



An MRI based Study of the Acoustic Effects of Sinus Cavities and its Application to Speaker Recognition

Tarun Pruthi, Carol Y. Espy-Wilson

Institute of Systems Research and Dept. of Electrical and Computer Engg.
University of Maryland, College Park, MD 20742, USA

{tpruthi, espy}@umd.edu

Abstract

The goal of this paper is to explore the effects of changes in velar coupling area and oral cavity configuration on the poles and zeros introduced in the nasalized vowel and nasal consonant spectra due to the sphenoidal and maxillary sinuses. MRI data for the vocal tract and nasal tract of one speaker was used to simulate the spectra of the nasalized vowels /a, æ, i, u/, and nasal consonants /m, n/ with different coupling areas. It is shown that during nasalized vowels, the frequencies of both poles and zeros due to the sinuses change with a change in the velar coupling area or the vowel. It is also shown that during nasal consonants, the zero frequencies are constant, and the pole frequencies are more stable as compared to nasalized vowels. This study, therefore, corroborates the use of nasal consonant spectra for speaker recognition and raises doubts on the potential benefits of using nasalization during vowels for that purpose.

Index Terms: speaker recognition, nasal, sinus, MRI.

1. Introduction

The nasal cavity is probably the most complicated structure involved in the production of speech. Unlike the oral cavity, the nasal cavity is divided into two parallel passages which end with the two nostrils. The nasal cavity also has several paranasal cavities called sinuses. Humans have four kinds of sinuses: Maxillary Sinus (MS), Frontal Sinus (FS), Sphenoidal Sinus (SS) and Ethmoidal Sinus (ES). These sinuses are connected to the main nasal passages through small openings called *ostia*. Coupling between the nasal tract and the vocal tract (oral cavity and pharyngeal cavity) is controlled by a movable fold called the *velum*. It has been shown that the asymmetry between the two nasal passages can introduce extra poles and zeros in the acoustic spectrum [1]. It has also been shown that the maxillary sinuses account for the lowest pole-zero pair seen in the acoustic spectrum (especially for low vowels) when nasalization is introduced [2, 3], and they are also very important in making speech sound nasal [4].

Despite several studies, the exact dynamics of the poles and zeros due to the sinuses are unclear. In this study, MRI data for the vocal tract and nasal tract of one speaker recorded by Story et al [5, 6] was used to simulate the spectral effects of SS and MS (since these were the only two sinuses for which data was recorded). This study is focused towards understanding the movement of the poles and zeros due to the sinuses with a change in the velar coupling area and the oral cavity configuration. Four vowels (/a, æ, i, u/) and two nasal consonants (/m, n/) were considered in this study.

Analysis of MRI data shows that not only the frequencies of the poles, but also the frequencies of the zeros due to sinuses dur-

ing the nasalized vowel regions change with a change in the velar coupling area and a change in the vowel. The frequencies of the zeros due to the sinuses, however, stay at the same location during nasal consonant regions. Several researchers in the past have shown the effectiveness of the nasal consonantal regions for speaker recognition. The power spectrum during the nasal consonant regions was used in [7] for the purposes of speaker recognition. Features extracted from nasal consonant spectra were also used in [8] for speaker recognition. In another paper [9], coarticulation between the nasal /m/ and the following vowel was used as a cue for speaker recognition. The authors showed that using their coarticulation measure worked better than using the nasal spectrum alone. Other studies on the relative speaker discriminating properties of phonemes [10, 11, 12, 13] have shown that nasals and vowels perform the best. Although several researchers have shown that nasal consonant regions give reliable cues for speaker recognition, no one has used nasality during the vowel regions as a cue. In light of the analysis in this paper, a question arises: Does nasalization during vowels provide a good cue for speaker recognition?

The rest of the paper is organized as follows: Section 2 gives details of the procedure used for MRI simulations. Simulated spectra for nasalized vowels /a, æ, i, u/, and nasal consonants /m, n/ with different coupling areas are presented in Section 3 along with an analysis of the movement of poles and zeros due to the sinuses. Section 4 outlines the most important conclusions of this study.

2. Method

In this experiment, VTAR [14], a computer vocal tract model, was used to simulate the spectra for nasalized vowels and nasal consonants. The nasal cavity data shown in Figure 1 was combined with the oral cavity data for the vowels /a, æ, i, u/ and the consonants /m, n/ to get the area functions for the nasalized vowels and the nasal consonants. Note that the MRI data for the nasal tract was recorded during normal breathing, and it will be assumed in this study that combining the data for the oral tract and the nasal tract gives an approximate model for the nasalized vowels and the nasal consonants. The coupling area was varied between 0.0 cm^2 and a maximum coupling area which is limited by the oral cavity area at the coupling location. Changes in coupling area were achieved as follows: the area for the first section of the *nasopharynx* (the small region of the nasal cavity before bifurcation into the left and right passages) was made equal to the velar coupling area. The areas of the rest of the sections of the nasopharynx were linearly interpolated to get a smooth variation in areas. The difference in the areas of the sections of the nasopharynx with the desired coupling area

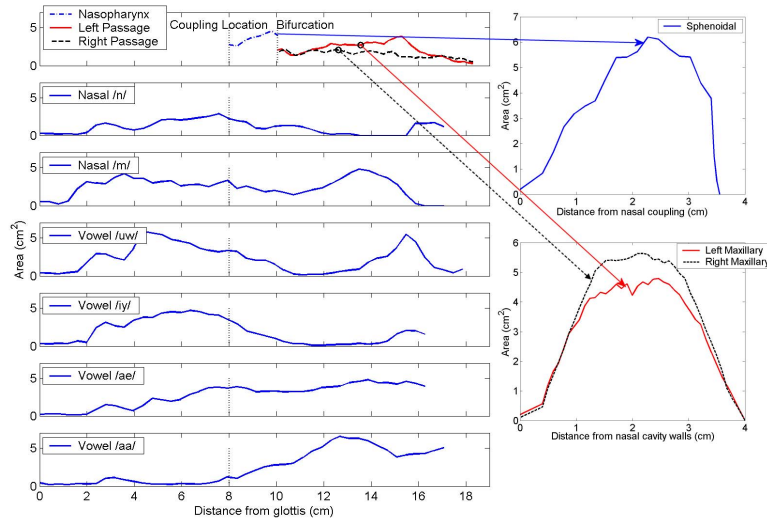


Figure 1: Areas for the oral cavity for vowels /a, æ, i, u/ and consonants /m, n/, nasal cavity, maxillary sinuses and sphenoidal sinus.

and the areas of the sections of the nasopharynx with no coupling (0.0 cm^2) was subtracted from corresponding sections of the oral cavity to model the effect of reduction in the areas of the oral cavity because of the falling velum. According to [4], this reduction in the oral cavity area is very important to get natural sounding nasalized vowels. Note that, although /æ/, /i/ and /u/ are more closed than /a/ in the oral cavity, they are much more open than /a/ at the coupling location. Hence, the possible range of coupling areas is much larger for them. Losses in the vocal tract and nasal tract were not included in the simulations in order to clearly show the frequencies of the 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.

3. MRI Simulations

Figures 2a-d show the transfer functions for different vowels for different coupling areas. Plots for coupling area = 0.0 cm^2 correspond to the non-nasalized vowels. Plots for the maximum coupling areas correspond to the case when the oral cavity is completely blocked off by the falling velum and the speech is output only from the nasal cavity. Thus, this is effectively like the case for the velar nasal consonant /ŋ/. The figures show significant changes in the spectra with the opening of the velar port. Five extra pole-zero pairs are introduced below 2500 Hz in the spectra with velar coupling areas lying in between 0.0 cm^2 and the maximum coupling area. According to our simulations and analysis based on susceptance plots, the extra pole-zero pairs whose frequencies are marked in Figures 2a-d (in the order of increasing frequencies) are due to: the Right Maxillary Sinus (RMS), the Left Maxillary Sinus (LMS), the SS and the asymmetry of the nasal passages. Note that the frequencies of the poles change with a change in the velar coupling areas, and also with a change in the vowel being articulated (see Figures 2a-d). This happens because the pole frequencies are decided by the locations where $B_n = -(B_p + B_o)$ (B_n = susceptance of the nasal cavity, B_p = susceptance of the pharyngeal cavity and B_o = susceptance of the oral cavity), and B_o and B_p change with a change in the vowel. B_o and B_n also change because of a change

in the area functions for the oral and nasal cavities due to a falling velum. This is in contrast to [3, Page 306] where it was suggested that sinuses introduce fixed-frequency prominences in the spectra of nasalized vowels and nasal consonants. The more interesting observation, however, is that even the frequencies of the zeros due to the sinuses change with a change in the velar coupling area and with a change in the vowel (see Figures 2a-d). A plausible explanation is as follows:

Consider Figure 3 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 $B_n = -(B_p + B_o)$), but different zeros [3, Page 306]. Zeros in the transfer function U_o/U_s occur when $B_n = \infty$, and zeros in the transfer function U_n/U_s occur when either $B_o = \infty$, or when the susceptance of the side cavity $B_s = \infty$. Without any loss of generality, let us assume for illustration purposes, that all these zeros are real. The transfer functions are given by:

$$\frac{U_n}{U_s} = \frac{(s - z_o)(s - z_s)}{D(s)}, \quad \frac{U_o}{U_s} = \frac{(s - z_n)}{D(s)} \quad (1)$$

$$\frac{(U_n + U_o)}{U_s} = \frac{s^2 - s(z_o + z_s - 1) + (z_o z_s - z_n)}{D(s)} \quad (2)$$

or,

$$\frac{(U_n + U_o)}{U_s} = \frac{(s - \alpha)(s - \beta)}{D(s)} \quad (3)$$

where, $D(s)$ is the common denominator, and α and β are obtained by the solution of the quadratic polynomial in the numerator of Equation 2. Clearly, α and β will change with a change in either z_o , z_s or z_n . Note that, z_o and z_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 coupling area. Therefore, the

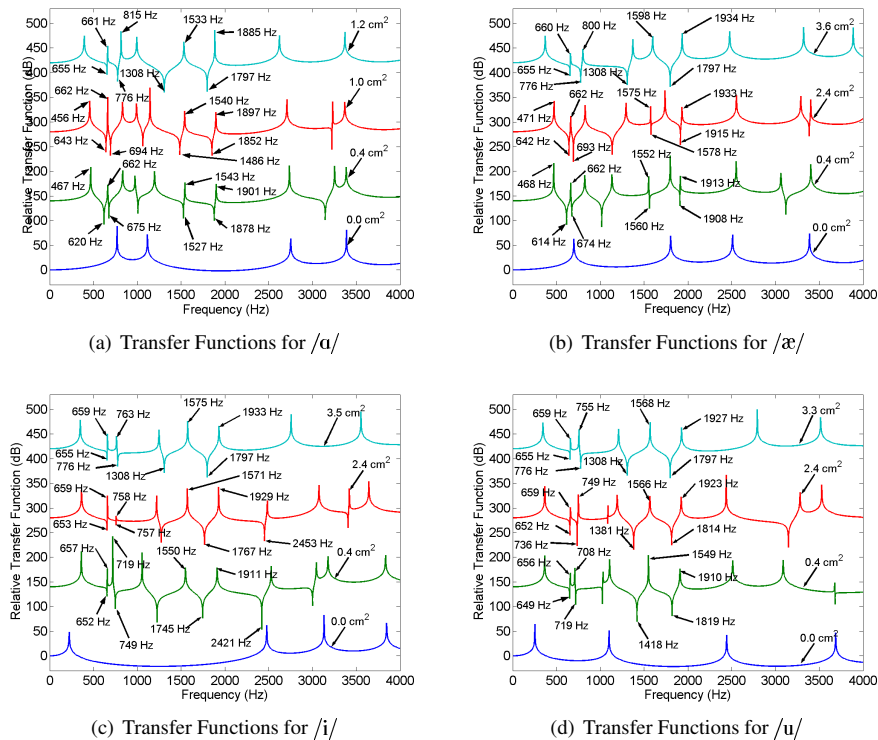


Figure 2: Plots of the transfer functions /a/, /æ/, /i/ and /u/ for different coupling areas.

effective frequency of the zero due to a sinus in the sum of the oral and nasal cavity outputs can vary with a change in the vowel and the coupling area. Thus, even though the sinus cavities themselves don't change, their effects on the nasalized vowel spectrum can be very different depending on the exact configuration in the oral and nasal cavities. This is in contrast to [3, Page 310] where it was suggested that sinuses introduce some fixed pole-zero pairs in the transfer function of a nasalized vowel.

Equations 1-3 also imply 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 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 will not change for nasal consonants, regardless of the area functions of the nasal cavity and the oral side branch. The invariance in the frequencies of zeros due to sinuses for nasal consonants is confirmed in Figure 4 which plots the transfer functions for the nasal consonants /m/ and /n/. According to our simulations and the analysis of susceptance plots, the pole-zero pairs whose frequencies are marked in Figure 4 (in the order of increasing frequencies) are due to RMS, LMS, SS and asymmetrical passages for /m/, and due to RMS, LMS, SS, and asymmetrical passages for /n/. Note that, the pole frequencies still change with a change in the velar coupling area and the nasal consonant, and the antiformant due to the oral side branch also changes with a change in the nasal consonant being articulated. Also note that the exact same analysis, as presented above for sinuses, would also be applicable for the pole-zero pair due to the asymmetrical nasal passages. The zero due to the asymmetry between the left and right nasal passages would be stationary for nasal consonants and variable for

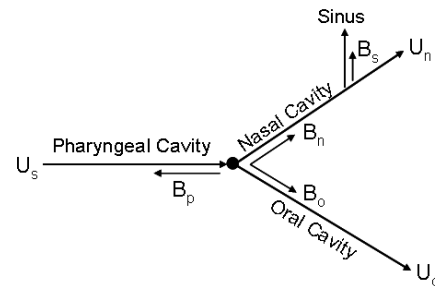


Figure 3: A simplified model of the vocal tract and nasal tract to explain the reason for the movement of zeros in the combined transfer function $(U_o + U_n)/U_s$.

nasalized vowels (see Figures 2a-d and 4).

As shown in Figures 2a-d, the frequencies of the zeros are exactly the same for the four vowels for maximum coupling area. The frequencies of the zeros due to the sinuses are also constant in the simulations for the two nasal consonants /m/, /n/, and are equal to the frequencies recorded for the curves for maximum coupling area in Figures 2a-d, which correspond approximately to the case for velar nasal consonant /ŋ/. Further, even though in Figure 4 the frequencies of the poles due to the sinuses and the asymmetrical passages change with a change in the coupling area and the nasal consonant, the change is not large and is localized to a small region in frequency (the maximum change observed across coupling areas and nasal consonants is 113 Hz). This localized region will depend on the exact area function of a particular person's

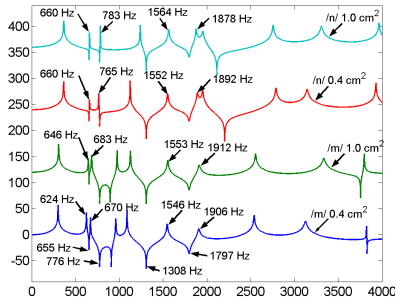
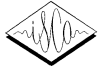


Figure 4: Plots of the transfer functions for /m/ and /n/ for different coupling areas. The zeros due to the sinuses and the asymmetrical passages are always at exactly the same frequencies, and therefore, are marked only once.

nasal cavity, and this area function will be fixed for a particular speaker since there are no moving parts in the nasal cavity (barring the possible changes due to the condition of mucous membrane). The possible variations in the frequencies of poles and zeros due to the sinuses and the asymmetrical nasal passages, however, is much higher for nasalized vowels (In Figures 2a-d the maximum change observed is 203 Hz for poles and 639 Hz for zeros). The variation is further complicated by the large number of vowels in any language. Hence, it could be much more difficult to extract acoustic cues corresponding to the anatomy of a speakers' nasal tract from nasalized vowel regions. Given this, it is our belief that the duration and degree of nasal coupling might be much better parameters to be extracted from nasalized vowel regions for speaker recognition since this would help in separating hypernasal and hyponasal speakers from normal speakers.

4. Conclusions

MRI data for the vocal tract and the nasal tract of one speaker was used to simulate the effects of the sinuses on nasalized vowel spectra and understand the movement of poles and zeros due to the sinuses with changes in the velar coupling area for the vowels /u, æ, i, u/, and the nasal consonants /m, n/. This study clearly supports the use of nasal consonants for speaker recognition and gives insights into why nasal consonantal regions should be good for speaker recognition. The relative stability of the nasal consonant spectra as compared to other phonemes is evident from the exact same frequencies of zeros due to sinuses, and only localized changes in the frequencies of the poles belonging to the nasal cavity. However, the possibility of wide variations in the spectra of nasalized vowels raises the following question: Is nasalization during the vowel regions a good cue for speaker recognition? The analysis presented suggests that it may not be the case, since the pole and zero frequencies due to the sinuses vary with a change in vowel or the coupling area even though the nasal tract anatomy is the same for the same speaker. Changes in the nasal cavity areas due to the shrinking and swelling of the speaker's mucous membrane would only complicate the picture for both nasal consonants and nasalized vowels. It is also proposed that it may be more beneficial to use the duration and degree of nasal coupling during nasalized vowels for the purposes of speaker recognition, rather than the acoustic properties of the static nasal tract anatomy.

The most obvious problem in this study is that the MRI data for only one speaker has been used. This analysis would be much more compelling if it was possible to look at the data for several other speakers. However, collection of MRI data is both time consuming and costly, and this was the only data available. Further, even though this study used MRI data for just one speaker's vocal tract and nasal tract, it has given important insights into the dynamics of nasality both during the nasalized vowels, and during the nasal consonantal regions.

5. Acknowledgments

We would like to thank Brad Story for providing us the MRI data. This work was supported by NSF grant no. BCS0236707.

6. References

- [1] J. Dang, K. Honda, and H. Suzuki, "Morphological and acoustical analysis of the nasal and the paranasal cavities," *J. Acoust. Soc. Am.*, vol. 96, no. 4, pp. 2088–2100, 1994.
- [2] J. Dang and K. Honda, "Acoustic characteristics of the human paranasal sinuses derived from transmission characteristic measurement and morphological observation," *J. Acoust. Soc. Am.*, vol. 100, no. 5, pp. 3374–3383, 1996.
- [3] K.N. Stevens, *Acoustic Phonetics*, MIT Press, Cambridge, Massachusetts, 1998.
- [4] S. Maeda, "The role of the sinus cavities in the production of nasal vowels," in *Proceedings of ICASSP*, 1982b, Vol. 2, pp. 911–914.
- [5] B. H. Story, *Physiologically-based speech simulation using an enhanced wave-reflection model of the vocal tract*, Ph.D. thesis, University of Iowa, 1995.
- [6] B.H. Story, I.R. Titze, and E.A. Hoffman, "Vocal tract area functions from magnetic resonance imaging," *J. Acoust. Soc. Am.*, vol. 100, no. 1, pp. 537–554, 1996.
- [7] J.W. Glenn and N. Kleiner, "Speaker identification based on nasal phonation," *J. Acoust. Soc. Am.*, vol. 43, no. 2, pp. 368–372, 1968.
- [8] J.J. Wolf, "Efficient acoustic parameters for speaker recognition," *J. Acoust. Soc. Am.*, vol. 51, no. 6, pp. 2044–2056, 1972.
- [9] L.S. Su, K.P. Li, and K.S. Fu, "Identification of speakers by use of nasal coarticulation," *J. Acoust. Soc. Am.*, vol. 56, no. 6, pp. 1876–1882, 1974.
- [10] R. Auckenthaler, E.S. Paris, and M.J. Carey, "Improving a GMM speaker verification system by phonetic weighting," in *Proceedings of ICASSP*, 1999, pp. 313–316.
- [11] J.P. Eatock and J.S. Mason, "A quantitative assessment of the relative speaker discriminating properties of phonemes," in *Proceedings of ICASSP*, 1994, pp. 33–36.
- [12] D.P. Delacretaz and J. Hennebert, "Text-prompted speaker verification experiments with phoneme specific MLPs," in *Proceedings of ICASSP*, 1998, pp. 777–780.
- [13] J.L. Le Floch, C. Montacie, and M.J. Caraty, "Investigations on speaker characterization from orphee system techniques," in *Proceedings of ICASSP*, 1994, pp. 49–52.
- [14] Z. Zhang and C.Y. Espy-Wilson, "A vocal-tract model of american english /l/," *J. Acoust. Soc. Am.*, vol. 115, no. 3, pp. 1274–1280, 2004.