JAIST Repository
https://dspace.jaist.ac.jp/

Title	A study on transvelar coupling for non-nasalized sounds
Author(s)	Dang, Jianwu; Wei, Jianguo; Honda, Kiyoshi; Nakai, Takayoshi
Citation	The Journal of the Acoustical Society of America, 139(1): 441-454
Issue Date	2016-01
Type	Journal Article
Text version	publisher
URL	http://hdl.handle.net/10119/13706
Rights	Copyright (C) 2016 Acoustical Society of America. Jianwu Dang, Jianguo Wei, and Kiyoshi Honda, and Takayoshi Nakai, The Journal of the Acoustical Society of America, 139(1), 2016, 441-454. http://dx.doi.org/10.1121/1.4939964
Description	

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

A study on transvelar coupling for non-nasalized sounds

Jianwu Dang,[a] Jianguo Wei, and Kiyoshi Honda
Tianjin Key Laboratory of Cognitive Computing and Application, Tianjin University, 92 Weijin Road, Nankai District, Tianjin 300072, China

Takayoshi Nakai
Department of Electrical and Electronic Engineering, Shizuoka University, Hamamatsu, Japan

(Received 5 July 2014; revised 24 December 2015; accepted 4 January 2016; published online 22 January 2016)

Previous studies have found that the velum in speech production may not only serve as a binary switch with on-off states for nasal and non-nasal sounds, but also partially alter the acoustic characteristics of non-nasalized sounds. The present study investigated the unique functions of the velum in the production of non-nasalized sounds by using morphological, mechanical, and acoustical measurements. Magnetic resonance imaging movies obtained from three Japanese speakers were used to measure the behaviors of the velum and dynamic changes in the pseudo-volume of the pharyngeal cavity during utterances of voiced stops and vowels. The measurements revealed no significant enlargements in the supraglottal cavity as subjects uttered voiced stops. It is found that the velum thickness varied across utterances in a way that depended on vowels, but not on consonants. The mechanical and acoustical observations in the study suggested that the velum is actively controlled to augment the voice bars of voiced stops, and nostril-radiated sound is one of the most important sources for voice bars, just as is laryngeal wall vibration. This study also proposed a two-layer diaphragm model that simulates transvelar coupling during the production of non-nasalized speech sounds. The simulation demonstrated that the model accurately represented the basic velar functions involved in speech production. © *2016 Acoustical Society of America.* [http://dx.doi.org/10.1121/1.4939964]

[CHS] Pages: 441–454

I. INTRODUCTION

Nasal and nasalized sounds have indispensable functions in speech, and such sounds are generated by coupling between the nasal and oral cavities via the velopharyngeal port. A number of studies have been conducted on the morphological and acoustical details of the nasal and paranasal cavities (Bjuggren and Fant, 1964; Dang and Honda, 1996a; Dang *et al.*, 1994). Of these, Dang *et al.* (1994) used magnetic resonance imaging (MRI) to investigate the relationship between the morphology and acoustics of the nasal and paranasal cavities (Dang and Honda, 1996a). Sundberg *et al.* (2007) used computerized axial tomography (CAT) scans to investigate the effect of the velopharyngeal opening on vowel formants during singing. Feng and Castelli (1996) investigated the coupling properties of the nasal and oral cavities for nasals and nasalized vowels in French.

The velum plays an important role during transitions between nasal and non-nasal sounds in controlling the velopharyngeal opening, which is a diaphragm of tissue that extends from the posterior end of the hard palate to the pharyngeal wall. According to Dickson and Maue-Dickson (1982), the velum, together with the lateral walls of the upper pharynx, controls the opening of the airway between the oral and nasal cavities by changing its position. The shape of the velum varies considerably with its position and with the state of the musculature. The posterior end of the velum comes into contact with the rear wall of the pharynx in the raised position, thereby effectively closing the air passage between the oral and nasal cavities. When the velum is lowered, an opening is created between the two cavities. Several muscles inserted into the velum raise or lower the diaphragm, and in the process, close or open the velopharyngeal port.

In the literature, the velopharyngeal port is commonly considered to be closed during non-nasal consonant production, while for voiced stops voicing often starts before the release of vocal tract closure. The voicing is resulted from audible glottal airflow oscillations, which is referred to as the voice bar hereafter. Note that the term of "voice bar" is also (but less frequently) used to represent the low frequency component of vowel spectra, as used in Fulop and Disner (2012), which differs in definition from what is used in this study. Although scholars generally agree that some mechanisms must provide for transglottal airflow during the production of voiced stops to maintain vocal fold vibrations, such mechanisms cannot fully explain all the phenomena of voiced stops in speech production. Rothenberg (1968) suggested that the differential pressure necessary for transglottal airflow might be created by any of three mechanisms: the active enlargement of the supraglottal cavity, the passive enlargement of the supraglottal cavity, or the passage of air through incomplete velopharyngeal closure. The first two mechanisms have been advocated by Kent and Moll (1969). Westbury (1983) investigated the temporal variations in supraglottal cavity volume during closures of voiced and

[a]Also at School of Information Science, Japan Advanced Institute of Science and Technology, Asahidai, Nomi, Ishikawa 923-1292, Japan. Electronic mail: jdang@jaist.ac.jp

voiceless stops by measuring the sagittal plane movements of the larynx, soft palate, and portions of the tongue. He used a high-speed cinefluorographic film of utterances produced by an adult male speaker of American English. He found that voiced stops were always accompanied by significant periods of vocal fold vibration during consonant closures and by relatively large increases in the supraglottal volume. In contrast, such a volume increase was not observed for voiceless stops. Svirsky *et al.* (1997) measured tongue displacement and estimated tongue tissue compliance in four subjects during the production of /aba/ and /apa/. All of their subjects demonstrated more tongue displacement during /aba/ than during /apa/, even though the peak intraoral pressure was lower for /aba/. Stevens (2000) summarized several studies on voicing for vocal tract obstruents in his work and pointed out that the maximum increase in the vocal tract volume by active expansion for an adult is probably about 10 cm^3, i.e., about 20% of the total vocal tract volume. This volume expansion is partly produced by lowering the larynx. Velar elevation also contributes to the volume increase. Despite the valuable results presented by these studies, most of them were conducted on English speakers. Therefore, it is not clear whether such observations are language dependent or not.

It is also necessary to consider the velopharyngeal opening as a possible voicing mechanism because nasal airflow has been observed during the production of non-nasal consonants (Subtelny *et al.*, 1969). Lubker (1973) measured the nasal airflow rate and intraoral pressure during the production of selected speech samples, and he pointed out that nasal airflow is common during the production of non-nasal stops by normal English speakers, whereas the observed airflow is most likely not due to the velopharyngeal opening, since the velopharyngeal leaks represent the least probable single mechanism for the generation of the difference in transglottal pressures. Thus, it was speculated that the observed nasal airflow may be due to active movements of the velum, which change the volume of the nasal cavity (Lubker, 1973). This observation indicated that velar movements can induce some air flow in the nasal cavity, where the same amount of air flow should also take place in the oral cavity. This action may be an alternative function to replace the pharyngeal cavity expansion to some extent.

Velar functions are generally more than a binary switch with on-off states during speech production. Rossato *et al.* (2003) measured velar movements in French and found that the velum was involved in the production of non-nasal vowels and consonants as well as in the production of nasal sounds. Suzuki *et al.* (1990) and Dang *et al.* (1992) observed that the amplitude of acoustic radiation from the nostrils changes with utterances of non-nasalized vowels, which was related to the magnitude of intraoral sound pressure variations. Dang *et al.* (1993) measured the intraoral pressure and acoustic signals radiated from the lips, nostrils, and laryngeal wall using a sound-isolated box, whose transmission loss was about 20 dB between the locations of the nostrils and laryngeal wall, and was 31 dB between the lips and two other places. Seven Japanese male subjects participated in the experiment and 270 syllables and words with voiced consonants of /b, d, g, z, j/ were used. They found that the frequency of occurrence of voice bars was 89% for voiced consonants in the initial positions of syllables and words, and 76% of the voice bars had a larger amplitude for nostril-radiated sounds than for laryngeal-wall-radiated sounds, while only 14% of them demonstrated the opposite tendency. The voice bars appeared 100% of voiced consonants in the word-medial positions, 69% of these had larger laryngeal-wall-radiated sounds, and 11% of these had larger nostril-radiated sounds. The results demonstrated that nostril-radiated sounds also play an important role in generating voice bars. They found that the ratio of nostril-radiated sounds to intraoral sound pressure (the AC component) exponentially decreased to $\frac{1}{4}$ of the maximal amplitude as the intraoral air pressure (the DC component) increased from -200 to 400 Pa. Dang and Honda (1996b) used multimodal measurement data to investigate velar functions in vowel production. Their results revealed that the nasal cavity is involved in the production of non-nasalized vowels via velar vibrations. However, this preliminary work was based on insufficient empirical observations.

Biomechanical measurements, on the other hand, have also been conducted to investigate the physical and physiological properties of the velum. Birch and Srodon (2009) measured the biomechanical properties of the velum and the variations across anatomic regions. They found that Young's modulus ranged from 585 Pa at the posterior free edge of the velum to 1409 Pa in the region of attachment. Cheng *et al.* (2011) used magnetic resonance elastography (MRE) to measure the viscoelastic properties of the tongue and velum of seven healthy volunteers during quiet breathing. The results showed that the storage shear modulus, which characterizes the elastic component of oscillation, of the velum was about 2537 Pa. These measurements provide us with references for modeling velar functions.

Several hypotheses have been proposed to explain the phenomena of nasal airflow and transglottal pressure differences during productions of voiced stops and non-nasalized vowels with nostril-radiated sounds, as the discussion above indicates. Thus far, however, no precise explanations have been offered for the mechanisms responsible for these phenomena. The main purpose of this study, then, is to validate these hypotheses by analyzing multimodal data. MRI is used to examine morphological changes in the velum in the production of non-nasalized sounds, and multimodal observations are adopted to investigate the behaviors of the velum to fully understand the mechanism of speech production. The multimodal data include radiated speech sounds, intraoral and intranasal sound pressures, and wall vibrations of speech organs. Acoustic measurements are utilized to examine the contribution of the nasal cavity to acoustic output via transvelar coupling, and wall vibrations are measured to estimate the components of sound radiated from the yielding wall during speech.

The rest of the paper is organized as follows. Section II introduces morphological measurements based on MRI movies. It also presents the results obtained from analysis of the morphological and time-varying characteristics of the velum and pharynx. Section III discusses the multimodal

measurements of radiated speech sounds, internal sound pressures, and wall vibrations. Such measurements are used to examine the transvelar coupling functions during the production of non-nasalized vowels and voiced consonants. Section IV proposes a two-layer diaphragm model of the velum, whose parameters are estimated on the basis of observations and optimized by numerical experiments. The proposed model is evaluated using experimental observations. The final section discusses some new perspectives on transvelar coupling.

II. MORPHOLOGIC MEASUREMENTS BASED ON MRI

Transvelar coupling is considered to be affected by morphological and mechanical changes in the velum during speech. Morphological changes in the state of the velum are measured based on MRI observations in this section, while mechanical and acoustical characteristics are investigated by using wall vibrations and acoustic measurements in Sec. III A.

A. Experimental setup

A cine MRI experiment was conducted using the synchronized sampling method (Masaki *et al.,* 1999) to investigate velar behaviors during short utterances of non-nasalized sounds. The MRI system used in this study was a Shimazu-Marconi Eclipse 1.5T Power-Drive 250 scanner. Three adult male Japanese subjects (S1, S2, and S3) participated in the experiment. The speech materials were meaningless Japanese words consisting of vowels and voiced stop consonants, including /baba/, /bibi/, /bubu/, /bebe/, /bobo/, /gaga/, and /gigi/. The frame rate of the reconstructed MRI movie files was 64 frames per second (fps) for the midsagittal and parasagittal layers.

Figure 1 shows two frames from the MRI motion images for /i/ in /bibi/ (left) and /o/ in /bobo/ (right) obtained from subject 1. One can see from this figure that the orientation and thickness of the velum differ to a certain extent. Six parameters of velar configurations were measured to investigate the velar mechanism in speech: the position and thickness of the velum, the anteroposterior dimensions and width of the velar diaphragm, and the anterior-posterior distance and length of the pharynx. These parameters were used as the basis for estimating the effective area of the velum and for modeling velar functions. The definitions of the parameters are given in Fig. 2.

The height of the velum during the production of non-nasal sounds is defined as the angle between the upper surface of the velum and the line tangent to the hard palate at its end point, as shown in Fig. 2(a). It has been known that the principal component analysis (PCA) parameters defined by Serrurier and Badin (2008) were used successfully to describe larger-scale velar movements during the production of both nasal and non-nasal sounds, while the velar angle may be sufficient for measuring the smaller scale movement, the height position of the velum. The velar angle is partly related to a speaker's efforts in controlling the velum because velar elevation is driven by muscle activation, but intraoral air pressure also contributes to elevating the velum. The anterior-posterior (A-P) distance of the pharynx is defined as the distance from the pharyngeal surface of the tongue on the midsagittal plane to the posterior pharyngeal wall at the level between the second and third cervical vertebrae. The pharyngeal length is defined as the vertical distance from the outlet of the larynx to the apex of the uvula, both of which can be easily identified.

Figure 2(b) illustrates how the velar thickness was calculated. The velar thickness determines the mass per unit area of the velum, which is a factor affecting velar vibration. The h_1 in the figure represents the distance between the lower and upper surfaces of the velum near the junction of the hard and soft palates. The h_2 is the height from the zero plane to the highest portion of the upper surface, where the zero plane is defined as the extended line of the lower surface of the anterior segment of the velum. The velar thickness is defined as the average of h_1 and h_2. Figure 2(c) shows a top view of the velum, which is based on Dickson and Maue-Dickson (1982) and Stevens (2000). The area with oblique lines denotes the diaphragm of soft tissue, and the dashed lines illustrate the velopharyngeal port with different degrees of opening determined by the descent of the velum. The dark gray region is used to approximate an elliptically shaped area of effective velar vibration. The long axis, W, on the ellipses denotes the width of the diaphragm, and the short axis, d, is its anteroposterior dimension, i.e.,

FIG. 1. Two movie frames for vowel (left) /i/ in utterance /bibi/ and (right) /o/ in utterance /bobo/. Images were obtained from subject 1.

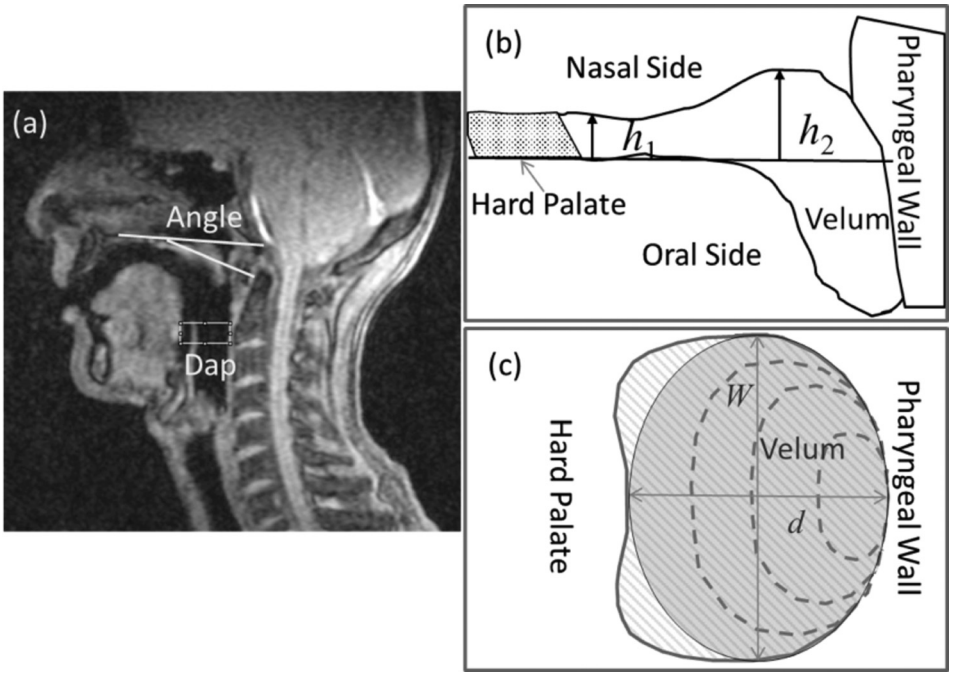

FIG. 2. Definitions of the morphological parameters in MRI measurements. (a) Angle and thickness of the velum and pharyngeal anterior-posterior distance (D_{ap}) of the vocal tract. (b) Thickness of the velum. (c) Top view of the velum, where W and d denote width and anterior-posterior dimension of velar diaphragm, respectively. MRI image was obtained when subject 1 was producing first /e/ in /bebe/.

the distance from the junction of the hard and soft palates to the rear wall of the upper pharynx.

B. Time-varying properties in non-nasalized utterances

To clarify the mechanism involved in the production of voiced stops, we measured variations in the velar angle, and the A-P distance (D_{ap}) and length of the pharyngeal cavity. Figure 3 plots the measurements for the three subjects during utterances of /baba/ (the upper panel) and /bibi/ (the lower panel). The shadowed portions in the figure represent the periods that the lips were closed in producing /b/, and the dark vertical lines indicate the onset of voicing, at which the frame number of voice onset was set to zero. The angular degree is four times the ordinate. The angular changes indicate that the velum rapidly rises to close the velopharyngeal port as the speakers produce /b/, and it then descends for the rest of the utterance. In most cases, the velar angle slightly increases when approaching the second voiced stop or during it.

The changes in the A-P distance of the pharynx indicate that at the beginning of the utterance, the tongue moves backward for /a/ and consequently narrows the pharyngeal cavity, while it moves forward for /i/ and enlarges the pharyngeal cavity. The A-P distance demonstrates no obvious changes in the second voiced stop for the utterances with either /a/ or /i/. These facts indicate that the tongue moves toward the target for vowels, but is not affected by the stops. Consequently, no significant increments were found in the anteroposterior dimension of the middle pharynx, even though such expansions of the pharynx would aid the production of voiced stops.

In this study, we are more interested in relative changes in the pharyngeal volume in the transitions between vowels and voiced stops. Although the left-right dimension of the pharynx possibly varies across vowels, the width can be expected to be invariant within the same vowel context such as that in /baba/. Since we did not compare the volumes between different vowel contexts in this study, a pharyngeal pseudo-volume could be represented by the product of the pharyngeal length and A-P distance with a unit width. The pseudo-volume (V_{ph}) is plotted in Fig. 3 with the bold lines, which is used for investigating relative changes in the pharyngeal volume. The V_{ph} value was estimated to be 1.5 times

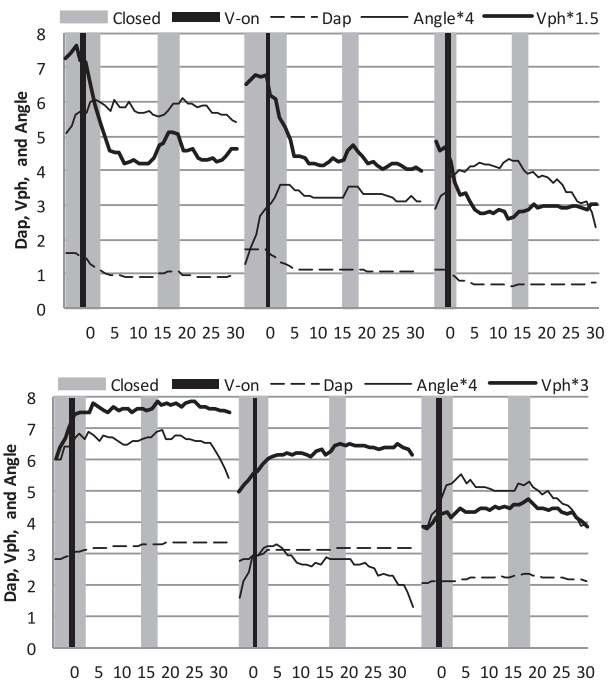

FIG. 3. Variations in A-P distance (D_{ap}), velar angle, and pharyngeal pseudo-volume (V_{ph}) during utterances of /baba/ (upper) and /bibi/ (lower) for the three subjects. Horizontal axis is in frame numbers (1 frame = 9.4 ms), and 0 is the voice onset indicated by vertical line (V-on). Gray vertical stripes indicate when the lips are closed. Angle*4 denotes that the real value of velar angle is 4.0 times ordinates in degree. V_{ph} *1.5 and V_{ph} *3 denote that their real values are 1.5 times and 3.0 times ordinates in cm^3.

the ordinate for /baba/ and three times the ordinate for /bibi/. The V_{ph} for utterance-initial voiced stops rapidly decreased during utterance /baba/ while it definitely increased for /bibi/. These changes basically reflect tongue movements toward the vocalic targets from a neutral position. The V_{ph} increased by 18% for the utterance of the medial /b/ in /baba/ with S1 and by 12% for that with S2. These increments were close to the estimates by Stevens (2000). The V_{ph} demonstrated no significant increments for the other utterances during the medial voiced stops. This is inconsistent with the observation (Westbury, 1983) of American English speakers, suggesting that other mechanisms may be involved in the production of voiced stops.

Stevens (2000) also estimated that the volume expansion of the pharynx during voiced stops was partly produced by lowering the larynx. To examine this possibility in Japanese utterances, two vertical distances were used to identify the relative position of the larynx. The vertical distances were defined from one superior fixed point, i.e., the junction of the hard and soft palates, to two laryngeal landmarks, i.e., the laryngeal outlet and the vocal folds. The distances were measured frame by frame in the vowel-consonant-vowel periods around the medial consonant for these three subjects. The two distance measures were only used for the sake of accuracy, since they must vary in parallel in voiced utterances. As a result, no significant changes in the distances were found in the periods within any utterances. This implies that no laryngeal descent takes place in producing the voiced stops. This is inconsistent with the estimate by Stevens (2000).

C. Articulation-related properties in non-nasalized speech

The velar thickness for each vowel is calculated with the average value over all the vocalic frames, and is the average over the initial and medial consonant frames for the stops. Figure 4 plots the results for the three subjects. The velar thickness monotonically increases in the order from /i/, /u/, /e/, /o/, to /a/, with an increase of greater than 13% from /i/ to /a/. The thickness of the velum for the consonants increases in the same order and by about the same amount as that for the succeeding vowels. The same tendency was

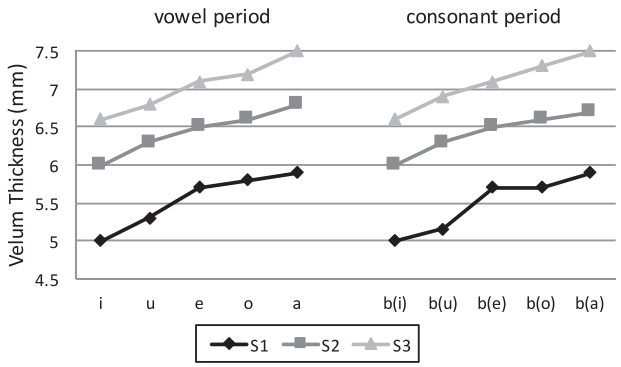

FIG. 4. Velar thickness averaged over all the vocalic frames for each vowel (left), and averaged over consonant frames of the initial /b/ and medial /b/ in utterances /bVbV/ (right), where V indicates one of five Japanese vowels.

observed in the production of voiced stop /g/. This implies that variations in thickness depend on vowels but not on consonants. The velum is thinner during utterances of close vowels and thicker during utterances of open vowels. The velar thickness during medial consonants increases by up to 3% more than that of the initial thickness in most cases. Similarly, the velar thickness also increases somewhat in the utterances of second vowels.

As suggested in previous studies (Dang *et al.*, 1993; Suzuki *et al.*, 1990), velar vibration is a possible cause of oral-nasal acoustic coupling for non-nasalized vowels. One of the most important parameters for velar vibration is the mass of the velum, which is determined by the effective area of the diaphragm and the thickness of the velum. The effective area is approximated by the ellipse shown in Fig. 2(c). The short axis, d, of the ellipse was measured on the basis of the MRI data obtained from the three subjects. Because the long axis, W, of the ellipse was unavailable in the current dataset, it was approximated as 3.6 cm for these three subjects, which was an average of different subjects in Dang and Honda (2001). The area of the ellipse was then calculated as

$$S = \pi/4Wd, \quad (1)$$

where W and d denote the long and short axes of the ellipse. The results are listed in Table I. On the basis of the morphological data, the mass per unit area of the velum ranges from 0.5 to 0.7 g/cm^2, where the density (ρ) of the velum is 1.06 g/cm^3 (Birch and Srodon, 2009).

III. ACOUSTIC MEASUREMENTS BASED ON MULTIMODAL APPROACH

A multimodal approach is used to measure the radiated speech sounds, internal sound pressures, and wall vibrations during the production of non-nasalized sounds, such as vowels and voiced consonants. The multimodal data are used to investigate the transvelar coupling functions of the velum in speech production.

A. Experimental setup

The experiment is aimed at measuring the acoustic coupling between the nasal and oral cavities during productions of non-nasalized vowels and voiced stops. The experimental setup is outlined in Fig. 5, where three microphones were used for recording sound pressure signals, and three accelerometers were used for detecting wall vibrations. In the acoustic measurements, a B&K 4003 microphone (M1) was placed in front of subjects (15 cm from the lips) to record

TABLE I. Effective area (S) of the velum for transvelar function, where d and W denote anterior-posterior dimension and width of velar diaphragm, respectively.

	d (cm)	W (cm)	S (cm^2)
S1	3.1	3.6	8.77
S2	2.9	3.6	8.20
S3	3.3	3.6	9.33

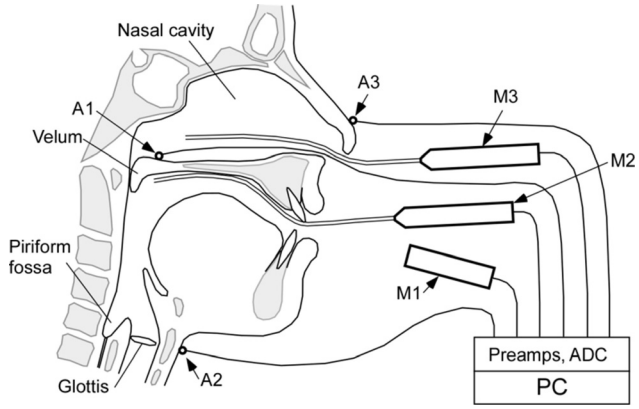

FIG. 5. Experimental setup for measuring internal and external sound pressures of vocal tract and acceleration of the velum and vocal tract wall, where M1–M3 are microphones and A1–A3 are accelerometers.

radiated sound pressures. Two B&K 4182 probe microphones (M2 and M3) were used to measure intraoral and intranasal sound pressures; the measurements was carried out via two identical 30-cm long flexible tubes. The probe tubes had a 0.165-cm outer diameter, a 0.076-cm inner diameter, and matched impedance to the microphones. The probe tube of M2 was inserted into the oral cavity and glued onto the hard palate, where the tip of the tube was placed beneath the velum. The probe tube of M3 was inserted along the nasal floor through one nostril into the nasopharynx about 7.5 cm from the nostrils. The probe microphone signals were continuously monitored throughout the experiments to prevent interference from clogging by mucus so that potential clogging could be quickly removed by injecting air into the probe tubes. The setup for the probe tubes was maintained during injecting operation, except when detaching and resetting them for the operation. High-pass filters were not used in the measurements to observe the DC component of the signals, but a high-pass filter with a cutoff frequency of 60 Hz was applied to obtain the AC component. Low pass filters were used in the measurements to meet the Nyquist frequency of 8 kHz.

The accelerometers used in this study were Endevco Model-22 devices weighing 0.14 g with a stiff coaxial cable. An accelerometer (A1) was used to measure velar vibrations, which was placed approximately on the center of the velum via one nostril, and it appressed the velar surface by utilizing the cable stiffness. The position of A1 was adjusted in the anterior-posterior direction to find a location with maximum vibration. Another accelerometer (A2) was placed on the surface of the neck to measure the vibration of the laryngeal wall. The third accelerometer (A3) was attached onto the surface of the nasal ala to estimate sound radiation from the nostrils. The centimeter-gram-second (CGS) system of units was adopted to compare the different channels; the unit for acoustic signals is Pascal ($Pa = 10\,dyn/cm^2$), and that for the vibration signals is cm/s^2.

Three male Japanese subjects participated in the experiment. The speech materials were meaningless Japanese words consisting of the vowels /a/, /i/, /u/, /e/, and /o/, and the consonants /b/, /d/, /g/, and /m/. Analysis of nasal coupling during vowels focused on vowel segments, whereas analysis of voice stops focused on voiced bars.

B. Oronasal coupling during vowel production

The measured multimodal data were analyzed for spectral and articulation-related variations. Figure 6 shows the spectra for the intraoral sound pressure (OP) recorded by M2

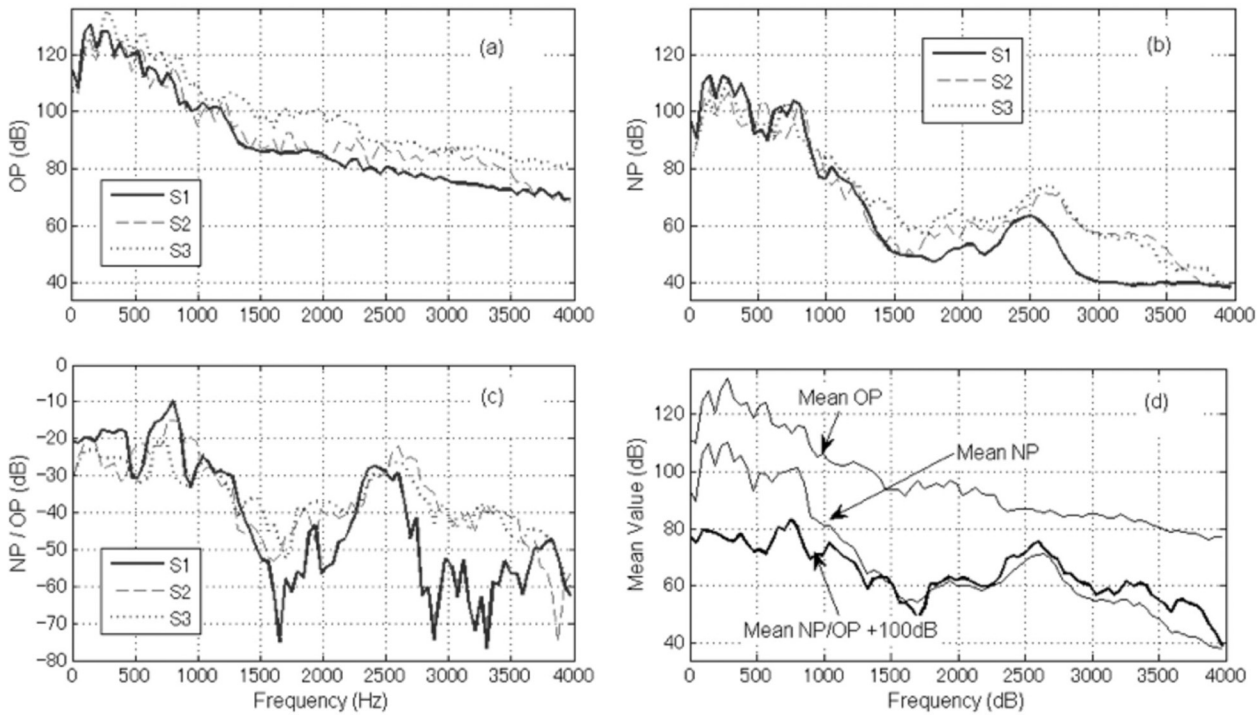

FIG. 6. Spectra averaged over five Japanese vowels for subjects S1, S2, and S3: (a) intraoral sound pressure (OP), (b) intranasal sound pressure (NP), (c) ratio of NP to OP, and (d) spectra averaged over all the vowel utterances and subjects.

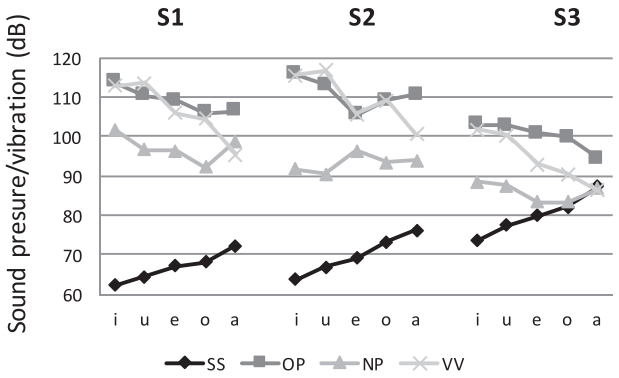

FIG. 7. Speech sound (SS), intraoral sound pressure (OP), intranasal sound pressure (NP), and velar vibration (VV) in vowel periods for three subjects S1, S2, and S3.

and the intranasal sound pressures (NP) recorded by M3, together with an index that indicates the relationship between them, during utterances of non-nasalized vowels by the three subjects. The spectra of the OP in Fig. 6(a) were averaged over the five Japanese non-nasalized vowels uttered by the three subjects. The frequency range in the figure was set to 4 kHz because there was less information of interests for this study above 4 kHz. The intraoral spectra monotonically decreased at about −12 dB/oct, in which the ripples correspond to the fundamental frequency and its harmonics. The average intraoral spectra reflected the properties of the glottal source rather than those of the vocal tract. The averaged intranasal spectra of NP in Fig. 6(b) have a relatively flat region below 0.8 kHz, and the amplitude rapidly decreases at around 1 kHz. Resonance peaks appear near 2.5 kHz; these peaks are the language-independent properties of the nasal cavity (Hyon *et al.*, 2014). The ratio of NP to OP is calculated by

$$H_{o2n} = \frac{NP(\omega)OP^{*}(\omega)}{\|OP(\omega)\|^{2}}, \tag{2}$$

where $NP(\omega)$ is the spectrum of intranasal sound, $OP(\omega)$ denotes the intraoral sound spectrum, and $OP^{*}(\omega)$ denotes its conjugate. Figure 6(c) plots the NP/OP ratio, which roughly demonstrates the behaviors of the velum. The spectral attenuations through the velum for the three subjects are approximately −20 dB below 1 kHz, and slightly greater than −30 dB at about 2.5 kHz. The attenuations are even greater in other frequency regions. Figure 6(d) illustrates the spectra that were averaged over all the vowel utterances and the subjects.

Figure 7 plots the amplitudes of speech sound (SS) recorded by M1 and the velar vibration (VV) recorded by A1, as well as the OP and NP for the three subjects. The amplitude of SS monotonically increases in the order from /i/, /u/, /e/, /o/, to /a/ for all the subjects, whereas the OP and VV roughly decrease in the same order. Variations in the NP demonstrate no obvious tendencies. The sound pressure amplitude is almost the same for S1 and S2, whereas S3 has an inverse tendency, i.e., a higher output (larger SS) with a lower input (smaller OP) compared with the other subjects. These results imply that S3 had less oronasal coupling during vowel production.

C. Velar functions in oronasal coupling

Transvelar coupling is regarded as an acoustic effect of velar vibration that takes place between the oral and nasal cavities. A closed velum can be roughly approximated as a circular drumhead. Asmar (2005) found that a drumhead has infinite vibration modes, and the number and the location of the nodes and antinodes depend on the shape of the drumhead and the vibration modes. Therefore, it is difficult to measure all of the vibration modes with a single accelerometer. The underlying principle of our setup was to record the most important components with the maximum vibration as shown in Sec. III A.

The vibration of the velum was recorded using A1 and analyzed separately for the periods of vowels and consonants. Figure 8 shows the spectra of velar vibrations during vowels and voice bars, which were averaged over the three subjects. The velar vibrations have a strong peak at approximately 160 Hz for all the vowels. However, the peak amplitudes differ from vowel to vowel and the difference between vowels /i/ and /a/ is greater than 20 dB. In contrast, the peaks during the voice bars are identical in terms of frequency and amplitude for all the cases, regardless of the succeeding vowels. The resonant frequency of the velum is close to that of the actual measurements of the wall vibrations for bilabial stop configurations (Fant *et al.*, 1976; Fujimura and Lindqvist, 1971), in which the resonant frequency is approximately 180 Hz for an adult male speaker and approximately 190 Hz for a female speaker (Stevens, 2000).

The spectra of the averaged OP have also been plotted in Figs. 8(a) and 8(b) as a reference with the thin dark lines for vowels and voice bars. Compared with the velar vibrations during voice bars, the averaged OP spectrum of the vowels demonstrates the richer frequency components; the velar vibrations also reflect the difference in OPs between the vowel and voice bar periods. During the production of voice bars, the averaged OP spectrum and velar vibrations have identical resonance peaks, which are much sharper than those found in the cases of vowels.

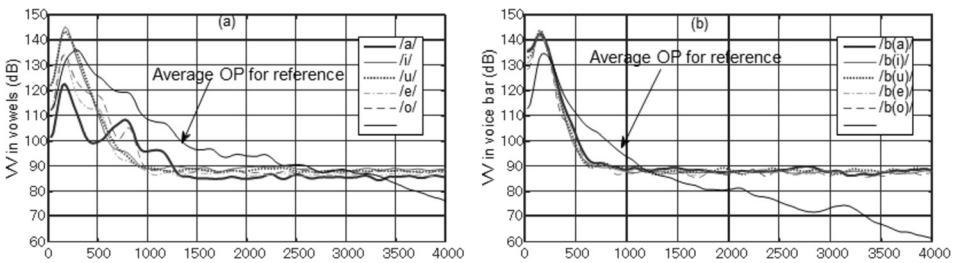

FIG. 8. Spectra of velar vibration during production of (a) vowels and (b) voiced stops averaged over the data from three subjects. Dark lines plot average OP as reference.

The amplitudes of the SS, OP, NP, and VV for the three subjects during the production of the voice bars are plotted in Fig. 9. One can see that the amplitudes of the SS and OP during voice bars are independent of the succeeding vowels. Both SS and OP are about 10 dB smaller than those of the close vowels produced by S1 and S2 (see Fig. 7 for details). The velum maintains its vibration (VV) almost constantly during stop periods, regardless of the succeeding vowels, and the amplitude of the VV is about 10 dB larger than that of the OP for S1 and S2. The VV during vowel productions possibly depends on both the intraoral sound pressure and vowel-dependent thickness of the velum. The VV is significantly higher during stop periods for all the subjects. Interestingly, the VV exhibits about the same amplitude during the production of the voice bars and close vowels /i/ and /u/, whereas the OP is considerably lower during stop periods. One of possible reasons is that speakers may actively control the condition of the velum as they produce the voice bar, while the other possible reason is that the location of the end of M2 may be near a node in the standing wave for /i/ and /u/, and near an antinode for voiced stops.

As can be seen from Fig. 9, S3 has the strongest voice bar with a smaller degree of velar vibration than S1 and S2. It is found that S3 had difficulty in producing a voiced stop when his nostrils were blocked. We speculated that for S3 the velopharyngeal port adjusted its opening to permit the velopharyngeal airflow that maintained vocal fold vibrations (Stevens, 2000). That is, S3 produces a nasalized voice bar. Even in this situation, the velum was more actively controlled than that in instances where the velum vibrated during vowel production.

D. Production of voice bar during voiced stops

A voice bar is generated when the oral tract is closed and the velopharyngeal port is supposed to be closed; thus, a question arises as to where the sound radiates externally. To address this issue, the sound source for the voice bars produced by the subjects was estimated by analyzing multimodal data. Since the relation of acceleration (a) to pressure (P) can be simply represented by $P = ma/S$, if assuming that related mass m and area S of the laryngeal wall are invariant when it vibrates, the surface pressure on the laryngeal wall is proportional to acceleration. Therefore, we can directly compare the sound pressure with the acceleration signal of the laryngeal wall, instead of the pressure on the wall surface. Figure 10 plots the relationship between speech sound (SS), intranasal sound pressure (NP), and laryngeal wall vibration (LV) during the production of the voice bar in the pronunciations of voiced stops /b/, /d/, and /g/ by S1 and S2, where the unit was 1/20 000 Pa for acoustic signals, and 1/2000 cm/s^2 for vibration. The SS and NP exhibit an almost linear relationship shown in the left panel of Fig. 10, whereas SS and LV exhibit less evident linearity in the right panel. The correlation coefficient is 0.86 between SS and NP, and is 0.37 between SS and LV. This indicates that the sound radiated during stop period primarily depends on nasal output. Thus, a conclusion can obviously be drawn that nostril-radiated sound is an important source for voice bars.

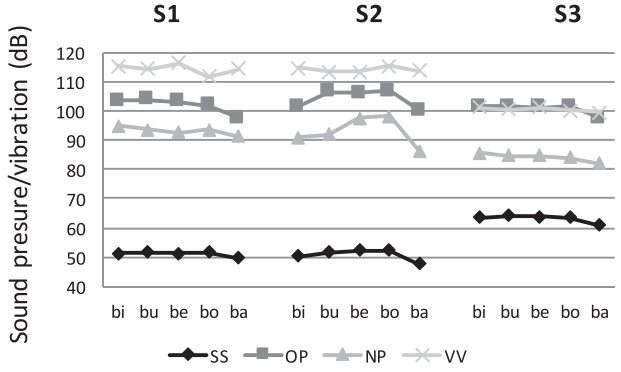

FIG. 9. Speech sound (SS), intraoral sound pressure (OP), intranasal sound pressure (NP), and velar vibration (VV) during stop periods by S1, S2, and S3.

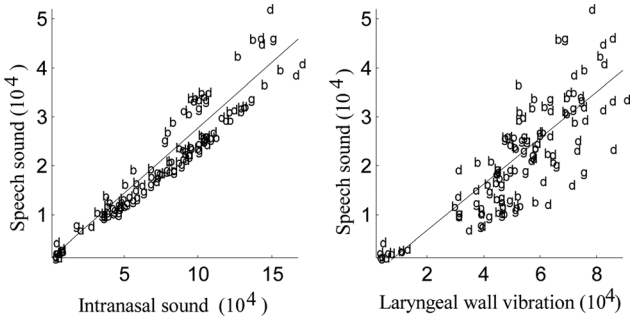

FIG. 10. Relationships between speech sound (SS) and intranasal sound (NP) (left), and between SS and laryngeal wall vibrations (LV) (right). All the data were obtained during stop periods of /b, d, g/ by subjects 1 and 2.

E. Oronasal coupling and velopharyngeal states

Velar movement is thought to switch the states of the velopharyngeal port between open or closed states in speech production. Another question that emerges is whether oronasal coupling continuously varies with state switching. To answer the question, we analyzed the measured sounds in both states during nasal sounds, nasalized sounds, and non-nasalized sounds.

Figure 11 plots the ratios of intranasal sound pressure (NP) to intraoral sound pressure (OP) in the frequency

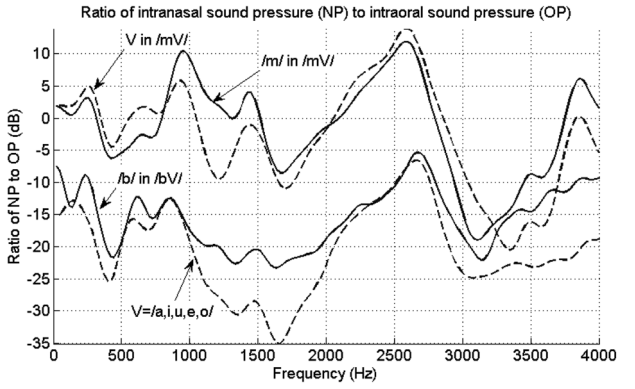

FIG. 11. Ratios of intranasal sound pressure to intraoral sound pressure for nasal consonant /m/, nasalized vowels, voiced stop consonant /b/, and non-nasalized vowels. Curves represent average frequencies obtained from subject 2.

domain for subject 2. The spectral curves for the NP/OP ratios correspond to the averages in the consonant period for /m/, in the voice bars for /b/, and in the vowel periods for the non-nasal and nasalized vowels. As expected, the nasal consonant /m/ has the largest value of the ratios. The other ratio values in descending order are: nasalized vowels, the voice bar of /b/, and non-nasalized vowels. The major peaks and troughs in the graph are consistent in the frequency region below 3 kHz for all the cases. The peaks correspond to the nasal resonance frequencies. Of these, the amplitude of the peak at around 1.5 kHz varies largely; its amplitude continuously increases as production changes from non-nasal sounds, voiced stops, and nasalized sounds to nasal consonants. These results demonstrate that oronasal coupling continuously varies, although the effect of the open-closed contrast in the velopharyngeal port is somewhat discontinuous. The findings confirm that the velum does not function as a binary switch with on-off states alone in speech production, and they also highlight the possibility of using a uniform model to describe velopharyngeal functions in the production of both nasal and non-nasal speech sounds.

IV. MODELING OF VELAR FUNCTIONS IN NON-NASAL SOUND PRODUCTION

On the basis of the previous analyses, velar vibration can be regarded as one of the factors that cause velopharyngeal acoustic coupling, and has an effect during the production of non-nasalized sounds. By simplifying the velum morphology, transvelar coupling is mediated by tissue vibration as the drumhead of the soft tissue of the velum. Acoustically, the velum can be treated as a section of the vocal tract, where three physical variables are involved: the input sound pressure to the section, the output sound pressure from the section, and sound propagation (via vibration) bifurcating at the section through either the open or closed velopharyngeal port. Across the velum, the intraoral sound pressure (OP) is considered to be input, and the intranasal sound pressure (NP) is output, which is also susceptible to the input impedance of the nasal cavity as observed from the velum.

A. Mechanical model of the velum

In producing non-nasal sounds such as vowels and voiced stops, the velum closes the velopharyngeal port and its tissues receive intraoral sound pressure variations. The velum can be regarded as a drumhead with a quasi-circular diaphragm and is driven by the intraoral sound pressure to vibrate. Because the velum is of a certain thickness, its upper and lower surfaces may exhibit different phases of vibration. Accordingly, we developed a simplified mechanical model with a two-layer diaphragm to represent the properties of the velum. There is a schematic of the model in Fig. 12. The two layers are connected to the surrounding wall with identical spring-dashpot pairs of R_m and K_m. The two layers are linked together by a spring-dashpot pair of R_{m0} and K_{m0}, and the layer is assumed to be a non-yielding circular plate. The model is driven by intraoral sound pressure.

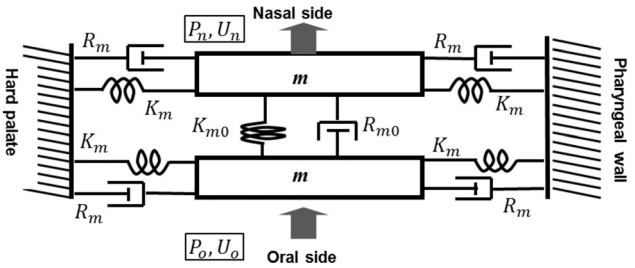

FIG. 12. Mechanical model of the velum for production of non-nasal sounds.

On the basis of the MRI observation, the average thickness of the velum is 0.634 cm and the standard deviation is 0.069. The average area of the velum is 8.77 cm^2 when the velopharyngeal port is closed. The density (ρ) of the velum is 1.06 g/cm^3 (Birch and Srodon, 2009) and its total mass is 5.894 g; thus, $m \approx 2.95$ g for each layer in the model. As obtained by the biomechanical measurements (Birch and Srodon, 2009; Cheng *et al.*, 2011), the storage shear modulus (E') of the velum is 25 370 dyn/cm^2, and the corresponding loss modulus (E'') is 9070 dyn/cm^2. The stiffness is calculated as

$$K_m = \frac{E' A_c}{L}, \tag{3}$$

where A_c is the cross-sectional area of the velum on the coronal plane, and L represents the length of the velum. Consequently, $K_m = 23\,769$ dyn/cm^2. The R_m can be derived from the loss modulus and relaxation time (τ). A relaxation time (τ) of 0.01145 s was adopted in this study, which followed the procedure by Kim *et al.* (1999). Mechanical impedance was estimated on the basis of the experimental data summarized in Table II. Acoustic impedance, which is also listed in Table II, was obtained using the relationship in the table. The K_{m0} and R_{m0} in the model were introduced to simulate the phase difference between the two layers. These two parameters were estimated and the other parameters were optimized by using the numerical experiments discussed in Sec. IV B.

B. Acoustical model of the velum

Figure 13 shows an equivalent electrical model of oronasal coupling in the pharyngeal segment, in which the upper branch with variable resistance and inductance is used to simulate transmission functions when the velopharyngeal port is open, and the lower branch implements transmission functions when the velopharyngeal port is closed. The lower

TABLE II. Mechanical and acoustic parameters of the velum estimated on the basis of experiments.

Mechanical parameters	m (g)	K_m (dyn/cm)	R_m (dyn × s/cm)
Values	2.95	2.38×10^4	61.96
Acoustical parameters	$L'_a = m/A$ ($dyn \times s^2/cm^3$)	$C'_a = A/K_m$ (cm^3/dyn)	$R'_a = R_m/A$ ($dyn \times s/cm^3$)
Values	3.43×10^{-2}	3.69×10^{-4}	7.07

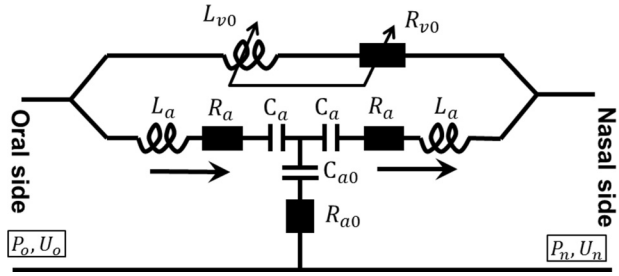

FIG. 13. Equivalent electric model of the velopharyngeal port for speech production. The upper branch with variable resistance and inductance was used to simulate acoustic functions when the velopharyngeal port was open, and the lower branch indicates transvelar function when the velopharyngeal port was closed.

branch is the acoustical model of the velum to simulate sound pressure propagation via velar vibration, which is the counterpart of the mechanical model of the velum in Fig. 12. The upper branch consist of two elements, i.e., variable resistance (R_{vo}) and variable inductance (L_{vo}), to characterize the acoustic properties of the velopharyngeal port. The values of R_{vo} and L_{vo} change according to the degree of velopharyngeal port opening. When the velopharyngeal port is completely closed, R_{vo} approaches infinity, and L_{vo} equals zero.

The parameters were estimated and optimized using the method of analysis-by-synthesis (AbS) based on the transmission line model of the vocal tract, where the velopharyngeal portion is represented by the equivalent electric model of the velum shown in Fig. 13. The area functions of close vowel /i/, open vowel /a/, and stop /b/ in /ba/ were used in the estimation (Dang and Honda, 1996b). The smallest lip radiation and the largest nasal radiation were generated in the production of close vowel /i/ in the AbS process, whereas the opposite was obtained in the production of open vowel /a/. The mechanical based parameters of R'_a, L'_a, and C'_a in Table II were used as the initial values in the AbS method. The parameters of L_a, C_a, R_a, C_{a0}, and R_{a0} were adjusted to match the simulation data to the multimodal data in the frequency domain. The procedure was implemented as follows. Acoustic inductance L_a was set to be identical to L'_a, which is the most reliable value we obtained from the morphological measurements. The R_a was estimated for the frequency region below 1 kHz, with reference to the difference between the oral and nasal sound pressures for /i/; C_{a0} and R_{a0} were estimated by reproducing the peak around 2.5 kHz

TABLE III. Acoustical parameters obtained from numerical experiment (units are the same as those in Table II for the acoustical parameters).

L_a	C_a	R_a	C_{a0}	R_{a0}
3.43×10^{-2}	4.15×10^{-4}	8.96	4.78×10^{-5}	0.90

(see Fig. 6). Finally, all the parameters were refined using the measured multimodal data.

The optimized parameters are listed in Table III, where the units are the same as those used in Table II. The optimized compliance C_a is about 10% larger than the mechanical based compliance C'_a, and the optimized acoustic resistance R_a is about 20% larger than the mechanical based resistance R'_a. Resistance R_{a0} between the two layers is about $\frac{1}{10}$ of R_a. Compliance C_{a0} is 15% larger than C_a. These differences imply that the horizontal stiffness is slightly larger than the vertical stiffness, and that the viscosities differ between the two directions. If taking into account the effects of velar thickness, the amplitude variations in the nostril-radiated sounds were within 4% across all the vowels. Therefore, the average velar thickness was used in the following simulation for simplifying the model.

C. Simulation of non-nasal sounds

The schematic of the structure of the transmission line model used in the simulation is shown in Fig. 14. The piriform fossa and paranasal cavities were included in the simulation but they have not been shown in the figure for simplicity. The velum model with the parameters in Table III was embedded in the transmission line model by using the dark lines in Fig. 14, and the other parameters were calculated based on the area functions of the vocal and nasal tracts (Dang and Honda, 1996b). A two-mass model of the glottis was used to generate the sound source (U_g). Five Japanese vowels and the syllables of voiced consonant /b/ with the vowels were used for the simulation. The velopharyngeal opening branch in the velum model was omitted because the simulation did not include nasal sounds.

The spectra of the sound pressure that radiated from the lips and nostrils, which were determined from the simulation, are shown in Fig. 15. The left panels in the figure show lip radiation, and the right panels plot nostril radiation. The energy is concentrated in the frequency region below 1 kHz for nostril radiation. The peaks between 2 and 3 kHz

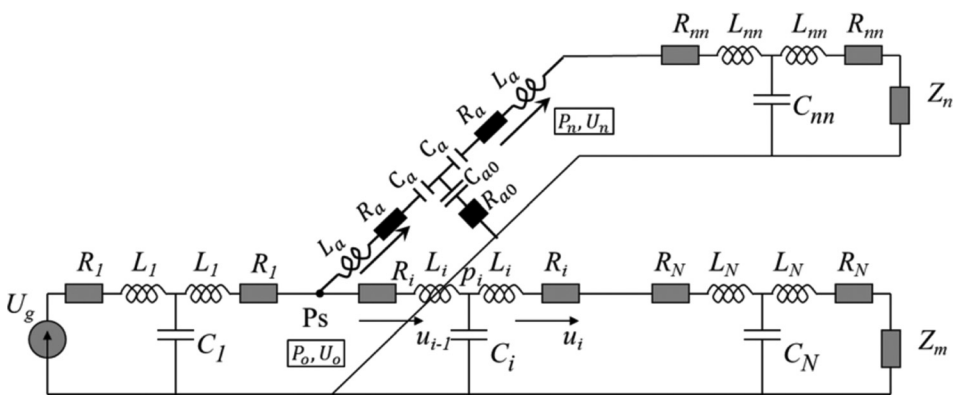

FIG. 14. Transmission line model used in simulations, where the velum model is illustrated without velopharyngeal opening branch. Ps indicates the node to obtain intraoral sound pressure.

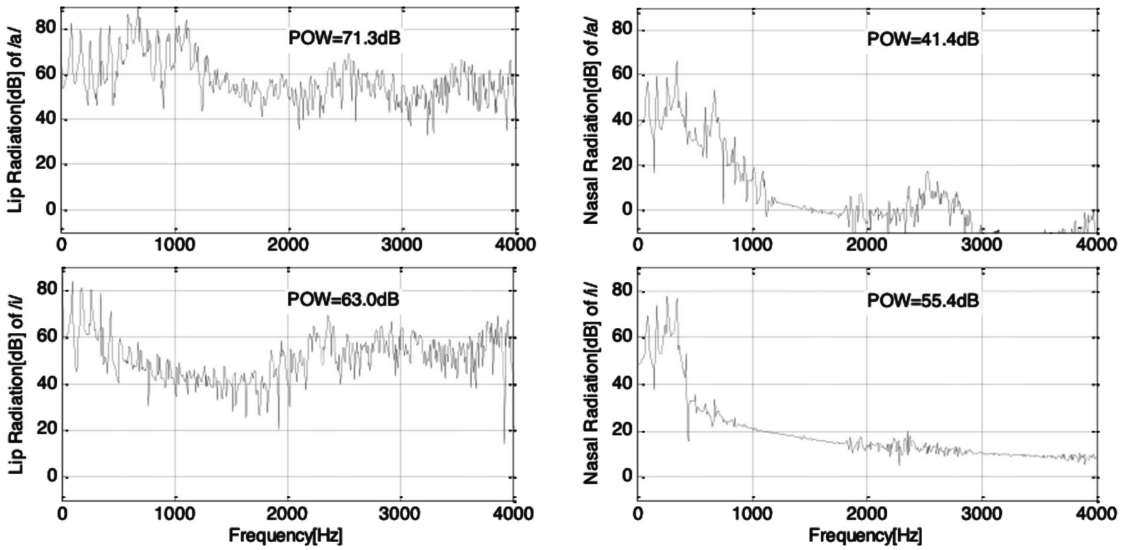

FIG. 15. Spectra of (left) lip radiation and (right) nostril radiation of sound pressures obtained at the resistance Z_m or Z_n shown in Fig. 14. Upper panels are for vowel /a/, and lower panels are for /i/. POW denotes sound pressure level.

correspond to nasal cavity resonances. The spectra of sounds radiated from the lips and nostrils are consistent with the results obtained by Dang *et al.* (1992), where they used a sound-isolation box with transmission loss of 31 dB between lips and nostrils.

The amplitudes of the lip radiation (LR) and nostril radiation (NR) signals during the simulation for the five Japanese vowels are listed in Table IV. The difference between LR and NR for vowel /i/ is only 7.6 dB, whereas that for vowel /a/ is about 30 dB. The amplitudes of the other vowels fall into the range between vowels /i/ and /a/. This simulation is consistent with the results obtained by Suzuki *et al.* (1990) for vowels /i/, /u/, and /e/. The difference between LR and NR for vowels /o/ and /a/ is about 2 dB larger than that observed by Suzuki *et al.* (1990). The difference between the simulation and observation may be due to limitations in the transmission loss of the sound-isolation box that was used in the observation.

Figure 16 plots the simulation results for the intraoral sound pressure and sound radiated from the nostrils for both the vowels and voice bars of voiced stop /b/, where the sound pressure at node Ps in Fig. 14 is treated as the intraoral sound pressure. The amplitude of subglottal air pressure during production of voice bars was set to half the value used for the vowels to simulate intraoral sound pressures that matched the actual observations as closely as possible. As a result, the intraoral sound pressure (OPb) of the voice bars was maintained at about 98 dB within a 2-dB variation, and the intraoral sound pressure (OPv) of the vowels varied from 89.8 to 96.7 dB across five Japanese vowels. The intraoral sound pressure at node Ps passed through the velum section via velar vibration and propagated to the nostrils along the nasal tract. Thus, the sound that radiated from the nostrils (NRb) during voice bars was proportional to the OPb. In contrast, the sound that radiated from the nostrils (NRv) during vowels attenuated from 41.4 to 55.4 dB, and its range of variation was twice that of the OPv. Given that the vocal tract bifurcates at the segment around the velum, nostril radiation depended on the impedance ratio of these two branches. A small impedance ratio of the nasal tract to the oral tract generates large nostril radiation, and vice versa. Since the oral tract impedance is larger for the close vowels than that for the open vowels, the former had a small impedance ratio. Thus, the close vowels generated higher amplitude for nostril-radiated sound than the open vowels.

TABLE IV. Sound amplitudes (dB) of lip radiation (LR) and nostril radiation (NR) obtained from simulations.

	/i/ (dB)	/u/ (dB)	/e/ (dB)	/o/ (dB)	/a/ (dB)
LR	63.0	67.2	67.3	71.9	71.3
NR	55.4	52.7	43.8	43.6	41.4
LR-NR	7.6	14.4	23.5	28.3	29.9

V. DISCUSSION AND CONCLUSION

As was reviewed in the earlier sections of this article, Rothenberg (1968) proposed three mechanisms that may create the differential pressure necessary for transglottal airflow to occur during stop period. These mechanisms are (1)

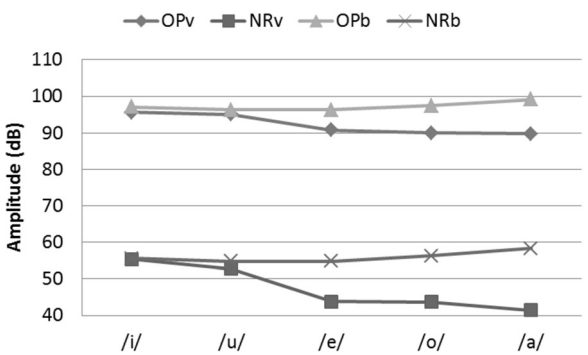

FIG. 16. Simulation results indicating the relationship between intraoral sound pressures (OPv and OPb) and sound radiated from nostrils (NRv and NRb), where the "v" denotes vowel period and "b" denotes stop period.

the active enlargement of the supraglottal cavity, (2) the passive enlargement of the supraglottal cavity, and (3) the passage of air through the velopharyngeal port with incomplete closure. A number of researchers have derived sufficient support for the first two hypotheses. The current work investigated the mechanism of voicing during the production of voiced stops with three Japanese speakers, using MRI movies. We used a pseudo-volume of the pharynx (V_{ph}), i.e., the product of the length and A-P distance of the pharynx, to investigate changes in the pharyngeal volume in the production of vowels and voiced stops. The V_{ph} monotonically decreased before the initial voiced stop as the tongue moves toward the target of the succeeding open vowels such as in the utterance /baba/. For the medial /b/ in /baba/, the V_{ph} increased by 18% for S1 and by 12% for S2. This result is close to the estimation by Stevens (2000). However, no significant increments in V_{ph} were seen during the production of voiced stops for most of the cases. In particular, the V_{ph} demonstrated no significant changes for the stop sounds in the /bibi/ utterance. Westbury (1983) found that voiced stops are always accompanied by relatively large increases in the supraglottal volume during vocal tract closure. The main difference between the two studies may be in the recruited subjects: Westbury (1983) conducted his study on an adult male speaker of American English, whereas the present research recruited three Japanese male speakers. From a linguistic point of view, the disparity in the results may be caused by cross-linguistic differences in the production of voiced stops, i.e., aspirated stops in English especially for the initial stops and unaspirated stops in Japanese. Aspiration requires intraoral air pressure to be augmented during closure.

Variations in laryngeal position were also measured in this study with reference to the position of the juncture of the hard and soft palates. Two vertical distances were measured from the juncture to the laryngeal outlet and to the vocal folds. No significant laryngeal descent was found from the two distance measures as the three subjects produced the voiced stops. Our findings do not support the hypothesis of Stevens (2000), who stated that laryngeal descent contributes to volume expansion of the pharynx.

The data from our subjects suggest that the incompletely closed velopharyngeal port of S3 caused air leakage to some extent during the production of voice bars. Such a phenomenon was also observed by Subtelny *et al.* (1969) and is basically in agreement with Rothenberg's (1968) third hypothesis. However, subjects S1 and S2 did not exhibit this behavior.

Lubker (1973) found that nasal airflow is commonly observed during the production of non-nasal stops by normal English speakers, but the observed airflow is most likely not due to velopharyngeal opening. Suzuki *et al.* (1990) and Dang *et al.* (1992) also noted that the amplitude of acoustic signals radiated from the nostrils changes with the production of non-nasalized vowels. We speculated that transvelar coupling may have been involved in the production of non-nasalized vowels and voiced stops. Motivated by this hypothesis, we designed multimodal measurements including morphological, mechanical, and acoustic observations to investigate velar functions. The MRI-based measurements revealed that velar thickness depended on vowels, but was independent of voiced stops. The velum is thick during open vowels and thin during close vowels. This study also used the velar angle as the index for evaluating speakers' efforts in producing non-nasalized sounds. An increase of the velar angle during voiced stops can be explained as the greater effort speakers actively exert in producing voiced stops. This supports the hypothesis by Lubker (1973) that nasal airflow during non-nasal sounds is caused by the elevation of the velum to a certain extent.

Velar vibration is considered to be a core mechanism for transvelar coupling during the production of non-nasal sounds. This study therefore measured such vibrations with accelerometers. The results demonstrated that the amplitude of velar vibrations in the production of voiced stops is about the same as that of velar vibrations in the production of close vowels, which is the maximum vibration of the vowels. However, the intraoral sound pressure of the voiced stops is about 10 dB lower than that of the close vowels. This implies that speakers actively control velar vibration to increase transvelar coupling during voiced stops.

Many studies (Lubker, 1973; Subtelny *et al.*, 1969; Westbury, 1983) have been conducted over the years on how speakers create the differential pressure necessary to maintain transglottal airflow during stop period, but few have investigated where voice bars are generated. Fant *et al.* (1976) measured vibrational amplitude externally on the walls of the neck and face, and found the maximum amplitude at the level of the larynx but a weaker maximum amplitude at the closed lips of the subject. However, they did not consider velar vibration that may cause sound to radiate from the nostrils. The conventional explanation is that voice bars originate from laryngeal wall vibration when vocal fold vibration takes place during voiced stops. Dang *et al.* (1993) found that the frequency of occurrence of voice bars was 89% for the initial voiced stops in Japanese syllables or words, and 100% for the medial voiced stops. For the initial voiced stops, 76% of voice bars had larger nasal radiation, and 14% of these had larger laryngeal-wall radiation. For the medial voiced stops, however, 69% of voice bars had larger laryngeal-wall radiation, and 11% of them had larger nasal radiation. Their results revealed that the contribution of nasal radiation to producing voice bars is almost the same as that of laryngeal-wall radiation. In this study, we adopted intranasal sound pressure and laryngeal-wall vibration to reexamine their contributions to voice bars. The correlations between radiated speech sounds, intranasal sound pressures, and laryngeal wall vibrations suggested that the amplitude of voice bars was less correlated to that of laryngeal-wall vibration and highly correlated to intranasal sound amplitude. This finding may challenge the conventional hypothesis that voice bars are caused by the vibration of the laryngeal wall when the vocal tract is completely closed for a stop. Note that this observation cannot refute the contribution of sound radiated from the laryngeal wall to the voice bars because the laryngeal wall vibration was obtained from only one point on the wall.

Since the sound radiated from the nostrils during voice bars results mainly from velar vibration, we developed a transvelar coupling model on the basis of the work of Dang (1992) to simulate the acoustic process in the production of non-nasal speech sounds. The velum was treated as a two-layer diaphragm, where each layer was connected to the surrounding tissues via dashpots and springs, and the two layers were coupled with a dashpot and a spring. The velum in the transmission line model was treated as a T-shaped RLC circuit. The parameters of the T-shaped RLC circuit were first estimated using MRI-based measurements and biomechanical findings (Birch and Srodon, 2009; Cheng *et al.*, 2011), and they were then optimized according to our multimodal data. The simulations revealed that the lip and nostril radiations in the model were consistent with the observations by Suzuki *et al.* (1990) and Dang *et al.* (1992). The differences in the average lip radiation between the vowels and voice bars was 12 dB in the simulations, and the difference ranged from 15 to 18 dB for the three subjects in our observations. This similarity indicated that the proposed velum model was a reasonable tool for describing the transvelar coupling in human speech production. We conducted simulations with and without velar transmission to clarify its effects. Our simulations showed that by taking velar transmission into account, a couple of pole-zeros appeared at about 2.5 kHz and/or 300 Hz on the spectra of non-nasalized vowels, and the amplitude of the first formant of /i/ increased by about 2 dB, while a smaller amplitude was observed for other vowels. These simulations were also consistent with our observations (Dang and Honda, 1996b).

The amplitude of the subglottal air pressure during the production of voice bars was set to half the value used for the vowels in the simulations to obtain the best match with the observation. Although the reasons were not clear for this phenomenon, our previous observations (Dang *et al.*, 1993) provided a supporting evidence to some extent. The observations indicated that when the intraoral air pressure ranged from −200 to 400 Pa during voice bars, the amplitude of voice bars increases as the intraoral air pressure decreases. This implies voice bars were easily generated under conditions with a relative lower intraoral air pressure.

The multimodal data on non-nasal sounds, voice bars, nasalized sounds, and nasal sounds suggested the potential effectiveness of using a uniform model to describe the general functions of the velopharyngeal segment in speech production. A framework was established (see Fig. 13) to acoustically model the velopharyngeal segment, but this study examined non-nasal sounds only. The observed data suggest that speakers actively controlled the velum in stop period, while this factor was not sufficiently taken into account in the current model. A more accurate description would require the consideration of active speaker control and variations in velar thickness. Prospective research directions would include investigations into the construction of a uniform model of the velopharyngeal segment and the mechanism for maintaining the difference in transglottal pressure when the vocal tract is completely closed during voiced stops.

ACKNOWLEDGMENTS

The authors appreciate the instructive comments and valuable suggestions given by the Associate Editor and the two reviewers. We would also like to express our appreciation to Dr. Hui Feng for her valuable comments. This work was supported in part by the National Basic Research Program of China (Grant No. 2013CB329301), and in part by the National Natural Science Foundation of China under contracts Nos. 61233009 and 61175016. This study was supported in part by JSPS KAKENHI Grant Nos. 25330190 and 25240026.

Asmar, N. H. (**2005**). *Partial Differential Equations With Fourier Series and Boundary Value Problems* (Pearson Prentice Hall, Upper Saddle River, NJ), 816 pp.

Birch, M. J., and Srodon, P. D. (**2009**). "Biomechanical properties of the human soft palate," Cleft Palate Craniofac J. **46**, 268–274.

Bjuggren, G., and Fant, G. (**1964**). "The nasal cavity structures," in *Speech Transmission Laboratory Quarterly Progress and Stains Report* (Royal Institute of Technology, Stockholm, Sweden), Vol. 5, pp. 5–7.

Cheng, S., Gandevia, S. C., Green, M., Sinkus, R., and Bilston, L. E. (**2011**). "Viscoelastic properties of the tongue and soft palate using MR elastography," J. Biomech. **44**, 450–454.

Dang, J. (**1992**). "A study on speech production model considering the vibrations of the velum and vocal tract wall," Ph.D. dissertation, in The School of Electronic Science (University of Shizuoka, Shizuoka, Japan), p. 150 (in Japanese).

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

Dang, J., and Honda, K. (**1996b**). "An improved vocal tract model of vowel production implementing piriform fossa resonance and transvelar nasal coupling," in *International Conference of Spoken Language Processing*, Philadelphia, PA, pp. 965–968.

Dang, J., and Honda, K. (**2001**). "A physiological model of a dynamic vocal tract for speech production," Acoust. Sci. Technol. **22**, 415–425.

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

Dang, J., Nakai, T., and Suzuki, H. (**1992**). "Investigation of sound radiations from the lips and the nostrils in nasalized speech," in *International Congress on Acoustics* (Beijing, China), pp. 2–7.

Dang, J., Nakai, T., and Suzuki, H. (**1993**). "Measurement and simulation of intraoral pressure and radiation of the stop consonants," J. Acoust. Soc. Jpn. **49**, 313–320 (in Japanese).

Dickson, D., and Maue-Dickson, W. (**1982**). *Anatomical and Physiological Bases of Speech* (Little, Brown, Boston, MA), 338 pp.

Fant, G., Nord, L., and Branderud, P. (**1976**). "A note on the vocal tract wall impedance," Speech Transmission Laboratory Quarterly Progress and Status Report, Royal Institute of Technology, Stockholm, Sweden, Vol. 4, pp. 13–20.

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

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

Fulop, S., and Disner, S. (**2012**). "Examining the voice bar," Proc. Meet. Acoust. **14**, 060002.

Hyon, S., Dang, J., Feng, H., Wang, H., and Honda, K. (**2014**). "Detection of speaker individual information using a phoneme effect suppression method," Speech Commun. **57**, 87–100.

Kent, R. D., and Moll, K. L. (**1969**). "Vocal tract characteristics of the stop cognates," J. Acoust. Soc. Am. **46**, 1549–1555.

Kim, S. M., McCulloch, T. M., and Rim, K. (**1999**). "Comparison of viscoelastic properties of the pharyngeal tissue: Human and canine," Dysphagia **14**, 8–16.

Lubker, J. (**1973**). "Transglottal airflow during stop consonant production," J. Acoust. Soc. Am. **53**, 212–215.

Masaki, S., Tiede, M., Honda, K., Shimada, Y., Fujimoto, I., Nakamura, Y., and Ninomiya, N. (**1999**). "MRI-based speech production study using a synchronized sampling method," J. Acoust. Soc. Jpn. **20**, 375–379.

Rossato, S., Badin, P., and Bouaouni, F. (**2003**). "Velar movements in French: An articulatory and acoustical analysis of coarticulation," in *15th International Congress of Phonetic Sciences*, edited by M.-J. Solé, D. Recasens, and J. Romero (Barcelona, Spain), pp. 3141–3144.

Rothenberg, M. (**1968**). *The Breath-Stream Dynamics of Simple-Released-Plosive Production* (Karger, Syracuse, NY), 117 pp.

Serrurier, A., and Badin, P. (**2008**). "A three-dimensional articulatory model of the velum and nasopharyngeal wall based on MRI and CT data," J. Acoust. Soc. Am. **123**, 2335–2355.

Stevens, K. (**2000**). *Acoustic Phonetics* (MIT Press, Cambridge, MA), 624 pp.

Subtelny, J. D., Kho, G. H., McCormack, R. M., and Subtelny, J. D. (**1969**). "Multidimensional analysis of bilabial stop and nasal consonants-cineradiographic and pressure flow analysis," Cleft Palate J. **6**, 263–289.

Sundberg, J., Birch, P., Gümoes, B., Stavad, H., Prytz, S., and Karle, A. (**2007**). "Experimental findings on the nasal tract resonator in singing," J. Voice **21**, 127–137.

Suzuki, H., Nakai, T., Dang, J., and Lu, C. (**1990**). "Speech production model involving subglottal structure and oral-nasal coupling through closed velum," in *International Conference of Spoken Language Processing* (Kobe, Japan), pp. 437–440.

Svirsky, M., Stevens, K., Matthies, M., Manzella, J., Perkell, J., and Wilhelms-Tricarico, R. (**1997**). "Tongue surface displacement during bilabial stops," J. Acoust. Soc. Am. **102**, 562–571.

Westbury, J. R. (**1983**). "Enlargement of the supraglottal cavity and its relation to stop consonant voicing," J. Acoust. Soc. Am. **73**, 1322–1336.