponding coupling areas in 5(c), 5(c), 6(c) and 6(d) respectively intersect.

Let us first consider the trapdoor coupling method. Stevens (1998, page 306), modeled this system as an acoustic mass, $M=\rho\ell_f/A_f$ (where ρ= density of air, l_f= length of the first section, and A_f= area of the first section), in series with the impedance of the fixed part of the nasal cavity, Z_{nf} (see Fig. 7(a)). This lumped approximation is valid until a frequency of 4000 Hz (the maximum frequency in consideration here), because $f=4000\ \text{Hz} \ll (c/\ell_f)=(35{,}000/0.34)=102{,}941$ Hz. Since losses have been removed, the circuit shown in Fig. 7(a) can be solved to obtain

$$B_n = \frac{B_{nf}}{1-\omega B_{nf} M}, \tag{1}$$

where $\omega=2\pi f$ and $B_{nf}=-1/Z_{nf}$. Thus, when $M=\infty$ (that is, the velar port is closed). $B_n=0$, and when $\omega M \ll 1/B_{nf}$, $B_n=B_{nf}$. Further, the zero crossings of B_n do not change with a change in the coupling area, but the singularities of B_n occur at frequencies where $1/B_{nf}=\omega M$. The static nature of the zero crossings can be confirmed in Figs. 5(c) and 5(d). Thus the frequencies of intersections of the susceptance plots

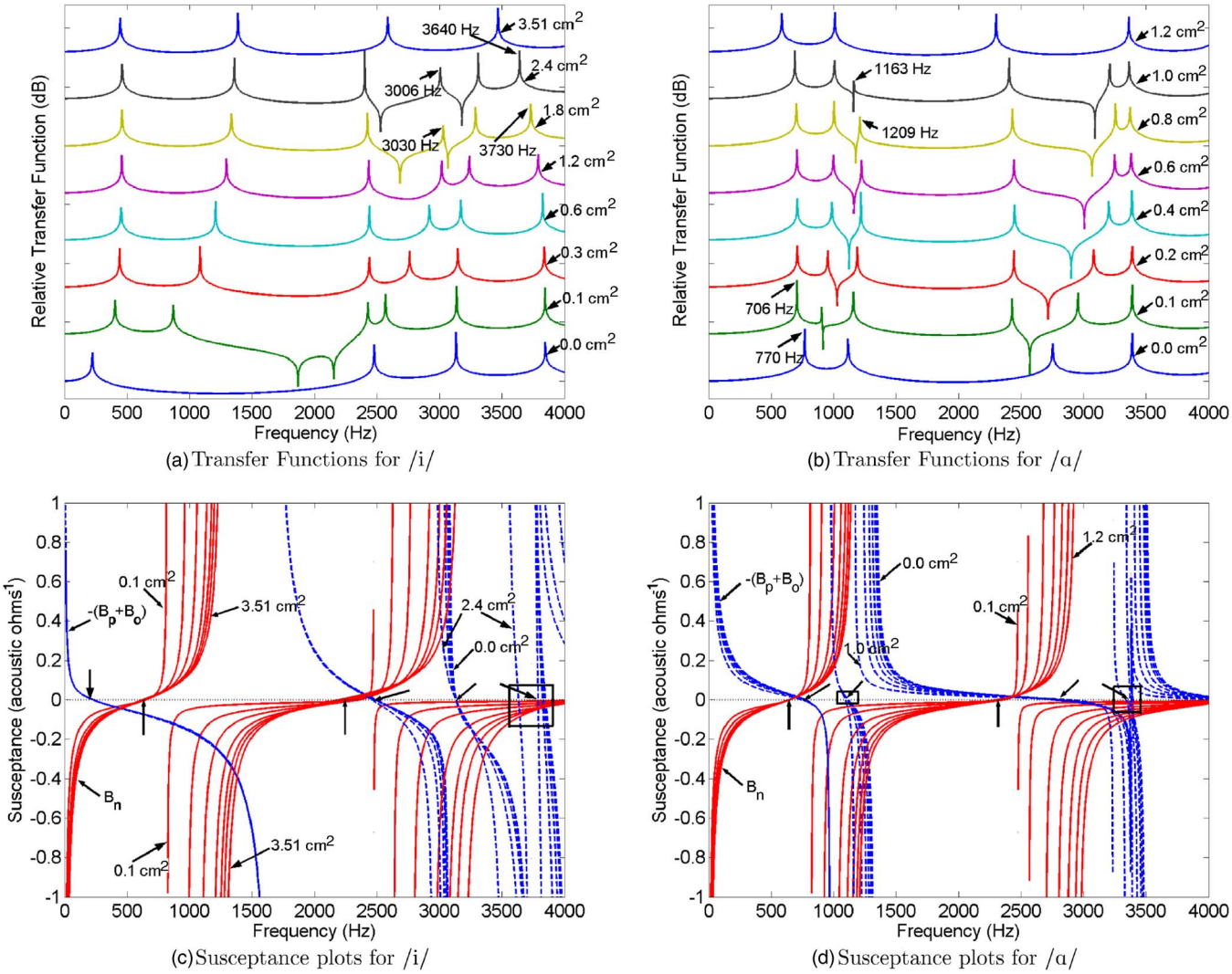

FIG. 6. (Color online) Plots of the transfer functions and susceptances for /i/ and /ɑ/ for the distributed coupling method as discussed in Sec. III. (a,b) Transfer functions for different coupling areas, (c,d) plots of susceptances $-(B_p+B_o)$ (dashed) and B_n (solid) for different coupling areas. The boxed regions highlight the regions where the zero crossings change. The arrows above the zero susceptance line mark the frequencies where $B_p+B_o=0$, and the arrows below the zero susceptance line mark the frequencies where $B_n=0$.

change with the coupling area while the zero crossings remain anchored. A pole in the uncoupled system (decided by the zero crossing of either $-(B_p+B_o)$ or B_n) will move to the frequency of the next intersection of $-(B_p+B_o)$ and B_n in the coupled system. This pole in the coupled system will be referred to as affiliated to the nasal cavity if the pole due to a zero crossing of B_n moved to this frequency, and as affiliated to the vocal tract if a formant due to the zero crossing of $-(B_p+B_o)$ moved to this frequency. For example, in Fig. 5(d), the first pole due to the zero crossing of B_n around 640 Hz moves to approximately 700 Hz in the coupled system, and the pole due to the zero crossing of $-(B_p+B_o)$ around 770 Hz moves to approximately 920 Hz in the coupled system. Thus the zero crossings of the plots for $-(B_p+B_o)$ and B_n determine the order of principle cavity affiliations of the poles in the coupled system. Further, the static nature of the zero crossings, along with the fact that susceptance plots are monotonically increasing functions of frequency, leads to the conclusion that the order of principle cavity affiliations of the poles of the system cannot change with a change in the coupling area (Fujimura and Lindqvist, 1971; Maeda, 1993) because, if for example, the zero crossing of B_n is before the zero crossing of $-(B_p+B_o)$, then the curves for B_n and $-(B_p+B_o)$ would intersect before the zero crossing of $-(B_p+B_o)$.Thus, according to this convention, the order of principle cavity affiliations of the six poles for nasalized /ɑ/ is N, O, O, N, O, and O, where N= nasal cavity, and O= vocal tract (i.e. either oral or pharyngeal cavities). Further

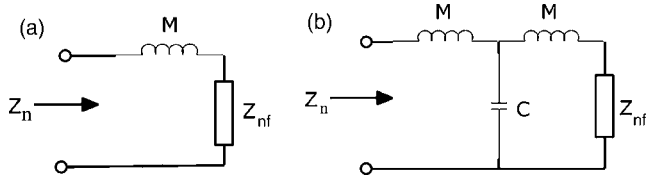

FIG. 7. (a) Equivalent circuit diagram of the lumped model of the nasal cavity. (b) Equivalent circuit diagram of a simplified distributed model of the nasal tract.

$$\frac{dB_n}{dM} = \frac{\omega B_{nf}^2}{(1-\omega B_{nf} M)^2} \geqslant 0, \tag{2}$$

which shows that B_n decreases as M decreases (or coupling area increases) except at frequencies where $B_{nf}=0$ (recall that B_{nf} is a function of f). Since B_n is a monotonically increasing function of frequency except at singularities, Eq. (2) explains the rightward shift of B_n curves with increasing coupling area along with the fact that this shift is not uniform across all frequencies, and it saturates as the coupling area increases (i.e., M approaches zero). Hence, increase in coupling area has the effect of increasing all the pole frequencies (see Figs. 5(a) and 5(b)). Because susceptance plots are monotonically increasing functions of frequency, and the zero crossings are always at the same location, limits can be placed on the movement of each pole. Thus, coupling between two cavities can only cause a pole to move between the frequency location of the zero crossing corresponding to the pole, and the frequency location of the next zero crossing. These limits on pole movements are illustrated by the markers above Figs. 5(c) and 5(d).

The behavior of the susceptance curves described above essentially outlined the rules proposed by Fujimura and Lindqvist (1971) and Maeda (1993, page 150). The rules, however, change for the more realistic case corresponding to the distributed coupling method. This case is shown in Figs. 6(c) and 6(d). The following changes occur for such a case:

- A simplified distributed system model for this case is shown in Fig. 7(b). This model corresponds directly to the lossless transmission line model used for the calculation of susceptance plots by VTAR. Note, however, that this is a simplified model because, in the simulations, several such T sections were concatenated to model the change in velar coupling area since the areas of the whole nasopharynx were changed with a change in the coupling area. In this case, the circuit shown in Fig. 7(b) can be solved to obtain

$$B_n = \frac{B_{nf}(1-\omega^2 MC) + \omega C}{B_{nf}(\omega^3 M^2 C - 2\omega M) - \omega^2 MC + 1}, \tag{3}$$

  where $C=(A_f \ell_f)/(\rho c^2)$. This equation shows that the frequencies of both the zero crossings and the singularities of B_n will change with a change in M and C corresponding to a change in coupling area. A similar analysis for B_o leads to the conclusion that a change in the coupling area will lead to a change in the frequencies of the zero crossings and singularities of $-(B_p+B_o)$. The change will be even more prominent when the areas of not just the first section, but the first few sections change with a change in the coupling area. This is clearly evident in the plots for $-(B_p+B_o)$ for both /i/ and /ɑ/ (see the boxed regions in Figs. 6(c) and 6(d)). Further. Eq. (3) also suggests that the change in the zero crossing frequency should be more prominent at higher frequencies which is again evident in the boxed regions in Figs. 6(c) and 6(d). The zero crossing frequency changes by about 200 Hz for /i/ around 3700 Hz, by about 30 Hz for /ɑ/ around 1100 Hz, and by 50 Hz for /ɑ/ around 3400 Hz. This change in the zero crossing frequency also happens for B_n although the change is much less evident in this case.
- In Figs. 6(c) and 6(d), plots of B_n move to the right, and plots of $-(B_p+B_o)$ move to the left with an increase in the degree of coupling. The zero crossings of B_n and $-(B_p+B_o)$ usually fall in frequency with an increase in the degree of coupling, although no consistent pattern was observed across all instances. Nothing, however, seems to suggest that there cannot be a case where the zero crossings of the two susceptance plots might cross over each other. That is, it is possible that while one of the zero crossings of $-(B_p+B_o)$ was below B_n for a particular coupling area, the zero crossing of B_n might be below the zero crossing of $-(B_p+B_o)$ for another coupling area. Therefore, we speculate that there might be cases where the order of principle cavity affiliations (as defined by the convention above) of the poles of the coupled system does change with a change in the coupling area, This change in the order of principle cavity affiliations is especially possible if the zero crossings of B_n and $-(B_p+B_o)$ are close to each other at a high frequency. Hence, the principle cavity affiliations can only be determined from the susceptance plot for that particular coupling area.
- Pole frequencies need not increase monotonically with an increase in coupling area. Pole frequencies may decrease with an increase in the coupling area when the increase in the nasal cavity area is more than compensated by a reduction in the oral cavity area. For example, the fourth formant for the nasalized /i/ in Fig. 6(a) falls from 3030 Hz at a coupling area 1.8 cm^2 to 3006 Hz at a coupling area of 2.4 cm^2 and the sixth formant falls from 3730 Hz at a coupling area of 1.8 cm^2 to 3640 Hz at a coupling area of 2.4 cm^2. Similarly, the third formant for the nasalized /ɑ/ in Fig. 6(b) falls from 1209 Hz at a coupling area of 0.8 cm^2 to 1163 Hz at a coupling area of 1.0 cm^2. This is an example of reduction in the formant frequency because of a change in the cavity configuration. This reduction was also observed by Maeda (1982b). Contrast this with Figs. 5(a) and 5(b) where formant frequencies never decrease.

It must be noted, however, that the very act of introducing coupling to a side cavity (i.e., changing the coupling area from 0.0 cm^2 to a finite value) cannot cause a pole frequency to decrease. That is so because the susceptance plots are monotonically increasing functions of frequency. Hence, introduction of any kind of coupling can only lead to an increase in the pole frequency. If the pole frequency decreases after the introduction of coupling, then it means that the pole at the lower frequency belongs to the side cavity (owing to a lower frequency of zero crossing for the susceptance plot for the side cavity). One such example is the first pole of the nasalized vowel /ɑ/ in Fig. 6(b). Introduction of coupling to the nasal cavity causes a reduction in the frequency of the first pole from 770 Hz at a coupling area of 0.0 cm^2 coupling to 706 Hz at a coupling area of 0.1 cm^2 coupling, because of a switch in the principle cavity affiliation of the first pole from the oral cavity to the nasal cavity. This switch is evident from the susceptance plot for /ɑ/ in Fig. 6(d) which shows the lower frequency of the zero crossing for B_n.

It is clear from Figs. 6(a) and 6(b) that coupling with the nasal cavity introduces significant changes in the spectrum. In the case of /i/, nasal coupling of 0.1 cm^2 introduces two pole-zero pairs between F1 and F2 of the non-nasalized vowel /i/. In the case of /ɑ/, nasal coupling of 0.1 cm^2 introduces a pole below F1, a zero between F1 and F2, and another pole-zero pair between F2 and F3 of the non-nasalized /ɑ/. With an increase in the coupling area, the distance between the nasal pole and zero increases and the nasal poles become more and more distinct. The nasal zero can get closer to an oral formant and reduce it in prominence. The reduction in the prominence of an oral formant is clearly visible for /ɑ/ in Fig. 6(b) at a coupling area of 0.1 cm^2. In this case, the lowest peak in the spectrum is due to a nasal pole. F1 is now around 900 Hz, however it is reduced in amplitude due to the close proximity of the nasal zero (note that in this case, according to the convention proposed above, the lowest pole of the transfer function is interpreted to be a nasal pole, and the weak second pole due to the presence of the zero nearby as the shifted oral F1), and again around 1200 Hz at a coupling area of 1.0 cm^2, when the nasal zero is close to oral F2. The advantage of using the susceptance plots to study the evolution of poles and zeros with changing coupling area is evident here. These plots provide a systematic method to affiliate the poles to the oral/nasal cavities and follow their evolution with changing coupling areas. Without following this convention there would be no way of judging whether the first pole in /ɑ/ is affiliated with the oral cavity or the nasal cavity.

Figure 6(a) shows that, as the coupling area for /i/ is increased from 0.1 to 0.3 cm^2, the two zeros around 2000 Hz seem to disappear, and then reappear at a coupling of 1.8 cm^2. This can be explained by the fact that the nasalized vowel configuration is equivalent to a parallel combination of two linear time variant systems which, in the case of nasalized vowels, have the same denominator. Therefore, at the output, the transfer function of the system from the glottis to the lips, AR_{lip}, will get added to the transfer function of the system from the glottis to the nostrils, AR_{nose}. The net effect of this addition is that the zeros of the resulting combined transfer function may become obscured. Figure 8 shows the plots for AR_{lip} (top plot), AR_{nose} (middle plot), and $AR_{\text{lip}}+AR_{\text{nose}}$ (bottom plot) for a coupling area of 0.3 cm^2 for /i/. This figure shows that even though the top and middle plots have zeros, the bottom plot does not. Thus, no zeros are seen in the log-magnitude transfer function plots for /i/ at a coupling area of 0.3 cm^2.

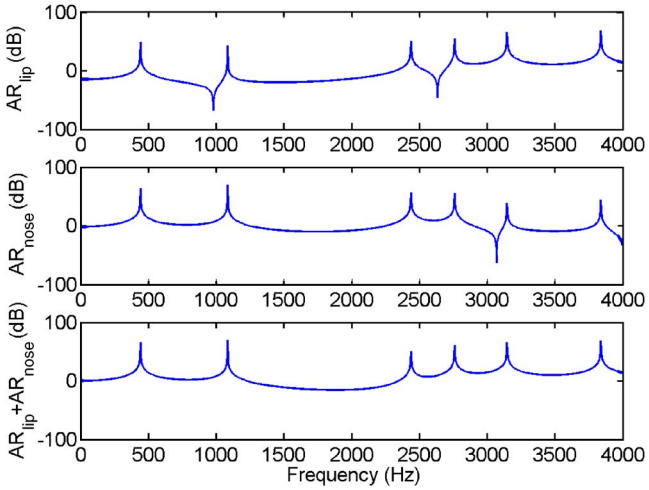

FIG. 8. (Color online) Plots of AR_{lip} (transfer function from the glottis to the lips) (top plot), AR_{nose} (transfer function from the glottis to the nostrils) (middle plot) and $AR_{\text{lip}}+AR_{\text{nose}}$ (bottom plot) at a coupling area of 0.3 cm^2 for vowel /i/.

B. Effect of asymmetry of the left and right nasal passages

When the acoustic wave propagates through two parallel passages, zeros can be introduced in the transfer function because of the following reasons:

- A branching effect, where one of the passages acts as a zero impedance shunt at a particular frequency, thus short circuiting the other passage and introducing a zero in the transfer function of the other passage. A single zero is observed in the combined transfer function of the two passages. The location of this combined zero is in between the frequencies of the zeros of the two passages (Stevens, 1998, page 307).
- A lateral channel effect, which is analogous to the case for /l/. Two kinds of zeros are observed in the transfer function in this case. The first type of zero occurs because of a reversal of phase with comparable magnitudes in the outputs of the two passages due to a difference in the lengths. A difference in the area functions of the two passages because of asymmetry can be treated as being equivalent to a difference in the length. The other type of zero occurs at a frequency corresponding to a wavelength equal to the total length of the two lateral channels (Prahler, 1998; Zhang and Espy-Wilson, 2004).

When the two passages are symmetrical, they can be treated as a single cavity by summing the areas of the two passages since none of the above phenomena would occur for such a case (Prahler, 1998). However, when the two passages are asymmetrical, as will be true generally, the reasons outlined above can lead to the introduction of zeros in the transfer function. It is not reasonable to treat this case as an analogue to the case for /l/ because the two nostrils have different opening areas (as can be seen from Fig. 1), leading to different radiation impedances, and hence, different pressure at the openings. In the case of /l/, the two parallel paths have the same output pressure since the parallel paths combine at the opening (Zhang and Espy-Wilson, 2004). Another important factor is that both the nostrils open into free space, and therefore, there is no more reason to treat them as "lateral channels" than it is to treat the oral and nasal tracts as lateral channels. Thus, it is more reasonable to treat the zero introduced by the asymmetrical nasal passages as being because of the branching effect.

So, the two nasal passages introduce their own zeros at frequencies f_l (frequency at which the susceptance of the right nasal passage $B_r=\infty$), and f_r (frequency at which the susceptance of the left nasal passage $B_l=\infty$). The susceptan-

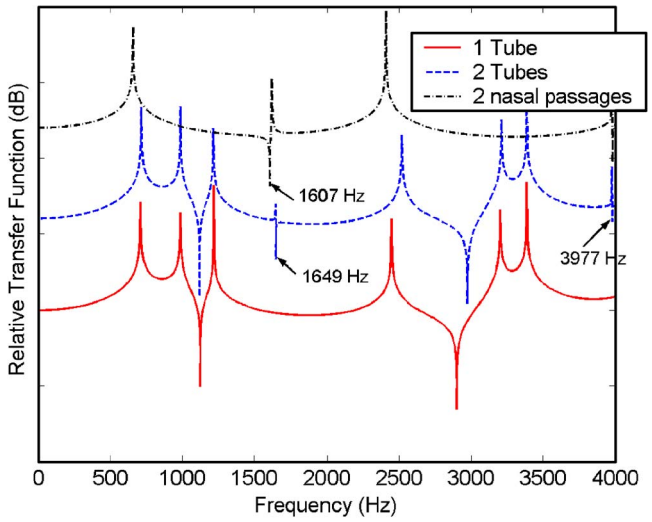

FIG. 9. (Color online) Simulation spectra obtained by treating the two nasal passages as a single tube, and by treating them as two separate passages, for vowel /ɑ/ at a coupling area of 0.4 cm^2. It also shows the transfer function from posterior nares to anterior nares.

ces B_r and B_l are marked in Fig. 3(b). A combined zero (as explained in Stevens (1998, page 307)) will be observed in the combined output of the two nasal passages at frequency f_z given by

$$f_z = f_l \sqrt{\frac{1+\frac{M_l}{M_r}}{1+\left(\frac{f_l}{f_r}\right)^2 \frac{M_l}{M_r}}}, \tag{4}$$

where

$$M_{r/l} = \text{acoustic mass of the right/left passage} = \sum_{\substack{\text{i=all sections of} \\ \text{right/left passage}}} \frac{\rho l_i}{A_i} \tag{5}$$

ρ is the density of air, l_i is the length of the ith section and A_i is the area of the ith section.

Figure 9 shows the transfer functions of the combined vocal tract and nasal tract for the vowel /ɑ/ at a coupling area of 0.4 cm^2, obtained by combining the left and right nasal passages into a single tube of area equal to the sum of the areas of the two tubes, and by treating the left and right passages as two different tubes. The transfer function plots show that the use of two tubes instead of one for the two asymmetrical nasal passages leads to the introduction of additional pole-zero pairs around 1649 Hz and around 3977 Hz. This figure also shows the combined transfer function of just the two nasal passages from the posterior nares to the anterior nares. The location of the first zero in this transfer function is 1607 Hz. Values of f_r and f_l were determined to be 1429 and 1851 Hz, respectively, from the susceptance plots of B_l and B_r. Further, from our calculations $M_l = 0.005\,653$ and $M_r = 0.006\,588$. Using these values in the formula above gives $f_z = 1615$ Hz which is close to the value (i.e., 1607 Hz) obtained from the simulated transfer function. Dang *et al.* (1994) observed the introduction of zeros around 2–2.5 KHz due to two asymmetrical nasal passages.

C. Effect of paranasal sinuses

Figures 10(a) and 10(b) show the transfer functions of the vocal tract with successive addition of the two asymmetrical nasal passages and the right maxillary sinus (RMS), left maxillary sinus (LMS) and SS to highlight the changes in the transfer functions of the nasalized vowels /i/ and /ɑ/ with every addition of complexity to the nasal cavity. The topmost curves in Figs. 10(a) and 10(b) show the transfer functions with all the complexity due to the sinuses and the asymmetrical passages added in. These curves correspond to the model shown in Fig. 3(c). Figures 10(c) and 10(d) show the susceptance plots corresponding to the topmost curves in Figs. 10(a) and 10(b), respectively. A comparison of the B_n curves in Figs. 10(c) and 10(d) with the B_n curves in 6(c) and 6(d) reveals the presence of four extra zero crossings in the B_n curves in Figs. 10(c) and 10(d), thus leading to four extra poles in the transfer function of the uncoupled nasal cavity, one each due to RMS, LMS, SS and the asymmetrical nasal passages. It must be noted that, in reality, the curves of B_n would be even more complicated since the human nasal cavity has eight paranasal sinuses (four pairs) whereas only three have been accounted for here. However, the effects of most of these extra, pole-zero pairs may be small in real acoustic spectra because of the proximity of poles and zeros, and because of losses.

Figures 10(a) and 10(b) clearly show that one extra pole-zero pair appears in the transfer functions of the nasalized vowels /i/ and /ɑ/ with the addition of every sinus. For /i/ the poles are at 580, 664, and 1538 Hz, and for /ɑ/ the poles are at 451, 662, and 1537 Hz for RMS, LMS and SS, respectively. The corresponding zeros are at 647, 717, and 1662 Hz for /i/ and 540, 665, and 1531 Hz for /ɑ/. Note that the pole frequencies due to the sinuses are different for the two vowels because the pole frequencies are decided by the locations where $B_n = -(B_p + B_o)$, and both B_p and B_o are different for the two vowels (see Figs. 10(c) and 10(d)). The pole frequencies due to the sinuses will also change with a change in the coupling area, since, a change in the coupling area corresponds to a change in both B_n and B_o. This observation is in contrast to Stevens (1998, page 306) where it was suggested that sinuses introduce fixed-frequency prominences in the nasalized vowel spectrum. The surprising observation, however, is that even the frequencies of the zeros due to the sinuses in the combined output of the oral and nasal cavities change. The change in the frequencies of the zeros is surprising because sinuses have always been thought of as Helmholtz resonators, branching off from the nasal cavity, which would introduce fixed pole-zero pairs in the nasal vowel spectrum (Dang and Honda, 1996; Dang *et al.*, 1994; Maeda, 1982b; Stevens, 1998). A plausible explanation is as follows:

Consider Fig. 11 which shows a simplified model of the vocal tract and nasal tract. In this figure, the nasal cavity is modeled as a single tube with only one side branch due to a sinus cavity. In this system both U_o/U_s and U_n/U_s will have the same poles (given by frequencies where $B_n = -(B_p + B_o)$),

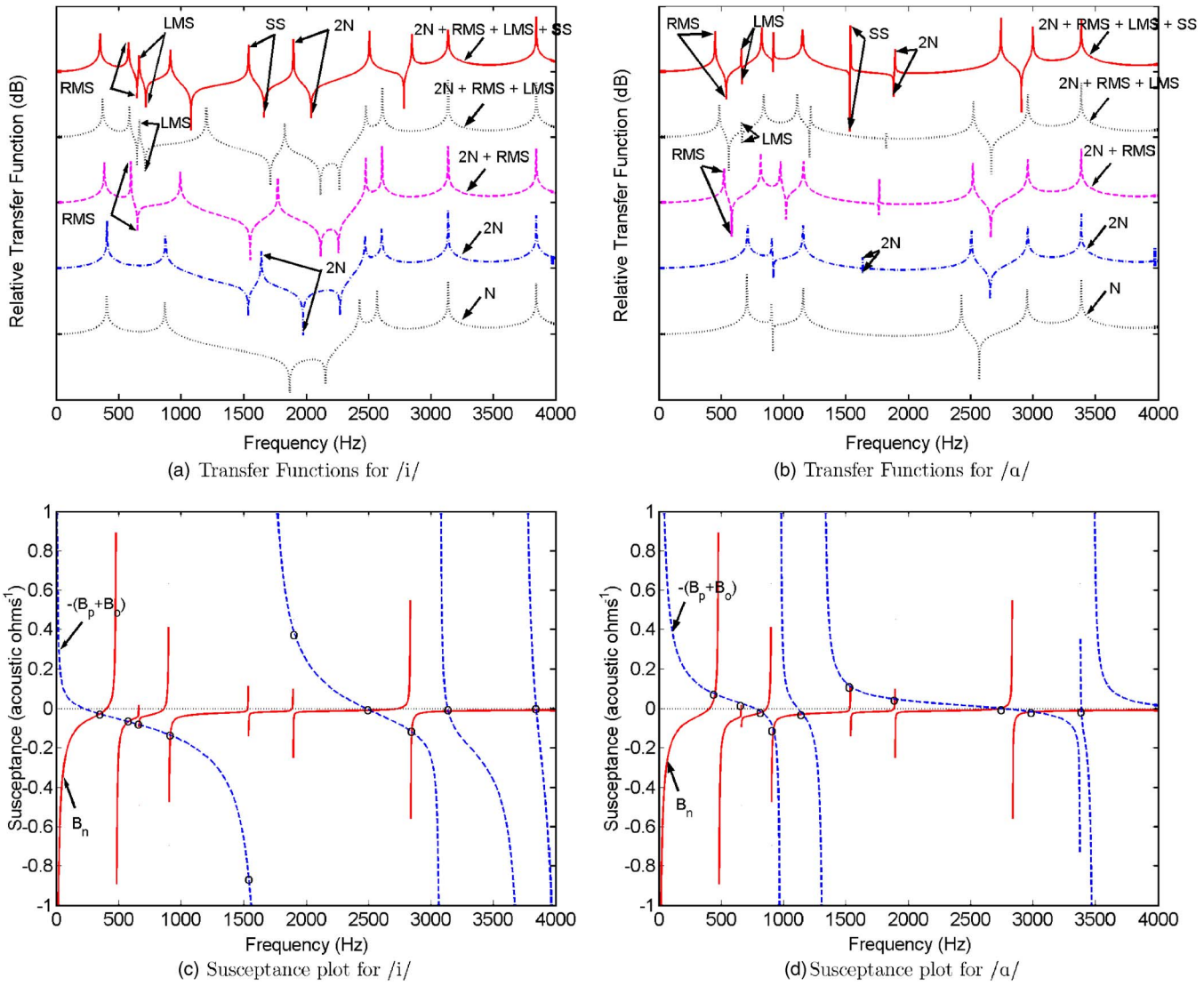

FIG. 10. (Color online) Plots for /i/ and /ɑ/ at a coupling of 0.1 cm^2. (a,b) Transfer functions with successive addition of the asymmetrical nasal passages and the sinuses (N= Nasal Cavity where the areas of the two asymmetrical nasal passages are added and they are treated as a single combined tube, 2N=2 Nasal passages, RMS= right maxillary sinus, LMS= left maxillary sinus, SS= sphenoidal sinus), (c,d) plots of $-(B_p+B_o)$ (dashed) with B_n (solid) for /i/ and /ɑ/ when all the sinuses are included. o's mark the locations of the poles for the coupled system.

but different zeros. Zeros in the transfer function U_o/U_s occur at frequency f_n at which $B_n=\infty$, and zeros in the transfer function U_n/U_s occur at frequency f_o at which $B_o=\infty$, and at frequency f_s at which the susceptance of the side cavity $B_s=\infty$. Then the overall transfer function $T(s)=(U_o+U_n)/U_s$ is given by

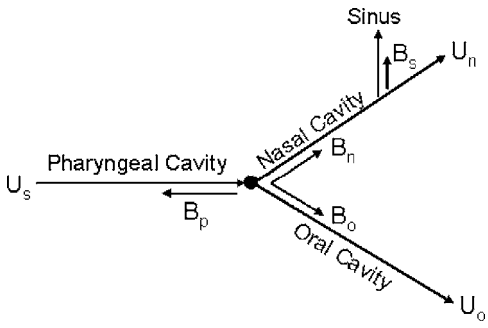

FIG. 11. An illustration to explain the reason for the movement of zeros in the combined transfer function $(U_o+U_n)/U_s$. The black dot marks the coupling location.

$$T(s) = a\frac{(s-s_n)(s-s_n^*)}{s_n s_n^*}P(s) + (1-a) \times \frac{(s-s_o)(s-s_o^*)(s-s_s)(s-s_s^*)}{s_o s_o^* s_s s_s^*}P(s), \quad (6)$$

where $s_n=j2\pi f_n$, $s_o=j2\pi f_o$, $s_s=j2\pi f_s$ and $P(s)$ is an all-pole component that is normalized so that $P(s)=1$ for $s=0$. Further, $a=M_n/(M_o+M_n)$, where M_n is the acoustic mass of the nasal cavity and M_o is the acoustic mass of the oral cavity as marked in Fig. 11 (note that other than the addition of a zero due to the sinus, this analysis is similar to that presented in Stevens (1998, page 307)). Equation (6) shows that the frequencies of the zeros in $T(s)$ will change with a change in either s_n, s_o, or s_s. Note that, s_o and s_n will change with a change in the oral cavity and nasal cavity area functions, respectively. A change in the oral cavity area function can either be due to a change in the vowel being articulated, or due to a change in the velar coupling area. A change in the nasal cavity area function can be due to a change in the velar

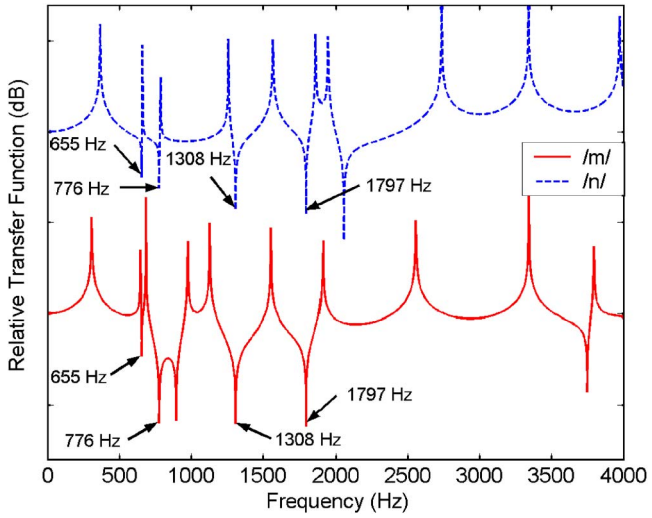

FIG. 12. (Color online) Transfer functions for nasal consonants /m/ (solid, at a coupling area of 1.04 cm^2) and /n/ (dashed, at a coupling area of 1.2 cm^2) showing the invariance of zeros due to the sinuses and the asymmetrical nasal passages. The zero frequencies are 665 Hz (RMS), 776 Hz (LMS), 1308 Hz (SS) and 1797 Hz (asymmetrical passages).

coupling area. The important point here is that even though the sinuses themselves are static structures, what we observe at the microphone is the combined output of the oral and nasal cavities, and the effective frequencies of the zeros due to the sinuses in this combined output can change with a change in the configuration of the oral and nasal cavities. Given this, it would not be correct to say that the effect of the sinus cavities is constant for a particular speaker. Therefore, although the configuration and area functions of the sinuses may be unique for every speaker, the acoustic effects of the sinus cavities on nasalized vowels may not be a very good cue for speaker recognition.

Equation (6), however, also implies that if the output from only one of the cavities, say the nasal cavity, was observed, then the frequencies of the zeros due to the sinuses in the nasal cavity output will be static as long as there is no change in the area function of the sinuses themselves. Therefore, it can be concluded that the frequencies of the zeros due to the sinuses in the nasal consonant spectra will not change regardless of the area functions of the nasal cavity and the oral side branch. The invariance in the frequencies of the zeros due to the sinuses for the nasal consonants is confirmed in Fig. 12 which plots the calculated transfer functions for the nasal consonants /m/ and /n/. The pole locations will still be different depending on the configuration of the vocal tract, and the antiformant due to the oral cavity will also change depending on which nasal consonant is being articulated (see Fig. 12). Thus, for the case of nasal consonants, the acoustic effects of the sinus cavities may be a much more robust cue for speaker recognition. A more detailed study of the implications of this result for speaker recognition was presented in Pruthi and Espy-Wilson (2006). The power spectrum during the nasal consonants was, in fact, used by Glenn and Kleiner (1968) for the purposes of speaker recognition. Using a simple procedure, they were able to obtain an accuracy of 93% for 30 speakers.

Note that Eq. (6) would become much more complicated if terms due to all the other sinuses are added to it. However, the argument presented above is still applicable. Further, this analysis is also directly applicable to the zero due to the asymmetrical nasal passages in the combined output of the oral and nasal cavities. The frequency of this zero in the combined output of the oral and nasal cavities would change with a change in the oral cavity configuration for nasalized vowels, and would not change for nasal consonants (see Figs. 10(a), 10(b), and 12). The analysis presented in Sec. IV B would still remain valid if the sinuses are added in to the model. The only change would be that the frequency location of the zero due to the asymmetrical nasal passages would now be governed by a much more complicated equation of the form of Eq. (6). Further, the analysis for changes in velar coupling areas presented in Sec. IV A would also remain valid, except that B_n would now be a lot more complicated than the B_n shown in Figs. 6(c) and 6(d).

As discussed in Sec. IV A, the principle cavity affiliation of each pole for a particular coupling area can only be determined from the susceptance plot for that particular coupling area. Thus, for the case shown in Fig. 10, the principle cavity affiliations for /i/ are O, N, N, N, N, N, O, N, O, O and the principle cavity affiliations for /ɑ/ are N, N, O, N, O, N, N, N, O, O. Note that, in Fig. 10(c) around 2500 Hz, the zero crossing of $-(B_p+B_o)$ occurs at a lower frequency than the zero crossing of B_n, and in Fig. 10(d) around 2700 Hz, the zero crossing of B_n occurs at a lower frequency than the zero crossing of $-(B_p+B_o)$. This means that in the case of nasalized /i/, the oral F2 always stays around 2500 Hz, and the extra nasal pole moves to 3000 Hz, whereas in the case of nasalized /ɑ/, the oral F3 moves to a frequency around 3000 Hz.

As observed by Chen (1995, 1997), we also find an extra pole due to nasal coupling in the 1000 Hz region. However, this does not mean that that this pole will always be in the vicinity of 1000 Hz since its location can change significantly with a change in the coupling at-area. In the simulations here, this pole was found to go as high as 1300 Hz in frequency for large coupling areas (see Fig. 6(a)). Thus using the amplitude of the highest peak harmonic around 950 Hz as an acoustic cue to capture the extra pole, as proposed by Chen (1995, 1997), might not be appropriate.

In the simulations here, the zeros due to MS were found to be in the range of 620–749 Hz, and the zeros due to SS were found to be in the range of 1527–1745 Hz. These values correspond well with the zero frequencies found by Dang and Honda (1996) which were in the range of 400–1100 Hz for MS, and 750–1900 Hz for SS.

V. GENERAL DISCUSSION

The above analysis has provided critical insight into the changes brought about by nasalization. Listed below are the acoustic changes that have been shown to accompany nasalization, and the reasons behind those changes from the point of view of knowledge gained in this study.

- **Extra poles and zeros in the spectrum**: Several researchers in the past have reported the introduction of extra poles and zeros in the spectrum as the most important and con-

sistent acoustic correlate of nasality (Fant, 1960; Fujimura and Lindqvist, 1971; Hattori *et al.*, 1958; Hawkins and Stevens, 1985; House and Stevens, 1956). Simulations in this study have shown that extra pole-zero pairs are introduced in the spectrum of a nasalized vowel because of (1) coupling between the vocal tract and the nasal tract, (2) asymmetry between the left and right passages of the nasal tract, and (3) the sinuses branching off from the nasal cavity walls. These pole-zero pairs move with a change in the coupling area, and the prominence of an extra pole for a particular coupling area depends on the frequency difference between the pole and an adjacent zero. Previous research has shown that the most prominent effects of these poles are in the first formant region. Hawkins and Stevens (1985) suggested that a measure of the degree of prominence of the spectral peak in the vicinity of the first formant was the basic acoustic property of nasality. It has also been suggested that the low frequency prominence characteristic of nasalized vowels is due to the sinuses (Chen, 1997; Dang and Honda, 1995; Lindqvist-Gauffin and Sundberg, 1976; Maeda, 1982b). Simulations presented above support these views by confirming that the most important change for /i/ is the appearance of the extra nasal pole around 1000 Hz, and for /ɑ/ it is the extra pole below 500 Hz due to MS.

- **F1 amplitude reduction**: Reduction in the amplitude of F1 with the introduction of nasalization has been reported in the past by Fant (1960) and House and Stevens (1956). The above analysis has shown that this effect should be expected more for the case of low vowels than for high vowels; the reason being that for low vowels, the sinus pole can occur below the first formant. With an increase in coupling, the pole-zero pair due to the sinus begins to separate, and as the zero gets closer to F1, the amplitude of F1 falls. For high vowels, however, if the pole-zero pair due to the MS is above F1, then an increase in coupling would only move the zero due to the sinus to a higher frequency, and thus, further awav from F1. This observation supports the view offered by Stevens *et al.* (1987) where it was suggested that the main reason behind the reduction of F1 amplitude was the presence of the nasal zero, not the increase in the bandwidth of poles.
- **Increase in bandwidths**: An increase in F1 and F2 bandwidths has also been cited as a cue for nasalization (Fant, 1960; House and Stevens, 1956). It has been confirmed by simulations that an increase in losses in the nasal cavity has little effect on the bandwidth of formants affiliated with the oral/pharyngeal cavities. Therefore, the bandwidths of all poles need not increase with the introduction of nasalization. However, the poles belonging to the nasal cavity would have higher bandwidths due to higher losses in the nasal cavity because of soft walls and a larger surface area. The bandwidths of other formants might appear to be higher because of an unresolved extra pole lying close by.
- **Spectral flatness at low frequencies**: Maeda (1982c) suggested that a flattening of the nasalized vowel spectra in the range of 300–2500 Hz was the principal cue for nasalization. We now know that the introduction of a large number of extra poles leads to the filling up of valleys between regular oral formants, and the larger prominence of extra poles in the first formant region leads to the spectral flatness effect being more prominent at low frequencies.
- **Movement of the low frequency center of gravity towards a neutral vowel configuration**: Arai (2004), Beddor and Hawkins (1990), Hawkins and Stevens (1985), and Wright (1986) noted a movement in the low frequency center of gravity towards a neutral vowel configuration with nasalization. The analysis above has shown that this effect should be expected both for low and high vowels, since, for low vowels extra poles are introduced below F1, and for high vowels the extra poles above F1 increase in prominence with nasalization. The extra poles would cause the low frequency center of gravity for low vowels to decrease and for high vowels to increase.
- **Reduction in the overall intensity of the vowel**: House and Stevens (1956) observed an overall reduction in the amplitude of the vowel. This reduction is most likely due to the presence of several zeros in the nasalized vowel spectrum as shown in the simulations above.
- **Shifts in pole frequencies**: It must be remembered that the nasal cavity is a large and complicated cavity, and also gives a volume velocity output. Therefore, even a tiny amount of coupling between the oral and nasal cavities can introduce large changes in the spectrum. Poles can suddenly switch their affiliation from the oral cavity to the nasal cavity. Thus, some of the prominent poles might now be affiliated with the nasal cavity instead of the oral cavity. It might as well be these nasal poles which seem to be moving in frequency as the degree of coupling is changed. Further, this effect need not be limited only to the low frequency poles. As seen in the simulations in this study, even F2 and F3 might also change significantly. Such shifts in formant frequencies have been observed in the past by Bognar and Fujisaki (1986), Dickson (1962) and Hawkins and Stevens (1985).

VI. SUMMARY AND CONCLUSIONS

This paper analyzed in detail the three most important sources of acoustic variability in the production of nasalized vowels: velar coupling area, asymmetry of nasal passages, and the sinuses. This analysis was based on real anatomical data, obtained by imaging the vocal tract of one American English speaker using MRI. Area functions obtained from the MRI data clearly show significant asymmetry between the left and right nasal passages, and the left and right maxillary sinuses of this speaker. A computer vocal tract model called VTAR (Zhang and Espy-Wilson, 2004) was used to simulate the spectra for nasalized vowels based on these area functions. A simple extension to VTAR to calculate susceptance plots was proposed and implemented for this study. These susceptance plots have been used extensively in this study to understand the introduction and the movement of poles with changes in the velar coupling area.

The susceptance plots were also used to propose a systematic method to affiliate the poles to either the nasal tract or the vocal tract (similar to Fujimura and Lindqvist (1971))

to follow their evolution with changing velar coupling areas. Analysis of pole movements with changing coupling area showed that the rules concerning the behavior of the poles of the transfer function (as proposed by Fujimura and Lindqvist (1971) and Maeda (1993)) change when a realistic model is assumed for velar coupling. Specifically, it was shown that: (1) the frequency of zero crossings of the susceptance plots changes with a change in the coupling area, and (2) pole frequencies need not shift monotonically upwards with an increase in coupling area. Further, as a consequence of (1), there could be cases where the order of principle cavity affiliations (as defined in this study) of the poles of the coupled system change.

Analysis using two asymmetric nasal passages showed that asymmetry between the left and right nasal passages introduces extra pole-zero pairs in the spectrum due to the branching effect where one of the passages acts as a zero impedance shunt, thus short circuiting the other passage and introducing a zero in the transfer function of the other passage. This result is in agreement with Dang *et al.* (1994). The exact location of the zero in the combined output of the two passages obtained through simulations was found to be a good match with the theoretical frequency calculated by assuming the distribution of the volume velocity into the two passages in a ratio of the acoustic mass of the two passages (as proposed in Stevens (1998, page 307)).

Simulations with the inclusion of maxillary and sphenoidal sinuses showed that each sinus can potentially introduce one pole-zero pair in the spectrum (maxillary sinuses produced the poles lowest in frequency), thus confirming the results of Dang and Honda (1996). The effective frequencies of these poles and zeros due to the sinuses in the combined output of the oral and nasal cavities change with a change in the oral cavity configuration for nasalized vowels. This change in the oral cavity configuration may be due to a change in the coupling area, or due to a change in the vowel being articulated. Thus, it was predicted that even if there was a way to find the frequencies of zeros due to sinuses, it would not be correct to use the effects of sinuses in the nasalized vowel regions as a cue for speaker recognition, although the anatomical structure of the sinuses might be different for every speaker. At the same time, it was also shown that the locations of zeros due to the sinuses will not change in the spectra of nasal consonants regardless of the area functions of the nasal cavity and the oral side branch. Hence, the effects of sinuses can be used as a cue for speaker recognition in the nasal consonant regions. A more detailed study of the application of the acoustic effects of sinus cavities to speaker recognition has been presented in Pruthi and Espy-Wilson (2006).

ACKNOWLEDGMENTS

We wish to thank the anonymous reviewers and Thomas Baer for their helpful comments on an earlier version of this manuscript. We would also like to thank Osamu Fujimura for valuable discussions on some of the issues presented in the paper. This work was supported by NSF Grant No. BCS0236707.

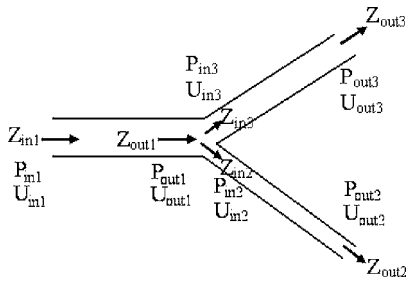

FIG. 13. An illustration to show the procedure to calculate the transfer functions and susceptance plots.

APPENDIX A: PROCEDURE TO CALCULATE THE SUSCEPTANCE PLOTS AND THE TRANSFER FUNCTIONS

The susceptance plots and the transfer functions shown in this work have been obtained using a model of the vocal tract called VTAR (Zhang and Espy-Wilson, 2004). A brief description is as follows:

The input and output pressures (p_{in} and p_{out}) and volume velocities (U_{in} and U_{out}) of a section of the vocal tract are related by the transfer matrix

$$\begin{bmatrix} p_{in} \\ U_{in} \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} p_{out} \\ U_{out} \end{bmatrix}, \tag{A1}$$

where A, B, C, and D depend on the properties of the air and the vocal tract walls and can be calculated by using the transmission-line model (as shown in Zhang and Espy-Wilson (2004)). The transfer function can then be calculated as

$$\frac{U_{out}}{U_{in}} = \frac{1}{CZ_{out} + D}, \tag{A2}$$

where $Z_{out} = p_{out}/U_{out}$. The impedance at a point in the vocal tract can be obtained as a by-product of the transfer function calculation. Hence,

$$Z_{in} = \frac{p_{in}}{U_{in}} = \frac{AZ_{out} + B}{CZ_{out} + D}. \tag{A3}$$

Every branch constitutes a parallel path. Therefore, $Z_{out1} = 1/(1/Z_{in2} + 1/Z_{in3})$ (see Fig. 13). Further, a branch coupling matrix can be used to relate the state variables cross the branching point. Therefore, for Fig. 13

$$\begin{bmatrix} p_{out1} \\ U_{out1} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1/Z_{in3} & 1 \end{bmatrix} \begin{bmatrix} p_{in2} \\ U_{in2} \end{bmatrix} \tag{A4}$$

and

$$\begin{bmatrix} p_{out1} \\ U_{out1} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1/Z_{in2} & 1 \end{bmatrix} \begin{bmatrix} p_{in3} \\ U_{in3} \end{bmatrix}, \tag{A5}$$

where Z_{in2} and Z_{in3} are obtained as shown in Eq. (A3). Thus, the impedance and transfer function at any point in the vocal tract can be found by starting at the output and successively considering each section of the vocal tract without any branches, finding Z_{in}, U_{in} and p_{in} for that section, adding the parallel contribution of any branches, and proceeding in that manner to obtain the required input impedance and the trans-

fer function from the input to that particular output. This procedure can, therefore, be used to obtain Z_{in1}, Z_{in2}, Z_{in3}, U_{out2}/U_{in1} and U_{out3}/U_{in1}. The susceptance B is equal to the imaginary part of the inverse of impedance Z (i.e., the admittance). Thus, plotting the values of impedance/susceptance and the transfer function with respect to frequency generates the impedance/susceptance and the transfer function plots.

To generate good impedance/susceptance plots, losses in the model need to be removed. Losses in the model can be removed by removing the resistive elements from the circuit, which can be achieved by assuming zero resistance due to flow viscosity, zero heat conduction, and infinite wall resistance to remove the loss due to wall vibrations. It should also be noted that for this lossless case,

$$B_{in} = \frac{1}{Z_{in}} = \frac{U_{in}}{p_{in}} = \frac{CZ_{out} + D}{AZ_{out} + B}, \tag{A6}$$

where the susceptance $B_{in}=\infty$ if either $CZ_{out}+D=\infty$ or $AZ_{out}+B=0$. Thus, Eqs. (A2) and (A6) also show that the transfer function does not necessarily have zeros when $B_{in}=\infty$. The transfer function will, however, have zeros when $CZ_{out}+D=\infty$.

Alwan, A., Narayanan, S., and Haker, K. (**1997**). "Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part II. The rhotics," J. Acoust. Soc. Am. **101**(2), 1078–1089.

Arai, T. (**2004**). "Formant shifts in nasalization of vowels," J. Acoust. Soc. Am. **115**(5), 2541.

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**(2), 799–828.

Beddor, P. S. (**1993**). *Phonetics and Phonology: Nasals, Nasalization and the Velum*, (Academic, New York), Chapter on The perception of nasal vowels, pp. 171–196.

Beddor, P. S., and Hawkins, S. (**1990**). "The influence of spectral prominence on perceived vowel quality," J. Acoust. Soc. Am. **87**(6), 2684–2704.

Bognar, E., and Fujisaki, H. (**1986**). "Analysis, synthesis and perception of French nasal vowels," *Proceedings of ICASSP*, Tokyo, pp. 1601–1604.

Chen, M. Y. (**1995**). "Acoustic parameters of nasalized vowels in hearing-impaired and normal-hearing speakers," J. Acoust. Soc. Am. **98**(5), 2443–2453.

Chen, M. Y. (**1997**). "Acoustic correlates of English and French nasalized vowels," J. Acoust. Soc. Am. **102**(4), 2360–2370.

Dang, J., and Honda, K. (**1995**). "An investigation of the acoustic characteristics of the paranasal cavities," in *Proceedings of the XIIIth International Congress of Phonetic Sciences*, Stockholm, pp. 342–345.

Dang, J., and Honda, K. (**1996**). "Acoustic characteristics of the human paranasal sinuses derived from transmission characteristic measurement and morphological observation," J. Acoust. Soc. Am. **100**(5), 3374–3383.

Dang, J., Honda, K., and Suzuki, H. (**1994**). "Morphological and acoustical analysis of the nasal and the paranasal cavities," J. Acoust. Soc. Am. **96**(4), 2088–2100.

Dickson, D. R. (**1962**). "Acoustic study of nasality," J. Speech Hear. Res. **5**(2), 103–111.

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

Feng, G., and Castelli, E. (**1996**). "Some acoustic features of nasal and nasalized vowels: A target for vowel nasalization," J. Acoust. Soc. Am. **99**(6), 3694–3706.

Fujimura, O., and Lindqvist, J. (**1971**). "Sweep tone measurements of vocal-tract characteristics," J. Acoust. Soc. Am. **49**(2), 541–558.

Glenn, J. W., and Kleiner, N. (**1968**). "Speaker identification based on Nasal phonation," J. Acoust. Soc. Am. **43**(2), 368–372.

Hattori, S., Yamamoto, K., and Fujimura, O. (**1958**). "Nasalization of vowels in relation to nasals," J. Acoust. Soc. Am. **30**(4), 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**(4), 1560–1575.

Hoffman, E. A., Gnanaprakasam, D., Gupta, K. B., Hoford, J. D., Kugelmass, S. D., and Kulawiec, R. S. (**1992**). "VIDA: An environment for multidimensional image display and analysis," Proc. SPIE **1660**, 694–711.

House, A. S., and Stevens, K. N. (**1956**). "Analog studies of the nasalization of vowels," J. Speech Hear Disord. **21**(2), 218–232.

Lindqvist-Gauffin, J., and Sundberg, J. (**1976**). "Acoustic properties of the nasal tract," Phonetica **33**(3), 161–168.

Maeda, S. (**1982b**). "The role of the sinus cavities in the production of nasal vowels," in *Proceedings of ICASSP*, Paris, Vol. **2**, pp. 911–914.

Maeda, S. (**1982c**). "Acoustic cues for vowel nasalization: A simulation study," J. Acoust. Soc. Am. **72**(S1), S102.

Maeda, S. (**1993**). *Phonetics and Phonology: Nasals, Nasalization and the Velum* (Academic, New York), Chapter on Acoustics of vowel nasalization and articulatory shifts in French Nasal Vowels, pp. 147–167.

Matsumura, M. (**1992**). "Measurement of three-dimensional shapes of vocal tract and nasal cavity using magnetic resonance imaging technique," in *Proceedings of ICSLP*, Banff, pp. 779–782.

Moore, C. A. (**1992**). "The correspondence of vocal tract resonance with volumes obtained from magnetic resonance imaging," J. Speech Hear. Res. **35**, 1009–1023.

Narayanan, S., Alwan, A., and Haker, K. (**1995**). "An articulatory study of fricative consonants using magnetic resonance imaging," J. Acoust. Soc. Am. **98**(3), 1325–1347.

Narayanan, S., Alwan, A., and Haker, K. (**1997**). "Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data, Part I. The laterals," J. Acoust. Soc. Am. **101**(2), 1064–1077.

Prahler, A. (**1998**). "Analysis and synthesis of American English lateral consonant," Master's thesis, MIT, Cambridge, MA.

Pruthi, T., and Espy-Wilson, C. (**2006**). "An MRI based study of the acoustic effects of sinus cavities and its application to speaker recognition," in *Proceedings of Interspeech* (Pittsburgh), pp. 2110–2113.

Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT Press, Cambridge, MA).

Stevens, K. N., Fant, G., and Hawkins, S. (**1987**). *In Honor of Ilse Lehiste* (Foris Publications), Chapter on Some acoustical and perceptual correlates of nasal vowels, pp. 241–254.

Story, B. H. (**1995**). "Physiologically-based speech simulation using an enhanced wave-reflection model of the vocal tract," Ph.D. thesis. University of Iowa.

Story, B. H., Titze, I. R., and Hoffman, E. A. (**1996**). "Vocal tract area functions from magnetic resonance imaging," J. Acoust. Soc. Am. **100**(1), 537–554.

Wright, J. T. (**1986**). *Experimental Phonology* (Academic, New York) Chapter on The bahavior of nasalized vowels in the perceptual vowel space, pp. 45–67.

Zemlin, W. R. (**1998**). *Speech and Hearing Science: Anatomy and Physiology*, 4th ed. (Allyn and Bacon, Boston).

Zhang, Z., and Espy-Wilson, C. Y. (**2004**). "A vocal-tract model of American English /l/," J. Acoust. Soc. Am. **115**(3), 1274–1280.