Estimation of the Acoustic Properties of the Nasal Tract during the Production of Nasalized Vowels

Xiaochuan Niu, Alexander Kain, Jan P. H. van Santen

Center for Spoken Language Understanding,
OGI School of Science & Engineering at OHSU
20000 NW Walker Road, Beaverton, Oregon 97006, USA
{xiaochua,kain,vansanten}@cslu.ogi.edu

Abstract

Accurate estimation of velar movements is useful for automatic speech recognition, speech enhancement, and diagnosis of certain speech disorders. This paper reports on initial results of a project on estimation of velar movements, for two-microphone setups where the microphones are differentially positioned to pick up nasal and oral speech output. Toward this goal, we propose a method that allows detailed estimation of the acoustic properties of the nasal tract in the simplified condition where the two microphone signals exhibit complete source separation. We successfully test the method against synthetic speech, generated by an articulatory synthesizer in which the acoustic properties of the simulated nasal tract are known.

1. Introduction

During the production of a normal vowel, the velopharyngeal (VP) port is closed, so that the airflow from the glottis passes through the mouth only. When the vowel is adjacent to a nasal, the VP port may be opened due to coarticulation, and the vowel may be perceived as nasalized. There are many speech disorders that involve inappropriate opening of the VP port. For example, deaf speakers may exhibit excessive nasalization due to the lack of auditory feedback. A cleft palate or certain forms of nasal surgery can also result in nasalization problems. Dysarthric speakers may produce hypernasal speech because of motor control impairments.

Estimation of velar movements can be used, for example, for better automatic speech recognition accuracy, for diagnosis of certain motor based speech disorders, and for speech intelligibility enhancement of hypernasal speech by enabling selective restoration of stops that are weakened by inappropriate VP opening.

Nasal tract properties have been a research focus for decades. The morphological measurement of the nasal tract with X-ray [1] or magnetic resonance imaging (MRI) [2] techniques has helped to establish the geometry model of the static nasal tract. Analysis-by-synthesis approaches [3, 4] have attempted to empirically simulate the acoustic and perceptual effects of nasalization. The transfer function of a certain segment of the nasal cavity has been measured either through sweep frequency experiments [5] or by analyzing sound pressure inside and outside the nose [6]. The transfer functions obtained in these experiments revealed intricate spectral patterns of the nasal tract. However, these experiments may involve placing either an oscillator or a microphone inside the nasal cavity. The least invasive method is to analyze the radiated sound pressure directly [7, 8]. This is difficult because the spectrum is highly variable due to the coupling of the pharyngeal-oral tract. Another less invasive method involves the measurement of the sound emitted from the nose and the mouth separately [9, 10]. This method has been used in clinical applications to measure nasalance of speech, defined as the acoustic energy ratio between sound pressures measured in front of the nostrils and the lips. It serves as a diagnostic index of nasality rather than a detailed description of velar movements.

In this paper, we present a new non-invasive method to analyze the spectral properties of the nasal cavity during the production of nasalized vowels. Given the separate measurements of both nasal and oral sound, the transfer ratio function from the volume velocity at the lips to the volume velocity at the nostrils can be estimated. The spectral properties of the nasal cavity are represented by the pole pattern of this transfer ratio function. The proposed method is validated by simulation experiments and we apply it to measure the change of VP opening.

2. Method

2.1. Theoretical formulation

For frequencies below 4 kHz, the acoustic system of the vocal tract can be approximated by a one-dimensional wave equation. The transmission properties of this system can be represented by a circuit network, in which the sound pressure and the volume velocity are analogous to voltage and current [1, 11].

During the production of nasalized vowels, the velum is lowered and a series of air puffs from the glottis propagate both to the mouth and to the nose. An equivalent network representation of a nasalized vowel is drawn in Figure 1, in which the pharyngeal, oral, and nasal cavities are represented by three two-port networks T_p, T_m, and T_n, respectively. Z_{sub}, Z_{ml}, and Z_{nl} denote the acoustic impedance of the subglottal system, the radiation load at the lips, and the radiation load at the nostrils, respectively. P_g, P_v, P_{ml}, and P_{nl} denote the sound pressures at the glottis, the VP port, the lips, and the nostrils, respectively. U_g is the excitation source at the glottis, U_{sub} is the volume velocity out of the trachea, and U_{pi}, U_{po}, U_{mi}, U_{mo}, U_{ni}, and U_{no} denote the volume velocity input and output of the pharyngeal, oral, and nasal cavities, respectively.

According to circuit theory [12], the input-output characteristics of a two-port network can be described by a 2×2 chain matrix (also called ABCD matrix) in the frequency domain. The matrix equations of the three cavities in Figure 1 are

$$\begin{bmatrix} P_g \\ U_{pi} \end{bmatrix} = \begin{bmatrix} A_p & B_p \\ C_p & D_p \end{bmatrix} \begin{bmatrix} P_v \\ U_{po} \end{bmatrix}, \quad (1)$$

This research was conducted with support from NSF Grant No. 0117911 (" Making Dysarthric Speech Intelligible"). We also thank Dr. Tim Bressmann in University of Toronto for inspiring discussions.

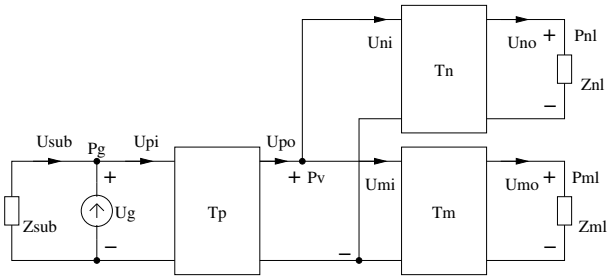


Figure 1: *A circuit network representation of nasalized vowels.*

$$\begin{bmatrix} P_v \\ U_{mi} \end{bmatrix} = \begin{bmatrix} A_m & B_m \\ C_m & D_m \end{bmatrix} \begin{bmatrix} P_{ml} \\ U_{mo} \end{bmatrix}, \quad (2)$$

$$\begin{bmatrix} P_v \\ U_{ni} \end{bmatrix} = \begin{bmatrix} A_n & B_n \\ C_n & D_n \end{bmatrix} \begin{bmatrix} P_{nl} \\ U_{no} \end{bmatrix}. \quad (3)$$

The equations of the oral cavity (2) and the nasal cavity (3) are coupled by $U_{po} = U_{mi} + U_{ni}$. The equations of the pharyngeal cavity and the subglottal tract (1) are coupled by $U_{sub} = U_g + U_{pi}$, which slightly affects the speech spectra when the glottal opening is small and thus the subglottal impedance is very high.

The chain matrix equations (2) and (3), implies that the transfer admittance functions from the air pressure at the VP port P_v to the volume velocity at the lips U_{mo}, and at the nostrils U_{no}, respectively, are given by

$$Y_{m/v}(\omega) =_{def} \frac{U_{mo}}{P_v} = \frac{1}{A_m Z_{ml} + B_m}, \quad (4)$$

$$Y_{n/v}(\omega) =_{def} \frac{U_{no}}{P_v} = \frac{1}{A_n Z_{nl} + B_n}. \quad (5)$$

Thus, the transfer ratio function of the volume velocity from the lips to the nostrils is

$$T_{n/m}(\omega) =_{def} \frac{U_{no}}{U_{mo}} = \frac{Y_{n/v}}{Y_{m/v}} = \frac{A_m Z_{ml} + B_m}{A_n Z_{nl} + B_n}. \quad (6)$$

Another transfer function is the ratio of the air pressure radiated from the nostrils P_{nr} to the air pressure radiated from the lips P_{mr}. This function involves radiation impedances Z_{nr} and Z_{mr}. It can be expressed as

$$T^p_{n/m}(\omega) =_{def} \frac{P_{nr}}{P_{mr}} = \frac{U_{no} Z_{nr}}{U_{mo} Z_{mr}} = T_{n/m}(\omega) \frac{Z_{nr}}{Z_{mr}}. \quad (7)$$

Under certain assumptions, the radiation impedance has a spectral zero at 0 Hz [11]. Thus $T^p_{n/m}$ has the same pole-zero pattern as $T_{n/m}$ except at 0 Hz.

It is of interest to note that the transfer ratio function $T_{n/m}$ is independent on the acoustic system below the VP port, because those effects are canceled by division. the transfer ratio function $T_{n/m}$ has the same poles and zeros as the transfer admittance function of the nasal cavity $Y_{n/v}$. It also has a set of zeros at the pole frequencies of the transfer admittance function of the mouth cavity $Y_{m/v}$. Due to the relatively simple structure of the mouth cavity during the production of vowels, including nasalized vowels, it is reasonable to assume that $Y_{m/v}$ has no zeros. As a result, all poles of $T_{n/m}$ stem from the nasal cavity. Therefore, if we can estimate $T_{n/m}$ from the acoustic signals of a nasalized vowel, we will be able to study the pole pattern of the nasal cavity, except when a pole of $Y_{n/v}$ is canceled by a pole of $Y_{m/v}$. In the next section, we will present a method for estimating the transfer ratio function $T_{n/m}$.

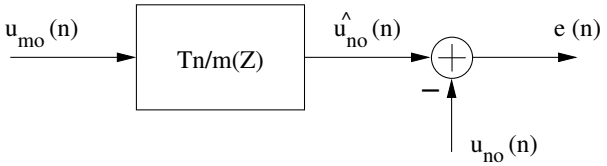


Figure 2: *A linear system for solving the transfer function from* $u_{mo}(n)$ *to* $u_{no}(n)$.

2.2. Analysis method

Suppose we can obtain the volume velocity signals, $u_{no}(t)$ and $u_{mo}(t)$, at the nostrils and the lips during the production of a nasalized vowel. The signals are sampled at a sampling frequency higher than the Nyquist frequency, so that we have the discrete signals, $u_{no}(n)$ and $u_{mo}(n)$. Assuming the transfer ratio function in the Z-domain, $T_{n/m}(Z)$, has the following form,

$$T_{n/m}(Z) = \frac{b_0 + b_1 Z^{-1} + b_2 Z^{-2} + \cdots + b_N Z^{-N}}{1 + a_1 Z^{-1} + a_2 Z^{-2} + \cdots + a_M Z^{-M}}, \quad (8)$$

given the order N and M, and L samples from signal $u_{no}(n)$ and $u_{mo}(n)$, respectively, the parameters in $T_{n/m}(Z)$ can be estimated by minimizing the mean square errors of the system in Figure 2. The solution to this problem is equivalent to the least mean square error solution of the following set of linear equations. For convenience, writing $u_{no}(i)$ as u^i_{no}, and $u_{mo}(i)$ as u^i_{mo},

$$\begin{bmatrix} u^{i+1}_{no} \\ u^{i+2}_{no} \\ \vdots \\ u^{i+L}_{no} \end{bmatrix} = \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_M \\ b_0 \\ \vdots \\ b_N \end{bmatrix}, \quad (9)$$

in which

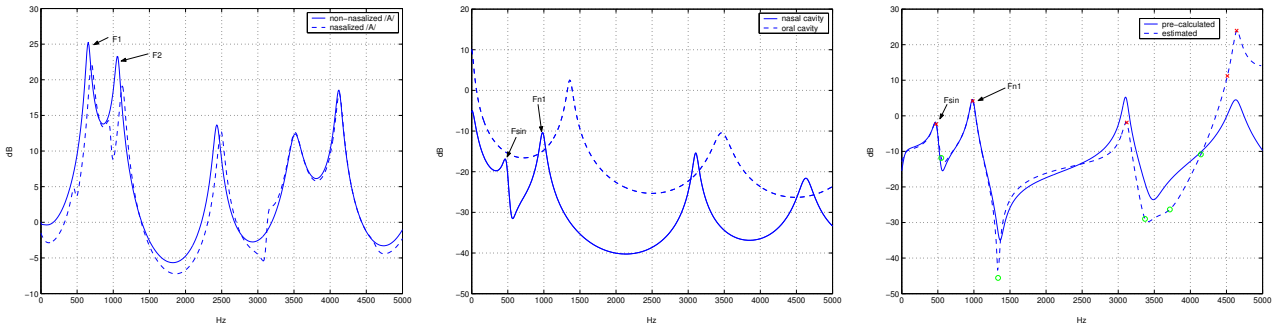
$$A = \begin{bmatrix} -u^i_{no} & -u^{i-1}_{no} & \cdots & -u^{i+1-M}_{no} \\ -u^{i+1}_{no} & -u^i_{no} & \cdots & -u^{i+2-M}_{no} \\ \vdots & \vdots & \vdots & \vdots \\ -u^{i+L-1}_{no} & -u^{i+L-2}_{no} & \cdots & -u^{i+L-M}_{no} \end{bmatrix}, \quad (10)$$

and

$$B = \begin{bmatrix} u^{i+1}_{mo} & u^i_{mo} & \cdots & u^{i+1-N}_{mo} \\ u^{i+2}_{mo} & u^{i+1}_{mo} & \cdots & u^{i+2-N}_{mo} \\ \vdots & \vdots & \vdots & \vdots \\ u^{i+L}_{mo} & u^{i+L-1}_{mo} & \cdots & u^{i+L-N}_{mo} \end{bmatrix}. \quad (11)$$

Poles and zeros can be determined by calculating the roots of the denominator and numerator polynomials. The analysis method remains the same for the transfer ratio function of the pressure $T^p_{n/m}$, after replacing the volume velocity signals with radiated air pressure signals, $p_{nr}(n)$ and $p_{mr}(n)$.

This analysis method is similar to LPC analysis. The difference is that the transfer ratio function here is in ARMA form. The order of the transfer ratio function, M or N, can be chosen in the same way as in LPC analysis. For example, when the sampling frequency is 10 kHz, M and N can be set to 12. L can be either fixed or pitch synchronous.



(a) Transfer functions of U_{mo}/U_g for a non-nasalized vowel and a nasalized vowel.

(b) Transfer admittance functions of the nasal cavity U_{no}/P_v and the oral cavity U_{mo}/P_v during the production of a nasalized vowel.

(c) Comparison of the pre-calculated and the estimated transfer ratio function of U_{no}/U_{mo}.

Figure 3: Transfer functions of synthesized speech.

2.3. Simulation

In order to validate the presented analysis method, an articulatory synthesizer was implemented to simulate the production of vowels or nasalized vowels. The synthesized outputs of the volume velocity from the nostrils and the lips, respectively, were analyzed to estimate the transfer ratio function. Since the articulatory configuration is known before synthesis, we can calculate the transfer ratio function directly from the configuration. By comparing this pre-calculated transfer ratio function with the estimated one, we can evaluate the estimation method.

The structure of the synthesizer is similar to the electrical circuit model in Figure 1, except that it assumes the subglottal impedance to be infinity for simplicity. The LF model [13] is utilized to generate the glottal source signal $u_g(n)$.

As the input of the system, two area functions that represent the uniform-tube approximations of the pharyngeal-oral tract and the nasal tract can be specified. The area function of the pharyngeal-oral tract is based on the X-ray data of vowels from [1] with 0.5 cm intervals of the cross-sectional areas. For the nasal tract, we adopt the empirical model used by [3], where the nasal tract is assumed to be 11 cm long and sampled at 1 cm intervals. In this model, The area of the first three sections from the velum can be changed to simulate the velar position, while the cross-sectional areas of the sections at a distance more than 3 cm from the velum are assumed to be constant. When the area of the first section from the velum is specified, the areas of the next two sections are calculated by a linear interpolation between the first and fourth sections. A sinus cavity is assumed to connect with the nasal tract at a 7 cm distance from the velum. The sinus cavity is modeled as a Helmholtz resonator with a characteristic frequency of 551 Hz. For simplicity, we also restrict the number of the sinus cavities to one. Inside the pharyngeal-oral tract, the VP port is at a 8.5 cm distance from the glottis. The velum can be set open or closed for the simulation of nasalized or non-nasalized vowels.

During synthesis, the area functions are converted into equivalent electrical circuits according to the transmission-line models [14]. Then the transfer functions of U_{no}/U_g and U_{mo}/U_g in frequency domain are calculated and discretized. Using the inverse FFT, these two transfer functions are transformed into impulse responses. As the outputs, the volume velocity signals at the lips $u_{mo}(n)$, and at the nostrils $u_{no}(n)$ are calculated by convolving the glottal source with the corresponding impulse response.

3. Experiments and results

In the following experiments, the settings of the LF parameters for $u_g(n)$ were fixed to typical modal voice values; the pitch and duration were set to 120 Hz and 250 ms with the sampling frequency of 10 kHz. During the analysis step, one pitch period was excised from the synthesized outputs $u_{mo}(n)$ and $u_{no}(n)$, and the transfer ratio is estimated by the method described in Section 2.2.

3.1. Evaluation of synthesized speech

In a first experiment, we synthesized a non-nasalized vowel /A/ as in (father), and a nasalized /A/ by using the same pharyngeal-oral configuration and opening the VP port to a 0.5 cm^2 area. The radiated pressure signals sounded natural to the authors; additionally, we also perceived nasality from the latter signal.

Figure 3(a) shows the log-magnitude spectra of the two transfer functions of U_{mo}/U_g. On the dashed line, a pole-zero pair appears at approximately 1 kHz between formant $F1$ and $F2$, caused by the the most significant resonance of the nasal cavity F_{n1}. Another pole-zero pair is located at approximately 500 Hz, caused by the sinus cavity. The figure also shows frequency increases of the formants and the amplitude decrease of $F1$ when the vowel is nasalized. These spectra confirm that the synthesizer is capable of simulating all the major observed effects of nasalization in real speech [8].

3.2. Transfer functions

In a second experiment, we compared the pre-calculated transfer ratio function U_{no}/U_{mo} with the estimated one. Before calculating U_{no}/U_{mo}, we had to first calculate the transfer admittance functions of the nasal cavity U_{no}/P_v and the oral cavity U_{mo}/P_v. Figure 3(b) shows the log-magnitude spectra of the two transfer admittance functions of the same nasalized /A/ simulated above. U_{no}/P_v (solid line) directly represents the properties of the nasal tract without any influence of other parts of the vocal tract. Each peak on the solid line in Figure 3(b) corresponds to a pole-zero pair of the dashed line

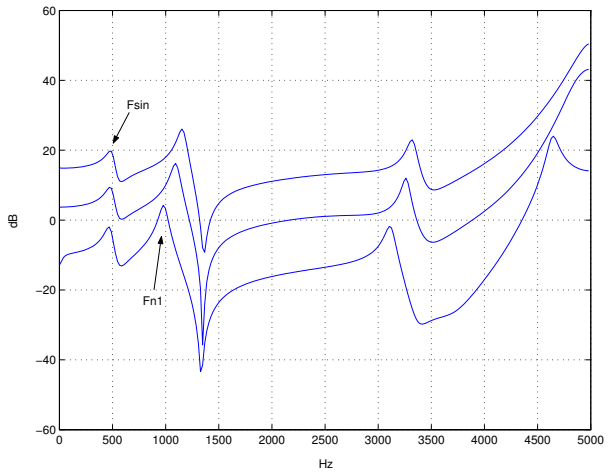


Figure 4: *The estimated spectra of the transfer ratio functions for three nasal configurations. The area of the VP opening increases from the lowest line to the highest line. A shift of 0, 20 and 40 dB has beed added to the lines respectively for better display.*

in Figure 3(a). From the theoretical analysis in Section 2.1, we know that U_{no}/U_{mo} involves the influence of the oral cavity, whose spectral properties are captured by U_{mo}/P_v (dashed line). U_{mo}/P_v has spectral peaks higher than $F2$ of the vowel, and the number of peaks is less than the number of formants of the vowel because of the short length of the mouth cavity.

The pre-calculated and estimated log-magnitude spectra of the transfer ratio function U_{no}/U_{mo} are drawn in Figure 3(c) for comparison. The pre-calculated spectrum (solid line) is the result of the spectral subtraction of U_{mo}/P_v from U_{no}/P_v in Figure 3(b), therefore its spectral peaks are at the same locations as that of U_{no}/P_v. This confirms that one can study the properties of the nasal tract if U_{no}/U_{mo} is available. The dashed line in Figure 3(c) represents the transfer ratio function estimated by applying the analysis method to the synthesized outputs of $u_{mo}(n)$ and $u_{no}(n)$. The estimated spectrum matches the pre-calculated one quite well in the low frequency range and in its peak locations.

3.3. Nasality measurement

In a third experiment, we explored three different configurations of the nasal cavity. During synthesis, the area of the VP opening was set to 0.5 cm^2, 1.5 cm^2, and 2.5 cm^2, while the area function of the pharyngeal-oral tract was configured for /A/. The output signals of each configuration were then analyzed. The estimated spectra of the transfer ratio functions for the three configurations are shown in Figure 4. Note that the pole caused by the sinus (F_{sin}) remains stationary, while the first frequency peak caused by the whole nasal cavity (F_{n1}) increases as the area of the VP opening increases. Thus, the estimated value of F_{n1} can be used to measure the degree of nasal coupling.

4. Conclusions

In this paper, we propose a new non-invasive method to analyze the spectral properties of the nasal cavity by estimating the transfer ratio function from the volume velocity at the lips to the volume velocity at the nostrils. The theoretical derivation shows that this transfer ratio function has the same pole pattern as the nasal cavity, therefore it can be used to analyze the detailed spectral properties of the nasal tract. The simulation experiments validate the proposed analysis method and also imply that this method can be used to measure velar movements.

We note that the real nasal tract may be more complicated than the model used in the simulation experiments. For example, both the asymmetry of the two passages in the nose and the existence of other sinuses can introduce more pole-zero pairs into the transfer admittance function of the nasal cavity. However, one can still use the proposed approach to study these effects in detail, since they are also reflected by the transfer ratio function.

Finally, the signals at the lips and the nostrils may not be perfectly separable. Our analysis method is a first step toward our goal of measuring velar movements and further utilizing this information in intelligibility enhancement.

5. References

[1] G. Fant, *Acoustic theory of speech production*, 2nd ed. Mounton, The Hague, 1960.

[2] J. Dang and K. Honda, "Morphological and acoustical analysis of the nasal and paranasal cavities," *J. Acoust. Soc. Am*, vol. 96, no. 4, pp. 2088–2100, Oct. 1994.

[3] S. Maeda, "The role of the sinus cavities in the production of nasal vowels," in *Proc. ICASSP*, 1982, pp. 911–914.

[4] S. Hawkins and K. N. Stevens, "Acoustic and perceptual correlates of the non-nasal-nasal distinction for vowels," *J. Acoust. Soc. Am*, vol. 77, no. 4, pp. 1560–1575, Apr. 1985.

[5] J. Lindqvist-Gauffin and J. Sundberg, "Acoustic properties of the nasal tract," *Phonetica*, vol. 33, pp. 161–168, 1976.

[6] J. Dang and K. Honda, "Acoustic characteristcs of the human paranasal sinuses derived from transmission characteristic measurement and morphological observation," *J. Acoust. Soc. Am*, vol. 100, no. 5, pp. 3374–3383, Nov. 1996.

[7] D. Cairns, J. H. L. Hansen, and J. Riski, "A noninvasive technique for detection hypernasal speech using a nonlinear operator," *IEEE trans. on Biomedical Engineer*, vol. 43, no. 1, pp. 35–45, Jan. 1996.

[8] M. Y. Chen, "Acoustic correlates of English and French nasalized vowels," *J. Acoust. Soc. Am*, vol. 102, no. 4, pp. 2360–2370, Oct. 1997.

[9] S. G. Fletcher, "Theory and instrumentation for quantitative measurement of nasality," *Cleft Palate J.*, vol. 7, pp. 601–609, 1970.

[10] S. N. Awan, "Development of a low-cost nasalance acquisition system," in *Pathologies of Speech and Language: Contributions of Clinical Phonetics and Linguistics*. International Clinical Linguistics and Phonetics Association, 1996, pp. 211–217.

[11] J. L. Flanagan, *Speech analysis synthesis and perception*, 2nd ed. Springer-Verlag, 1972.

[12] N. Balabanian, T. A. Bickart, and S. Seshu, *Electrical network theory*. John Wiley and Sons, Inc., 1969.

[13] G. Fant, "Glottal flow: models and interaction," *J. Phonetics*, vol. 14, pp. 393–399, 1986.

[14] D. G. Childers, *Speech processing and synthesis toolboxes*. John Wiley and Sons, Inc., 2000.