

Modeling of oropharyngeal articulatory adaptation to compensate for the acoustic effects of nasalization

Panying Rong, David P. Kuehn, and Ryan K. Shosted

Citation: *The Journal of the Acoustical Society of America* **140**, 2145 (2016);

View online: <https://doi.org/10.1121/1.4963065>

View Table of Contents: <http://asa.scitation.org/toc/jas/140/3>

Published by the *Acoustical Society of America*

Articles you may be interested in

[Influence of vocal tract geometry simplifications on the numerical simulation of vowel sounds](#)

The Journal of the Acoustical Society of America **140**, 1707 (2016); 10.1121/1.4962488

[Managing the distinctiveness of phonemic nasal vowels: Articulatory evidence from Hindi](#)

The Journal of the Acoustical Society of America **131**, 455 (2012); 10.1121/1.3665998

[Understanding the relationship between acoustics and articulation of nasal and oral vowels](#)

The Journal of the Acoustical Society of America **140**, 3219 (2016); 10.1121/1.4970150

[The effect of oral articulation on the acoustic characteristics of nasalized vowels](#)

The Journal of the Acoustical Society of America **127**, 2543 (2010); 10.1121/1.3294486

[Stimulus-directed attention attenuates lexically-guided perceptual learning](#)

The Journal of the Acoustical Society of America **140**, 1727 (2016); 10.1121/1.4962529

[Associations between tongue movement pattern consistency and formant movement pattern consistency in response to speech behavioral modifications](#)

The Journal of the Acoustical Society of America **140**, 3728 (2016); 10.1121/1.4967446

Modeling of oropharyngeal articulatory adaptation to compensate for the acoustic effects of nasalization

Panying Rong^{a)}

Department of Speech-Language-Hearing: Sciences and Disorders, University of Kansas, Lawrence, Kansas 66045, USA

David P. Kuehn

Department of Speech and Hearing Science, University of Illinois at Urbana-Champaign, Champaign, Illinois 61820, USA

Ryan K. Shosted

Department of Linguistics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

(Received 29 October 2015; revised 24 August 2016; accepted 4 September 2016; published online 30 September 2016)

Hypernasality is one of the most detrimental speech disturbances that lead to declines of speech intelligibility. Velopharyngeal inadequacy, which is associated with anatomic defects such as cleft palate or neuromuscular disorders that affect velopharyngeal function, is the primary cause of hypernasality. A simulation study by Rong and Kuehn [*J. Speech Lang. Hear. Res.* **55**(5), 1438–1448 (2012)] demonstrated that properly adjusted oropharyngeal articulation can reduce nasality for vowels synthesized with an articulatory model [Mermelstein, *J. Acoust. Soc. Am.* **53**(4), 1070–1082 (1973)]. In this study, a speaker-adaptive articulatory model was developed to simulate speaker-customized oropharyngeal articulatory adaptation to compensate for the acoustic effects of nasalization on /a/, /i/, and /u/. The results demonstrated that (1) the oropharyngeal articulatory adaptation effectively counteracted the effects of nasalization on the second lowest formant frequency (F2) and partially compensated for the effects of nasalization on vowel space (e.g., shifting and constriction of vowel space) and (2) the articulatory adaptation strategies generated by the speaker-adaptive model might be more efficacious for counteracting the acoustic effects of nasalization compared to the adaptation strategies generated by the standard articulatory model in Rong and Kuehn. The findings of this study indicated the potential of using oropharyngeal articulatory adaptation as a means to correct maladaptive articulatory behaviors and to reduce nasality. © 2016 Acoustical Society of America.

[<http://dx.doi.org/10.1121/1.4963065>]

[ZZ]

Pages: 2145–2161

I. INTRODUCTION

Hypernasality is a speech disorder that occurs when the velopharyngeal port does not close properly, resulting in excessive emission of sound energy through the nose and an imbalance of oral-nasal resonance.¹ Hypernasality is typically caused by velopharyngeal inadequacy (VPI), which can result from anatomic defects such as cleft palate or neuromuscular disorders due to the impairments of the vagus nerve and relevant muscle functions.² For individuals with cleft palate, hypernasality is the primary speech disturbance accompanied by articulatory imprecision and voice problems, all of which contribute to deterioration of speech intelligibility.³ Even following surgical repair, about 20%–30% of individuals with cleft palate still develop hypernasal speech.⁴ For individuals with motor speech disorders, resonatory disturbances characterized by hypernasality and excessive nasal emission of air have been found to be one of the most detrimental speech features that lead to declines in

speech intelligibility.^{5,6} Therefore, reducing hypernasality becomes a prioritized goal of speech therapy for resonance disorders.

Treatments of VPI include behavioral intervention, prosthetic management, and surgical intervention. Behavioral intervention is appropriate for patients who are able to compensate for VPI through speech behavioral modifications. Prosthetic management (e.g., palatal lift, nasal obturator) is appropriate for patients with consistently present VPI with moderate-severe effect on speech intelligibility and unable to make behavioral compensation. Surgical intervention (e.g., pharyngeal flap) is typically considered only after behavioral and prosthetic management have been tried and failed. Among different treatments of VPI, behavioral intervention is the most commonly used therapeutic approach, which involves a variety of strategies such as changing speaking rate, increasing vocal effort, exaggerating articulatory movements, and making speech adaptation based on biofeedback of nasal resonance.⁷ The goal of these strategies is to establish appropriate oropharyngeal articulation, oral airflow direction, and intraoral pressure build-up. Prior studies on manipulation of speaking rate and vocal

^{a)}Electronic mail: prong@ku.edu

effort have led to various findings that show inconsistent effects of rate and vocal effort adjustments on the nasality of speech.^{7–10}

Although information on direct assessment of the effect of oropharyngeal articulation on hypernasal speech is scant, studies of oral-nasal vowel contrasts in various languages demonstrate that speakers adapt both velar position and oropharyngeal articulatory placement for achieving the acoustic contrast between oral and nasal vowels.^{11–15} Oropharyngeal articulatory differences between oral and nasal vowels are evidenced in a variety of languages with phonemic nasalization. These oropharyngeal articulatory differences were found to enhance the acoustic effects of nasalization by shifting vowel formants toward the nasal antiformants such that the formants are weakened or annihilated.¹⁶ For example, Carignan and co-workers^{12,13} studied Northern Metropolitan French and identified a counter-clockwise chain-shift in the oral articulation of nasal vowels, which was demonstrated to have enhanced the effect of nasalization on F1 and F2. Shosted *et al.*¹⁵ presented evidence of tongue lowering in back nasal vowels of Hindi, which enhanced the acoustic salience of nasality. In a later study of Brazilian Portuguese, Shosted¹⁷ found a variety of differences in tongue and labial positions in nasal versus oral vowels, among which tongue body raising for the nasal vowel /ã/ was identified as one of the most robust differences that enhanced the acoustic effect of nasalization. In a larger study, Shosted *et al.*¹⁸ showed that this effect is due primarily to jaw position. The same study also suggests that the reduction of acoustic distinctiveness of Brazilian Portuguese nasal vowels as compared to their oral counterparts (i.e., centralization of nasal vowel space) is mirrored in their articulation.

In contrast to phonemic nasal vowels, a study of phonetically nasalized vowels in English found that the tongue body was raised in nasal versus oral /i/, which counteracted the raising of F1 associated with high vowel nasalization.¹¹ This difference in tongue placement between nasal and oral /i/ was argued as a compensatory strategy to attenuate the acoustic effect of nasalization and thus, resist the phonologization of vowel nasalization in English.¹¹ Maeda¹⁶ explained this kind of compensation as the outcome of motor equivalence such that the acoustic disturbances caused by one articulator (e.g., velum) can be compensated through adaptation of other articulators (e.g., tongue, jaw, lips). Although hypernasal speech is a much more complex phenomenon than nasal vowels, which affects connected speech in both production and perception mechanisms, the theoretical and experimental findings of vowel nasalization provide fundamental support for using oropharyngeal articulatory adaptation as a compensatory strategy for counteracting the acoustic effects of inappropriate velopharyngeal opening (VPO).

Rong and Kuehn^{19,20} conducted a series of exploratory studies to investigate the effects of oropharyngeal articulatory adaptation on vowel nasalization. Specifically, they simulated a variety of nasal /i/ vowels with inappropriate VPO using an articulatory synthesis model and identified a set of optimal oropharyngeal articulatory adjustments that maximally compensated for the acoustic disturbances caused by inappropriate VPO in these vowels.²⁰ These articulatory

adjustments were also demonstrated to have effectively reduced the percept of nasality for these synthetic /i/ vowels.¹⁹ As the first few studies that quantitatively examined the effect of articulatory adaptation to velopharyngeal perturbation, Rong and Kuehn^{19,20} demonstrated the possibility of using oropharyngeal articulatory strategies to reduce nasality. However, the great anatomic (e.g., size and shape of vocal tract) and articulatory (e.g., habitual way of articulation) variability across different individuals makes it difficult to generalize the articulatory strategies derived from a model based on a standard vocal tract such as the one used by Rong and Kuehn^{19,20} to a different speaker. For example, the mismatch of vocal tract size between the model and the speaker could lead to acoustic discrepancies, which provide an additional source of “acoustic error” apart from the acoustic disturbances caused by inappropriate VPO and thus, could confound the inverse mapping during articulatory adaptation. In this case, it is difficult to determine the extent to which the articulatory adaptation is targeted at compensating for inappropriate VPO versus anatomic mismatch. To address this issue, a speaker-adaptive articulatory model is needed to match the size and configuration of the speaker’s vocal tract and the range of the speaker’s articulatory movement space. The construction of this model can be achieved through two steps: (1) selection of a “standard” articulatory model to serve as the framework of the speaker-adaptive model; (2) adaptation of the standard model framework to individual speakers.

There are a variety of “standard” articulatory models, which were built based on two hypotheses: (1) the complex activities of the articulatory organs can be organized into a limited number of independently controllable functional blocks and (2) the shape of the vocal tract is determined by the state of these blocks.^{16,21} Depending on the way of partitioning the functional blocks, these articulatory models can roughly be classified into two categories. The first category uses factor analyses to derive the functional blocks as a set of element components that combined account for the bulk of variance [e.g., >90% (Ref. 16)] of a large speech inventory. The Maeda model¹⁶ is one of the most commonly used models belonging to this category, which is comprised of seven statistically extracted element components that correspond to jaw position, tongue dorsum position and shape, tongue apex position, lip opening and protrusion, and larynx height in the midsagittal plane. The second category extracts the functional blocks from vocal tract images by geometrical means. The Mermelstein model²² belongs to this category, which is comprised of seven geometrically extracted element components in control of tongue body and blade positions, jaw position, lip opening and protrusion, velar position, and hyoid height in the midsagittal plane.

Both categories of articulatory models as described above are constructed with fixed posterior-superior vocal tract outline (e.g., posterior pharyngeal wall, hard palate) and fixed constraints on the movement range of the element articulatory components. To build a speaker-adaptive model, prior studies have developed a variety of approaches to adapt the fixed parameters of a standard model to individual speakers.^{23–30} This adaptation, in general, involves two steps:

(1) adjusting the element articulatory components and the posterior-superior outline of the standard vocal tract model to match the configuration of a speaker's vocal tract in the midsagittal plane, which was defined as inverse normalizing map (NM) in McGowan and Cushing,²⁵ and (2) adjusting the parameters orthogonal to the space of NM (PONM) (e.g., length and shape of the vocal tract in the transverse plane) to match the acoustic outputs of the model and the speaker.

In this study, we first aimed to develop a speaker-adaptive articulatory model, which was customized with speaker-dependent parameters including the size and configuration of the vocal tract and the range of articulatory movement space. Specifically, we adapted the Mermelstein model²² to a speaker with a different vocal tract from the model based on measurements of eight fleshpoints on the vocal tract during a variety of speech tasks designed to sample the limits of the articulatory movement space. We selected the Mermelstein model as the framework for adaptation because its element articulatory components were geometrically extracted, which can be measured directly by means of fleshpoints on the corresponding articulators using electromagnetic tracking devices [e.g., Articulography (Carstens, Germany) and WAVE (NDI, Canada)]. Using the speaker-adaptive articulatory model developed in the first aim, we achieved the second aim, which was to derive and simulate speaker-customized oropharyngeal articulatory adaptation to compensate for the acoustic disturbances caused by inappropriate VPO.

II. METHODS

Articulatory, acoustic, and nasal aerodynamic data were collected simultaneously from a speaker during a variety of oral and nasal speech tasks. Based on the recorded articulatory and acoustic data, the Mermelstein model was adapted to the speaker following two steps. First, the element articulatory parameters of the model were adjusted to match the measured fleshpoint positions for each speech sound and the constraints of each parameter were derived accordingly based on the adaptation of all speech sounds. Second, the PONM of the vocal tract model were adjusted to match the three lowest formant frequencies of the simulated acoustic output of the model with that of the recorded acoustic signal. Given the speaker-specific articulatory constraints and PONM, the adapted articulatory model was used to synthesize three types of vowels that corresponded to oral vowels, nasal vowels, and adjusted nasal vowels (defined below). Specifically, oral and nasal vowels were synthesized through inverse mapping of the recorded acoustic signals of the corresponding vowels. Adjusted nasal vowels were synthesized to reflect oropharyngeal articulatory adjustments that aimed at compensating for the acoustic disturbances caused by inappropriate VPO. This compensation was achieved by adjusting the oropharyngeal articulatory parameters of the model given an open velopharyngeal port to minimize the discrepancies in the three lowest formant frequencies relative to that of oral vowels with a closed velopharyngeal port. The three types of vowels were compared in terms of articulatory configuration and acoustic features to assess the

effectiveness of oropharyngeal articulatory adaptation for counteracting the acoustic disturbances caused by inappropriate VPO.

A. Participant and speech tasks

An adult male speaker who has normal speech and hearing functions, is free from oral-nasal anomalies, and speaks American English as the first language participated in the study. Speech materials included /CV/, /VC/, /CVC/, and /VCV/ syllables, where “V” corresponded to three corner vowels /a/, /i/, and /u/, and “C” corresponded to stops, fricatives and nasals with different places of articulation (i.e., /p/, /b/, /t/, /d/, /k/, /g/, /f/, /v/, /s/, /z/, /m/, /n/, /ŋ/). Each consonant was involved in all possible syllable positions (i.e., initial, medial, final) with the exception of /ŋ/, which was only used in syllable final positions. A variety of consonantal contexts were included to (1) account for coarticulatory effects of consonants on vowels, (2) increase the variability of vowel articulation, and (3) explore the range of articulatory movement space. Nasal consonants, in particular, were coarticulated with the preceding/following vowels such that the velopharyngeal port stayed open during partial or entire duration of the vowel. The nasalized part of the vowel was determined based on the nasal airflow collected with a mask (discussed below in Sec. II B 3). All syllables were recorded within one of the following carrier phrases: “Say...again,” “Say...six times,” “I said...again,” and “I said...six times.”

B. Data collection

Articulatory movement, acoustic signal, and nasal airflow were collected simultaneously using an electromagnetic articulometer, a microphone, and a nasal mask, respectively (specifications of instruments are found in Secs. II B 1, II B 2, and II B 3). The three channels of signals were synchronized based on the synchronizing signal sent out by the Sybox-Opto4 unit included with the electromagnetic articulometer system.

1. Articulatory data acquisition

Articulatory movements during the assigned speech tasks were recorded in three dimensions at a sampling rate of 200 Hz using the electromagnetic articulography (EMA) AG 500 system (Carstens, Germany). Eleven electromagnetic sensors were attached to the face and articulators of the participant using surgical glue to record their movements during speech.^{11,31} Specifically, eight sensors were attached to the articulators of the participant: (1) four sensors on the tongue, including one on the tongue tip (TT), another on the tongue blade (TB) (1 cm posterior to TT), one on the posterior tongue dorsum (PTD) (the deepest position where the sensor can be placed on the tongue without causing physical discomfort of the participant), and one on the anterior tongue dorsum (ATD) (midpoint between TB and PTD), (2) two sensors on the teeth [upper incisor (UI) and lower incisor (LI)], and (3) two sensors on the lips [upper/low border of vermilion in the midline (UL/LL)]. Three additional sensors were affixed to bony and cartilaginous structures [Nose:

nose bridge, left/right tragus (LT/RT)] to serve as reference landmarks for head movement correction.

After all speech tasks were completed, all sensors except the three reference landmarks on the head were removed from the participant. An additional sensor was glued on the fingertip of the experimenter to trace the midsagittal outline of the participant's hard palate (extending posteriorly from the intersection of the soft and hard palates and anteriorly to the upper central incisors). The fingertip movement was recorded as a time-varying signal such that the position of the fingertip at each sampled point comprised the midsagittal trace of the hard palate.

2. Acoustic data acquisition

Acoustic signals were recorded at a sampling rate of 16 kHz through a Countryman Isomax E6 directional microphone placed about 5 cm from the corner of the mouth. The signal gain was modulated using an M-Audio Fast Track Pro preamplifier to an appropriate level so that the signal did not clip during the recording.

3. Aerodynamic data acquisition

Nasal airflow was collected using a vented Scicon NM-2 nasal mask.³² The mask was secured around the participant's nose to form an airtight seal. The open outlet of the mask was connected to a tube (3 m long, 4 mm ID), which passed the airflow to a Biopac TSD160A pressure transducer. The voltage output of the transducer was amplified, filtered, and then digitized at 1 kHz. The digitized signal was calibrated using a calibration function, which was determined during the calibration before the data acquisition, and then recorded by a custom MATLAB program. A more detailed description of instrumentation and data acquisition procedures was provided in Carignan *et al.*¹¹

C. Data processing

1. Head movement correction

A custom MATLAB program was developed to decouple the speech-unrelated translational and rotational movements of the head from the movements of the sensors on the articulators tracked by the AG 500. The decoupling was achieved through coordinate transformation, which converted sensor positions from the earth coordinate system to the head coordinate system, using the three sensors on the head as references. Movements of the sensors with respect to the head were regarded as speech articulatory movements and were used for further analysis.

2. Annotation

Three landmarks that corresponded to the onset, midpoint, and offset of the vowel were marked in each token based on the acoustic waveform. For nasal tokens, two additional landmarks related to the nasalization of the vowel were marked based on the nasal airflow. Specifically, a threshold at 20% above the average of the low-pass filtered nasal airflow of each token was calculated and used to

determine the timing of nasalization. In anticipatory nasalization (e.g., /bam/), the onset of nasalization was marked to be the first positive peak above the threshold that occurred after the voice onset of the vowel (see Fig. 1). The midpoint between the onset of anticipatory nasalization and offset of vowel was marked as the midpoint of nasalization. In carry-over nasalization (e.g., /mab/), the offset of nasalization was marked as the last positive peak above the threshold before the vowel offset. The midpoint between the onset of vowel and the offset of carry-over nasalization was marked as the midpoint of nasalization. If both anticipatory and carry-over nasalization occurred in a vowel (e.g., /man/), the portion with the nasal airflow above the threshold was marked to be the nasalized part of the vowel. If the nasal airflow was below the threshold during the entire vowel, the vowel was regarded as denasalized. If the nasal airflow was above the threshold during the entire vowel, the vowel was considered as completely nasalized.

Based on the articulatory signal, the following landmarks were annotated: (1) maximum vertical position of TT for /t/, /d/, /n/, /s/, /z/, (2) maximum vertical position of PTD for /k/, /g/, /ŋ/, and (3) maximum and minimum vertical distance between UL and LL for /p/, /b/, /m/. In addition, the maximum vertical position of PTD was marked for the high vowels /i/ and /u/ and the minimum vertical position of PTD was marked for the low vowel /a/.

D. Formant measurement

The three lowest formant frequencies (F1, F2, F3) were measured at the onset, midpoint, and offset of each oral vowel and at the onset, midpoint, and offset of nasalization for each nasal vowel using the speech analysis function in WaveSurfer (1.8.8p4).³³ Occasionally, spurious peaks were detected for some vowels, which were manually logged by (1) generating the spectrum of the vowel and (2) manually measuring the frequency of the harmonic closest to each formant to serve as the estimates of F1, F2, and F3. For nasal vowels, in particular, spurious nasal peaks were sometimes detected (e.g., a nasal peak close to F1 for nasal /i/ and /u/). For these nasal vowels, we compared the frequencies of all peaks detected by WaveSurfer with the F1, F2, F3 of their oral vowel counterparts. The three lowest formants of these

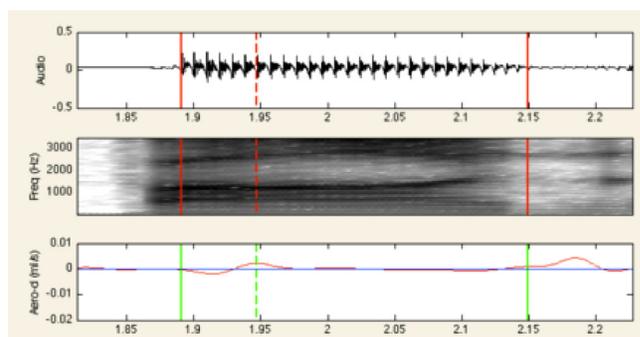


FIG. 1. (Color online) Acoustic and aerodynamic annotation scheme using /baɪ/ as an example. The three panels from top to bottom correspond to the acoustic waveform, spectrogram, and nasal airflow. The two vertical solid lines mark the onset and offset of the vowel /a/ and the vertical dashed line marks the onset of anticipatory nasalization.

nasal vowels were determined as the peaks that were closest to the F1, F2, F3 of their oral vowel counterparts, respectively, and the formant frequencies were manually logged accordingly.

E. Model adaptation

As shown by the flow chart in Fig. 2, model adaptation involved three steps: (1) adaptation of midsagittal configuration, (2) adaptation of PONM, and (3) adaptation of articulatory movement space. In the following, we explained how the three steps of model adaptation were achieved.

1. Adaptation of midsagittal configuration

The first step of model adaptation was to adapt the midsagittal configuration of the Mermelstein model to match the speaker's vocal tract. This was achieved through inverse normalizing mapping²⁵ by (1) rescaling the model and aligning the static model structures with the speaker's vocal tract and (2) adapting the 2D positions (i.e., vertical, anterior-posterior) of the fleshpoints on the primary articulators measured by the AG 500 to the model. We selected five frames of each nasal token (i.e., onset, midpoint, and offset of nasalization; maximum vertical positions of the primary articulator for consonants; maximum/minimum vertical position of PTD for high/low vowels) and five frames of each oral token (i.e., onset, midpoint, and offset of vowels; maximum vertical positions of the primary articulator for consonants; maximum/minimum vertical position of PTD for high/low vowel) to comprise the database for model adaptation.

Using the hard palate as a reference structure, we measured the length of the speaker's hard palate (pal_leng_s) based on the midsagittal trace of the palatal outline and compared it with the length of the hard palate of the model (pal_leng_m). The ratio of pal_leng_s/pal_leng_m was used as a scaling factor to rescale the model to obtain an approximate

match with the size of the speaker's vocal tract. We then applied a 2D coordinate transformation to align the positions of the posterior edge of the hard palate [marked as "M" in Fig. 3(a)] and the upper central incisors [marked as "U" in Fig. 3(a)] of the model with the onset and offset of the midsagittal palatal trace measured by the AG 500, respectively. Based on the transformation matrix, the outer contour of the model except the velar and lip regions was aligned with the speaker's vocal tract in the midsagittal plane. The inner contour of the model was determined by adapting the fleshpoint positions to the model.

In the Mermelstein model,²² the entire vocal tract was divided into six sections from the glottis to the lips [see Fig. 3(b)]. Two pharyngeal sections (AR1, AR9) extended from the glottis to the velopharynx. One posterior oral section (AR2) covered the region from the velopharynx to the posterior dorsum of the tongue. One main oral section (AR23) covered the oral cavity from the posterior tongue dorsum to tongue tip. One front oral section (AR4) covered the space between tongue tip and upper/lower central incisors. One lip section (AR5) extended from the incisors to the lips.

To adapt the fleshpoints to the model, the shape of each vocal tract section was adjusted accordingly. Specifically, the inner contours of AR1, AR9, and AR2 were determined by laryngeal height and PTD position. Because the larynx is an internal structure, which could not be tracked by the AG 500, we used the default positions of the hyoid bone for /a/, /i/, and /u/ in Childers³⁴ as estimates of laryngeal height for these three vowels. The outer contours of AR1 and AR9 were comprised of the posterior pharyngeal wall, which was considered as a static structure that was determined during the alignment of model structures. The outer contour of AR2 was determined by velar position. As the velum was not tracked by the AG 500, we made a simplified assumption on the velopharynx: (1) during oral vowels, the velopharynx stayed closed and (2) during nasal vowels, the velopharynx

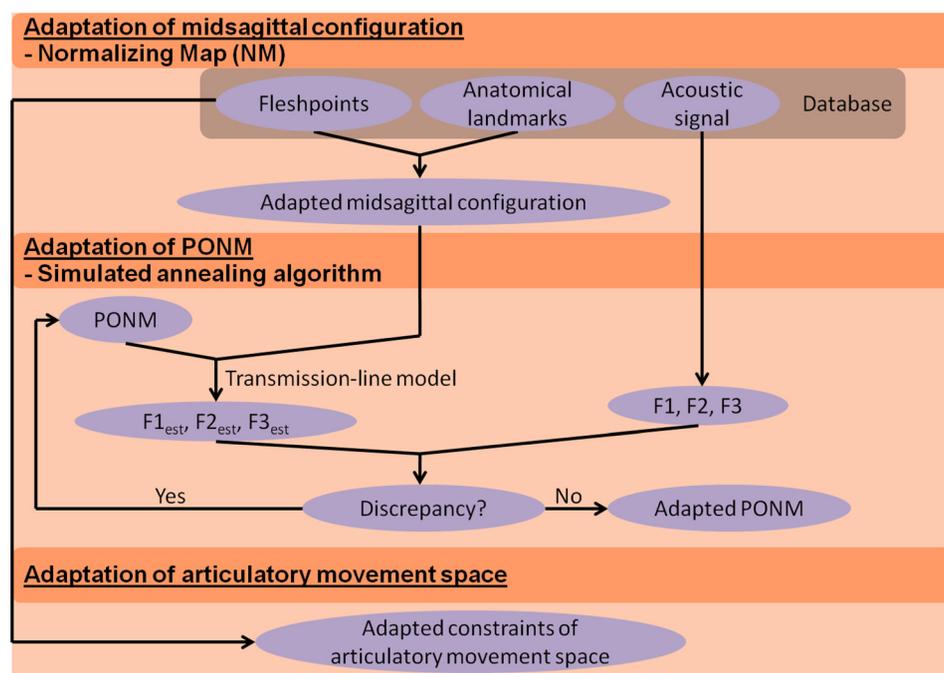


FIG. 2. (Color online) Flow chart of model adaptation.

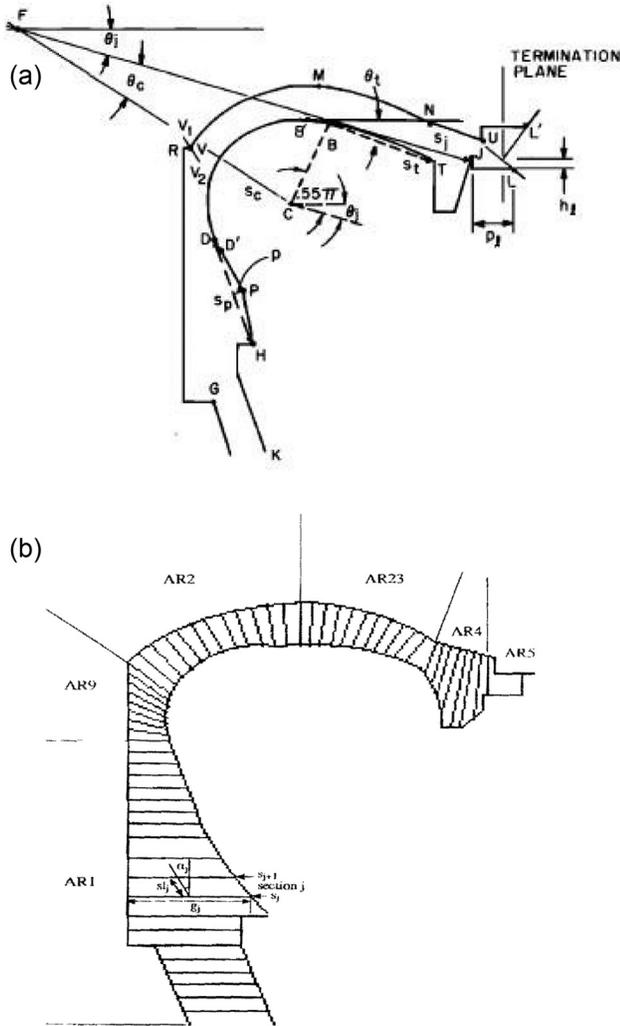


FIG. 3. The Mermelstein articulatory model: (a) model framework [p. 1071 in Mermelstein (Ref. 22)], (b) model sections [p. 400 in Childers (Ref. 34)].

stayed open with a constant VPO of 200 mm². The inner contour of AR23 was determined by the positions of four fleshpoints on the tongue (PTD, ATD, TB, TT). The inner contour of AR4 was determined by the positions of two fleshpoints (TT, LI). The outer contours of AR23 and AR4 were comprised of the hard palate, which was derived from the midsagittal palatal trace. AR5 was modeled as a tube with its inner and outer contours determined by the positions of LL and UL, respectively.

2. Adaptation of PONM

Given the adapted midsagittal configuration of the vocal tract for each speech sample based on inverse NM, the second step of model adaptation was to adapt the parameters orthogonal to the midsagittal plane of the model (i.e., PONM).²⁵

As the vocal tract (i.e., from the glottis to the lips) was simulated as 60 consecutive tubes in the Mermelstein model, the shape of the vocal tract was represented by the cross-sectional area and length of each tube, which comprised the area and length functions (each function represented by a 60-by-1 vector with each element in the vector corresponding to

a tube), respectively. The total length of the vocal tract was controlled by two PONMs (*coef_phar*, *coef_oral*). Specifically, *coef_phar* corresponded to the ratio of the pharyngeal cavity length (i.e., sum of the first 30 elements of the vector that correspond to the lengths of the tubes in the pharyngeal cavity) to the default length of the pharyngeal cavity in the model; *coef_oral* corresponded to the ratio of the oral cavity length (i.e., sum of the last 30 elements of the vector that correspond to the lengths of the tubes in the oral cavity) to the default length of the oral cavity in the model. The cross-sectional area of the vocal tract was determined by the area of the coronal sections, which was projected onto the midsagittal plane to form a grid system as shown in Fig. 3(b). The midsagittal projections were determined based on the adapted fleshpoint positions in Sec. II E 1. Cross-sectional areas were converted from the midsagittal projections using a transformation based on nine PONMs (*slp1*, *int1*, *slp2*, *slp23*, *slp4a*, *slp4b*, *slp4c*, *slp5*, *int5*) (see Appendix A for details about the transformation).

For each vowel sample, the 11 PONMs as described above were adapted to match the three lowest formant frequencies (F_1 , F_2 , F_3) calculated from the transfer function of the model with the three lowest formant frequencies of the recorded vowel at the selected time point (e.g., onset, midpoint, offset of oral vowels; onset, midpoint, offset of nasalization for nasal vowels), using the simulated annealing algorithm (SAA).³⁵ The SAA adjusted the PONMs iteratively in a trial-by-error manner, aiming at reducing the error cost as defined in Eq. (1). The iterations were stopped when the error cost was minimized; and the corresponding PONMs were regarded as the optimal solution that corresponded to the best match of PONMs between the model and the speaker. The average of the PONMs across all vowel samples were calculated to serve as the adapted PONMs of the model,

$$\text{cost} = 0.4 * \frac{\|\tilde{F}_1 - F_1\|}{F_1} + 0.4 * \frac{\|\tilde{F}_2 - F_2\|}{F_2} + 0.2 * \frac{\|\tilde{F}_3 - F_3\|}{F_3}, \quad (1)$$

where F_1 , F_2 , F_3 are the three lowest formant frequencies measured from the target vowel produced by the speaker and \tilde{F}_1 , \tilde{F}_2 , \tilde{F}_3 are the three lowest formant frequencies calculated from the transfer function of the model.

3. Adaptation of articulatory movement space

In addition to the adaptation of PONM, the constraints of the element parameters of the Mermelstein model were also adapted to match the range of the articulatory movement space of the speaker. Specifically, given the adapted articulatory positions at all five time frames for each syllable (i.e., for nasal tokens: onset, midpoint, and offset of vowel nasalization; maximum vertical positions of the primary articulator for consonants; maximum/minimum vertical position of PTDs for high/low vowels. For oral tokens: onset, midpoint, and offset of vowel; maximum vertical positions of the primary articulator for consonants; maximum/minimum vertical

position of PTDs for high/low vowels), the five element parameters of the model, which correspond to “tongue body position” ($Tbody-x$ and $Tbody-y$), “tongue tip position” ($Ttip-x$ and $Ttip-y$), “jaw position” ($Jaw-x$ and $Jaw-y$), “lip protrusion” ($Lpro$), and “lip opening” ($Lopn$), were derived accordingly. These parameters (i.e., $ART = \{Tbody-x, Tbody-y, Ttip-x, Ttip-y, Jaw-x, Jaw-y, Lpro, Lopn\}$) across all speech samples comprised a database, based on which the maximum and minimum values of each element of ART were calculated to serve as the upper and lower constraints of the corresponding parameter in the adapted model. Due to the lack of flesh-points on the larynx and velum, the constraints of hyoid bone height (H) and velic opening (VPO) remained the default values.

F. Simulation of oropharyngeal articulatory adjustment to compensate for VPO

1. Oropharyngeal articulatory adjustment

With the speaker-adaptive articulatory model as developed above, oropharyngeal articulatory placement was adjusted to compensate for the acoustic disturbances caused by inappropriate VPO. The articulatory adjustment was derived using the simulated annealing algorithm by coordinating the articulatory parameters (i.e., ART) of the speaker-adaptive model when the velopharyngeal port remained open to match the acoustic features (i.e., F1, F2, F3) calculated from the transfer function of model with that of a target oral vowel with a closed velopharyngeal port. Although the acoustic cues of nasalization are much beyond the three lowest formant frequencies,^{36–38} incorporating these additional acoustic cues into the articulatory adjustment process largely increased the computation time. Due to the exploratory nature of this study, we only considered the three lowest formant frequencies, which can be computed in an efficient manner based on the transfer functions. Specifically, given the three lowest formant frequencies of an oral vowel sample at the midpoint as the target, we first set $VPO = 200 \text{ mm}^2$ and then adjusted ART within the constraints in a trial-by-error manner to minimize the error cost defined in Eq. (1). The corresponding ART was regarded as the articulatory mapping of an adjusted nasal vowel (AN).

In addition to the adjusted nasal vowels, the articulatory mapping of two other types of vowels (i.e., oral vowels, nasal vowels) were derived as well. Specifically, given the three lowest formant frequencies of the same oral vowel sample as that used in the simulation of AN to serve as the target, we first set $VPO = 0$ and then adjusted ART using the SAA to minimize the error cost defined in Eq. (1). The corresponding ART was regarded as the articulatory mapping of an oral vowel (O). Similarly, given the three lowest formant frequencies of a nasal vowel sample at the midpoint of nasalization as the target, we first set $VPO = 200 \text{ mm}^2$ and then adjusted ART using the SAA to minimize the error cost defined in Eq. (1). The corresponding ART was regarded as the articulatory mapping of a nasal vowel (N).

2. Speech synthesis

Based on the source-filter theory,³⁹ speech was synthesized with an excitation source generated at the larynx and a filtering function determined by the shape of the vocal tract. We selected an excitation source with a fundamental frequency of $f_0 = 120 \text{ Hz}$ (corresponding to a typical male voice), a duration of 1 s, a gain of 60 dB, and a glottal waveform with an open quotient of 72.1%.³⁴ The source was applied to different filtering functions, which were determined by the articulatory mappings for O, N, and AN as well as the PONM of the model. The corresponding vowels were synthesized in MATLAB (R2011a), which yielded 186 samples (62 O + 62 N + 62 AN) of /a/, /i/, and /u/.

G. Assessing the effect of articulatory adjustment on the acoustics of nasal vowels

To assess the articulatory adjustments made by the model, we derived a set of empirical orthogonal modes that represented the articulatory patterns of AN and compared them with the orthogonal modes representative of the articulatory patterns of N. To determine the effect of the articulatory adjustments on the acoustics of nasal vowels, we calculated and compared the two lowest formant frequencies of AN and that of N.

1. Decomposition of orthogonal modes

We first derived the midsagittal configurations of the vocal tract for AN and O, which were represented by the coordinates of the 60 consecutive sections of the vocal tract. These coordinates were used to decompose a series of orthogonal modes that corresponded to the principal pattern of articulatory differences between AN and O, following a factor analysis similar to the one used by Story and Titze⁴⁰ (see Appendix B for details about the factor analysis). Each AN sample was assigned a coefficient for each orthogonal mode, which indicated the extent to which the midsagittal configuration of the AN sample differed from the average midsagittal configuration of O following the pattern of the orthogonal mode. Similarly, we also derived the midsagittal configurations of N and used them along with the midsagittal configurations of O to decompose a series of orthogonal modes that represented the principal pattern of articulatory differences between N and O, based on the same factor analysis. The coefficient assigned for each N sample for each orthogonal mode indicated the extent to which the midsagittal configuration of the N sample differed from the average midsagittal configuration of O following the pattern of the orthogonal mode. We compared the orthogonal modes for the AN-O and N-O pairs to assess what articulatory adjustments were made by the model to compensate for VPO.

2. Formant frequencies of synthetic vowels

We calculated the two lowest formant frequencies (F1, F2) of the synthetic AN, N, and O following the same procedures as in Sec. IID. ANOVAs followed by Tukey’s *post hoc* analyses were applied to compare the formant frequencies across the three types of vowels. In addition, the vowel

space for adjusted nasal vowel, nasal vowels, and oral vowels were derived and compared to assess the overall effect of articulatory adjustments across all vowels.

III. RESULTS

A. Model adaptation

1. Adaptation of midsagittal configuration

Figure 4 provides six examples of the results from the adaptation of midsagittal configuration as describe in Sec. II E 1. The midsagittal configurations of the vocal tract estimated based on the adaptation of the fleshpoints for oral /a/, nasal /a/, oral /i/, nasal /i/, oral /u/, and nasal /u/ from syllables with bilabial consonantal contexts were shown in (a)–(f), respectively.

2. Effect of model adaptation on the acoustic compensation for VPO

To assess whether the speaker-adaptive model provided better acoustic compensations for inappropriate VPO than

did the standard model used by Rong and Kuehn,^{19,20} Fig. 5 shows the spectra of 12 synthetic vowel samples, which included (1) a randomly selected oral /a/ sample, its nasal counterpart, and two adjusted nasal /a/ generated with the speaker-adaptive model and the standard model, respectively, based on the oral /a/ as the target, (2) a randomly selected oral /i/ sample, its nasal counterpart, and two adjusted nasal /i/ generated with the speaker-adaptive model and the standard model, respectively, and (3) a randomly selected oral /u/ sample, its nasal counterpart, two adjusted nasal /u/ generated with the speaker-adaptive model and the standard model, respectively.

By comparing the spectra of oral, nasal and two types of adjusted nasal vowels, we found that, in general, the spectra of both types of adjusted nasal vowels showed a closer match with the spectra of oral vowels than did the spectra of nasal vowels. Between the two types of adjusted nasal vowels, those synthesized with the speaker-adaptive model presented better compensations for the acoustic effects of inappropriate VPO, especially in the higher-frequency range of the spectra.

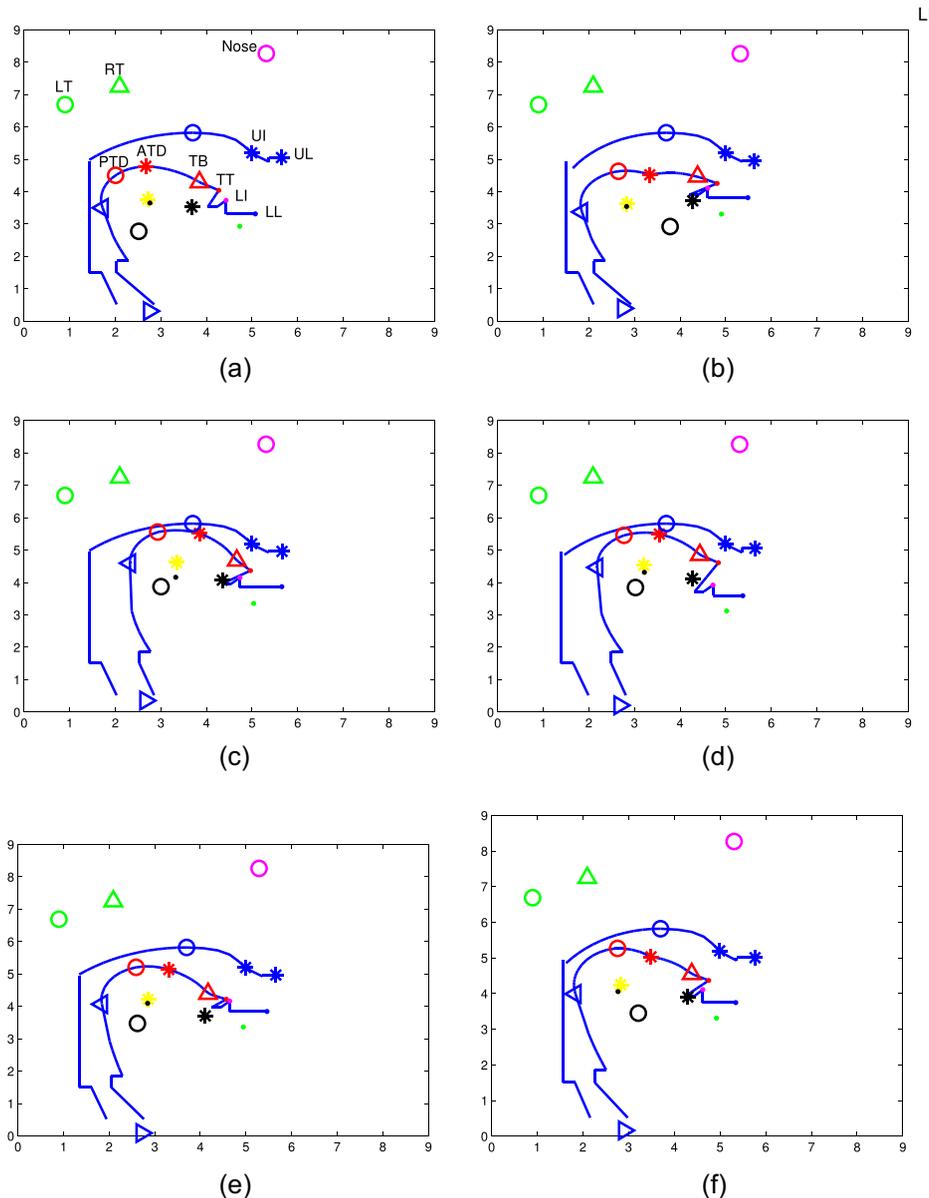


FIG. 4. (Color online) Examples of midsagittal configuration of the vocal tract estimated based on the adaptation of fleshpoints. The four fleshpoints on the tongue correspond to PTD (circle), ATD (asterisk), TB (upper triangle), and TT (dot). The two fleshpoints at the opening of the vocal tract correspond to UL (asterisk) and LL (dot), respectively. The asterisk posterior to UL represents UI and the dot posterior to LL corresponds to LI. The upper triangle and two circles above the palate correspond to three reference structures (LT, RT, Nose). The rest of the markers are the landmarks of the Mermelstein model.

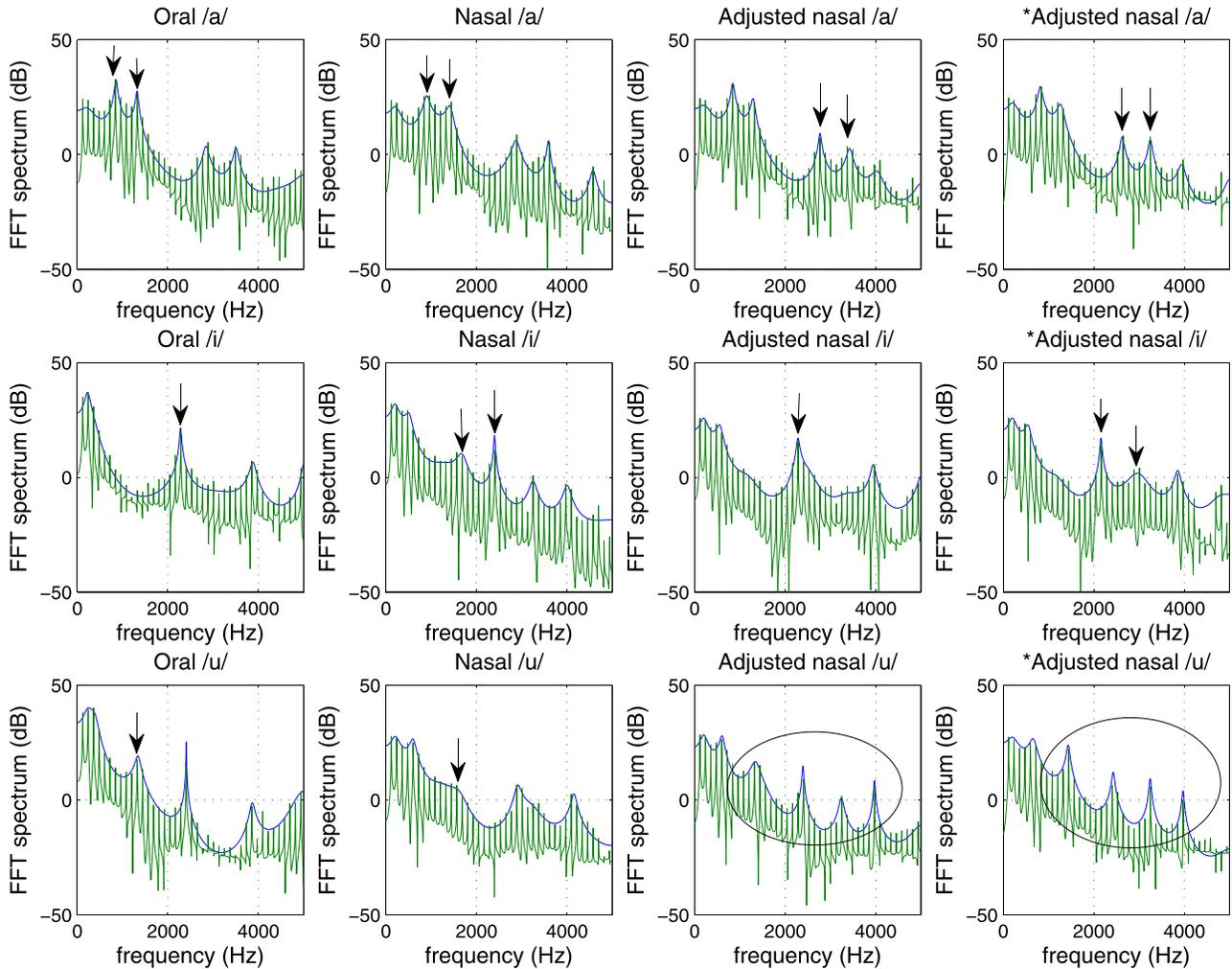


FIG. 5. (Color online) Spectra of 12 synthetic vowel samples, including (1) in the first row, an oral /a/, a nasal /a/, and two adjusted nasal /a/'s synthesized with the speaker-adaptive model and the standard model as used by Rong and Kuehn (Refs. 19 and 20) (marked with asterisks in the figure), respectively; (2) in the second row, an oral /i/, a nasal /i/, two adjusted nasal /i/'s synthesized with the speaker-adaptive model and the standard model (marked with asterisks), respectively; and (3) in the third row, an oral /u/, a nasal /u/, and two adjusted nasal /u/'s synthesized with the speaker-adaptive model and the standard model (marked with asterisks), respectively. The envelope of each spectrum was derived from the linear predictive coding analysis. The arrows in the spectra of oral and nasal vowels mark the spectral differences between each pair of oral and nasal vowels. The arrows and ellipses in the spectra of the two types of adjusted nasal vowels mark the spectral differences between the adjusted nasal vowels generated by the speaker-adaptive and standard models, respectively.

Specifically, nasal /a/ was characterized by reduced amplitudes and increased bandwidths of both F1 and F2 as compared to oral /a/, whereas the oropharyngeal articulatory adjustments in both types of adjusted nasal /a/ led to increases of F1 and F2 amplitudes and reductions of F1 and F2 bandwidths relative to nasal /a/, resulting in closer spectral matches with oral /a/. Between the two types of adjusted nasal /a/, the one synthesized with the speaker-adaptive model showed a better spectral match with oral /a/, which was not only reflected in the formant structures of F1 and F2 but also in the frequencies of higher formants such as F3 and F4 ($F3_O = 2891$ Hz, $F3_{adaptedAN} = 2773$ Hz, $F3_{standardAN} = 2651$ Hz, $F4_O = 3496$ Hz, $F4_{adaptedAN} = 3496$ Hz, $F4_{standardAN} = 3252$ Hz).

Compared to oral /i/, nasal /i/ was characterized by an increase of F2 frequency and an additional peak between F1 and F2, which corresponded to the resonance of the nasal tract. With oropharyngeal articulatory adjustments, both types of adjusted nasal /i/ showed a reduction of F2 frequency

and attenuation of the nasal peak between F1 and F2 relative to nasal /i/ and thus, presented closer spectral matches with oral /i/. However, the adjusted nasal /i/ synthesized with the standard model shows (1) an over-compensation for F2, resulting in a further reduction of F2 frequency relative to oral /i/ ($F2_O = 2290$ Hz, $F2_{standardAN} = 2168$ Hz, $F2_{adaptedAN} = 2290$ Hz), and (2) failure of compensation for the extra peak at 3013 Hz as marked by the arrow in Fig. 5. In contrast, the oropharyngeal articulatory adjustments generated by the speaker-adaptive model were able to provide better compensations from both aspects.

Compared to oral /u/, nasal /u/ was characterized by attenuation of F2, whereas both types of adjusted nasal /u/ showed an increase of F2 amplitude relative to nasal /u/, resulting in closer spectral matches with oral /u/. However, a comparison of the higher-frequency spectra of the two types of adjusted nasal /u/ suggested that the one synthesized with the standard model presented a larger discrepancy relative to oral /u/, which was marked by the ellipses in Fig. 5.

B. Articulatory configurations for O, N, and AN

Figure 6 shows the first two orthogonal modes that jointly accounted for over 85% of the variance in the difference of midsagittal configuration between AN and O. Figure 7 shows the first two orthogonal modes that corresponded to over 85% of the difference of midsagittal configuration between N and O. Table I summarized the articulatory differences between each pair of vowels (i.e., adjusted nasal /a/ vs oral /a/; nasal /a/ vs oral /a/; adjusted nasal /i/ vs oral /i/; nasal /i/ vs oral /i/; adjusted nasal /u/ vs oral /u/; nasal /u/ vs oral /u/), as represented by each of the orthogonal modes. The relationship between the articulatory differences of oral and nasal vowels and the acoustic characteristics of nasalization was discussed below in Sec. IV A. The relationship between the articulatory differences of nasal and adjusted nasal vowels and the acoustic effects of oropharyngeal articulatory adaptation was discussed in Sec. IV B.

C. Acoustic features of O, N, and AN

1. Formant frequencies

Because the two lowest formant frequencies are known to be the most important acoustic cues that disambiguate different vowels,³⁹ we analyzed F1 and F2 to assess the effects of nasalization and oropharyngeal articulatory adjustments on the acoustics of vowels. Figure 8 shows the bar plots of F1 and F2 for each type of vowel (i.e., AN, N, O). Based on a significance level of 0.05 with Bonferroni correction (i.e., $\alpha=0.05/3=0.0167$), group comparisons of F1 and F2 among O, N, and AN showed (1) significant differences in F1 among the three types of vowels for /i/ ($p<0.001$) and /u/ ($p<0.001$) and (2) significant differences in F2 among the three types of vowels for /a/ ($p<0.001$), /i/ ($p<0.001$), and /u/ ($p<0.001$). Tukey's *post hoc* tests suggested (1) significantly higher F2 for nasal /a/ with respect to oral and adjusted nasal /a/, (2) significantly

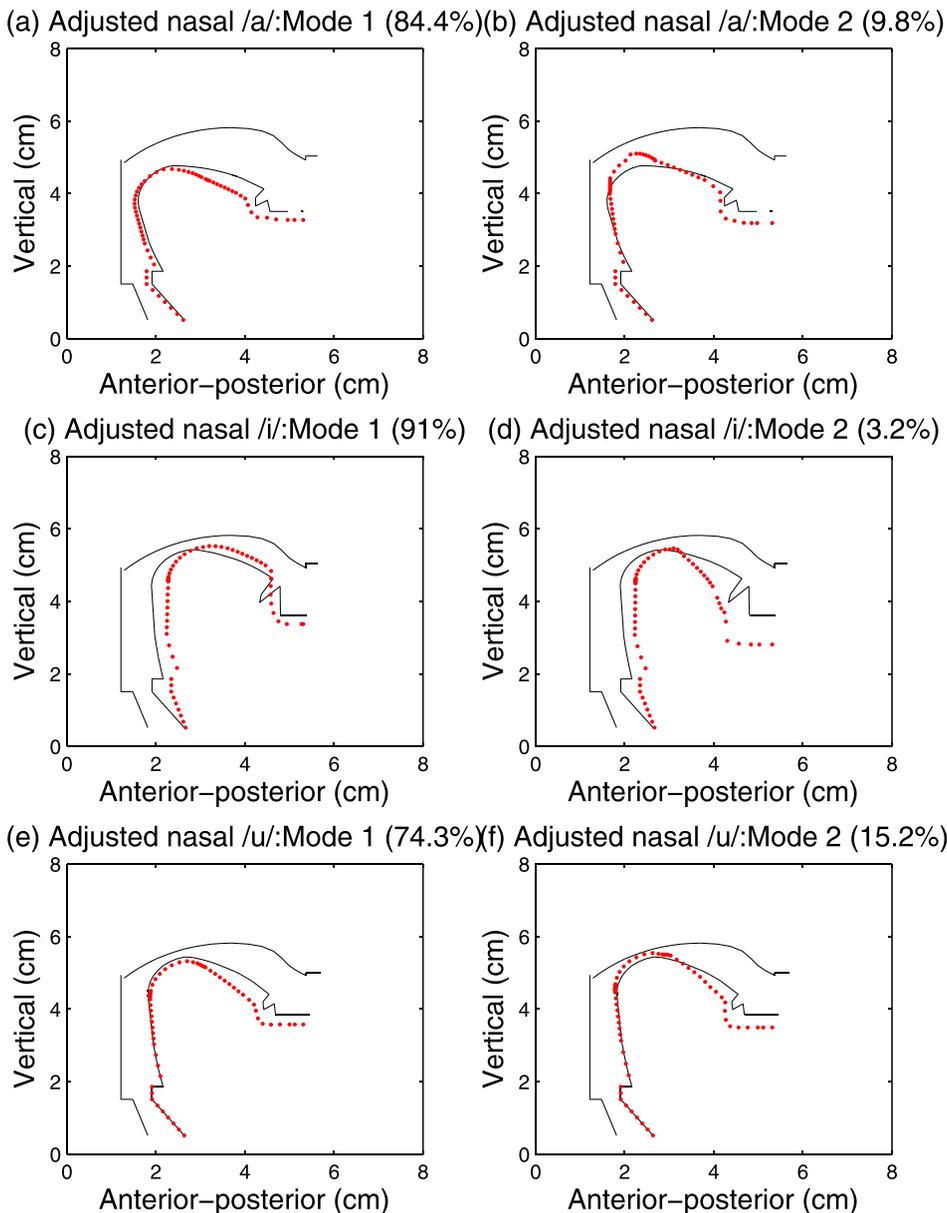


FIG. 6. (Color online) First two orthogonal modes of the articulatory difference between (a), (b) adjusted nasal /a/ (dashed) and oral /a/ (solid); (c), (d) adjusted nasal /i/ (dashed) and oral /i/ (solid); (e), (f) adjusted nasal /u/ (dashed) and oral /u/ (solid). The percentage numbers within the parentheses correspond to the percentage of variance accounted for by each orthogonal mode.

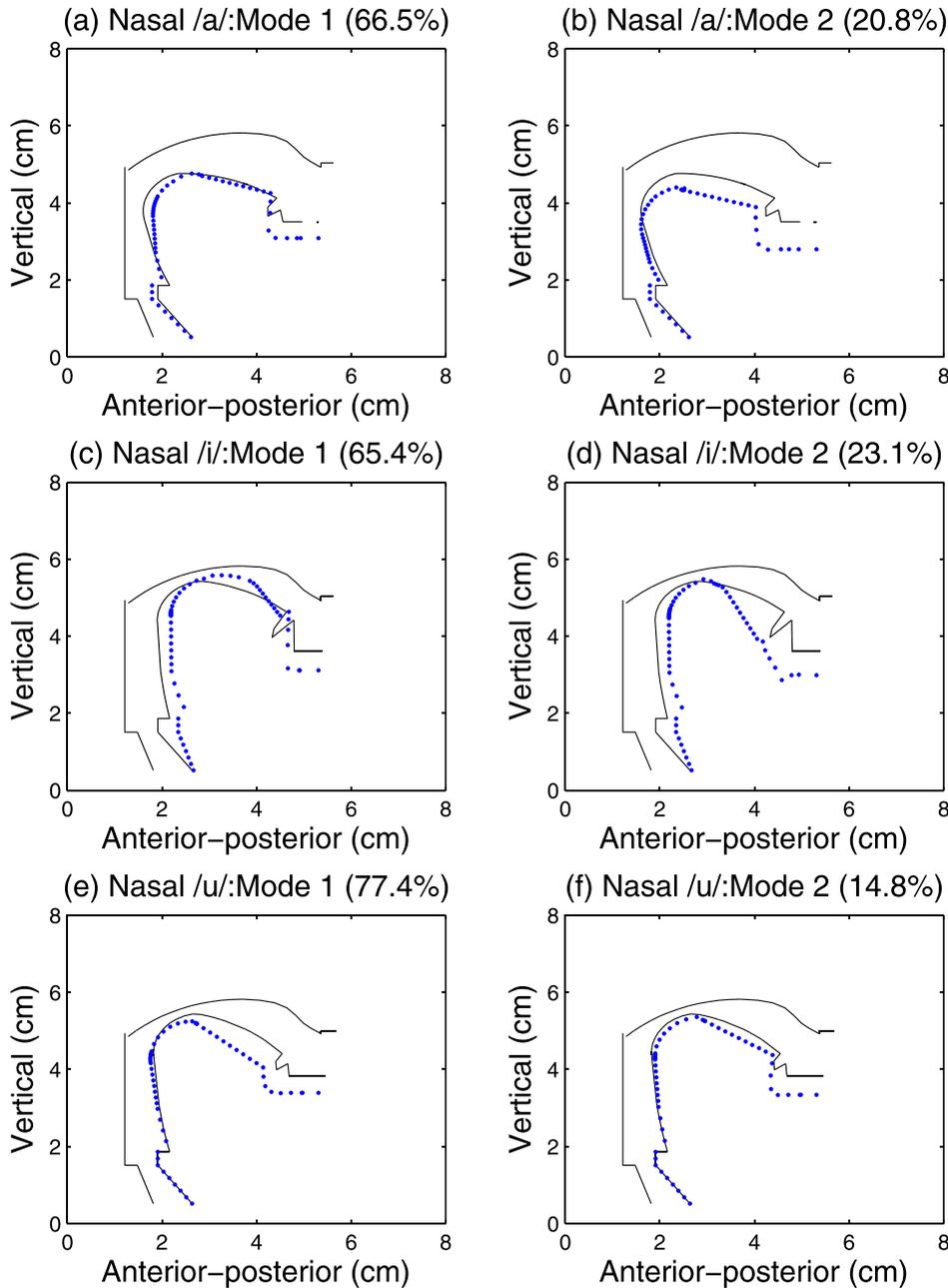


FIG. 7. (Color online) First two orthogonal modes of the articular difference between (a), (b) nasal /a/ (dashed) and oral /a/ (solid); (c), (d) nasal /i/ (dashed) and oral /i/ (solid); (e), (f) nasal /u/ (dashed) and oral /u/ (solid). The percentage numbers within the parentheses correspond to the percentage of variance accounted for by each orthogonal mode.

higher F1 for nasal /i/ and adjusted nasal /i/ relative to oral /i/, (3) significantly higher F2 for nasal /i/ compared to oral and adjusted nasal /i/, (4) significantly higher F1 for nasal and adjusted nasal /u/ with respect to oral /u/, (5) significantly higher F2 for nasal /u/ compared to oral and adjusted nasal /u/.

2. Vowel space

In addition to individual formants, we also examined the vowel space for each type of vowel (i.e., AN, N, and O) based on the averages of F1 and F2 across all samples of /a/, /i/, and /u/, respectively (see Fig. 9). The comparison of

TABLE I. Primary articulatory differences between (adjusted) nasal vowels and oral vowels as represented by each of the first two orthogonal modes.

Vowel	First orthogonal mode	Second orthogonal mode
Adjusted nasal /a/	lowering of tongue blade and jaw; retraction of tongue root	raising of tongue body
Nasal /a/	fronting of tongue root; larger lip opening	lowering of tongue; larger lip opening
Adjusted nasal /i/	raising and fronting of tongue	bulging of tongue; larger lip opening
Nasal /i/	fronting of tongue root; raising of tongue blade; larger lip opening	bulging of tongue; larger lip opening
Adjusted nasal /u/	lowering of tongue blade	raising of tongue body; lowering of tongue tip
Nasal /u/	lowering of tongue blade and jaw; larger lip opening	larger lip opening

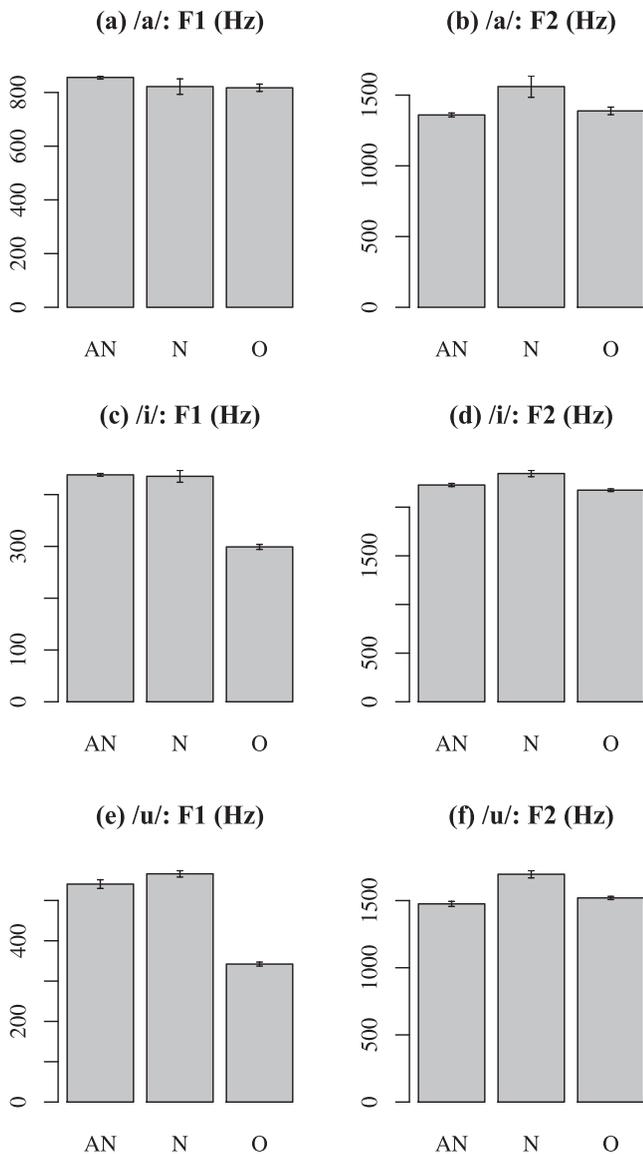


FIG. 8. The two lowest formant frequencies for synthetic /a/, /i/, and /u/: (a) F1 for /a/, (b) F2 for /a/, (c) F1 for /i/, (d) F2 for /i/, (e) F1 for /u/, (f) F2 for /u/.

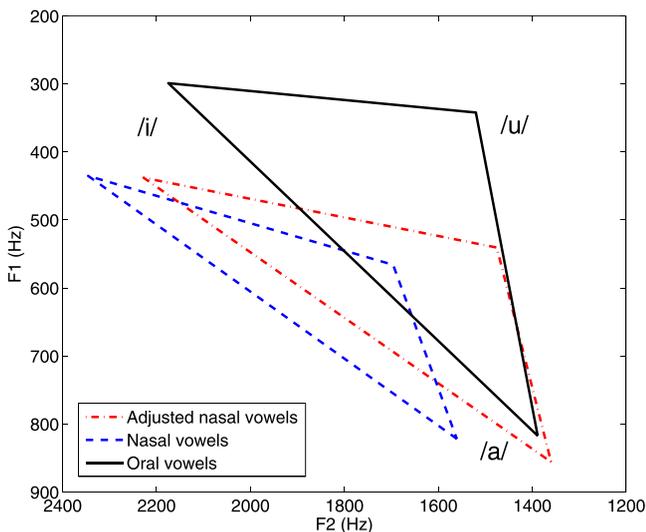


FIG. 9. (Color online) Vowel space for synthetic O (solid), N (dashed), and AN (dash-dotted).

vowel space between O and N indicated the acoustic effect of nasalization, which was characterized by an overall shift of the vowel space toward higher F1 and F2 as well as a contraction of vowel space area. Comparison of vowel space between AN and N suggests that the articulatory adjustments for AN resulted in a shift of the vowel space toward lower F2 and an increase of vowel space area relative to that of N, but the F1 of high vowels (i.e., /i/ and /u/) were comparable between N and AN, which were both higher than the F1 for O.

IV. DISCUSSION

In this study, we developed a speaker-adaptive articulatory model through (1) adaptation of fleshpoints on the primary articulators of a speaker to the framework of the Mermelstein model²² to match the midsagittal configurations of the model with the speaker's vocal tract, (2) adjusting the parameters orthogonal to the midsagittal plane of the model to match the three lowest formant frequencies of the simulated acoustic output of the model with that of the speaker through inverse filtering, and (3) adjusting the constraints of the model parameters to match the articulatory movement space of the model with that of the speaker. With the speaker-adaptive model, we simulated oral and nasal vowels and compared their articulatory and acoustic characteristics. To compensate for the acoustic effects caused by inappropriate VPO, we adjusted the oropharyngeal articulatory configuration of the speaker-adaptive model given an open velopharynx to minimize the acoustic discrepancy relative to the oral vowel targets. The corresponding oropharyngeal articulatory adjustments were used to generate the articulatory mapping for adjusted nasal vowels. The oropharyngeal articulatory placement and acoustic features of the adjusted nasal vowels were compared with that of oral and nasal vowels to assess the effectiveness of oropharyngeal articulatory adjustments for counteracting the acoustic effects of nasalization.

A. Articulatory and acoustic characteristics of nasal vowels

To determine the acoustic correlates of nasality, previous studies have explored a variety of acoustic features, among which altered formant structures (e.g., reduced amplitude, increased bandwidth, shifted frequency) in the low-frequency spectrum, increased nasal resonances (e.g., extra peaks in the spectrum), and reduced overall intensity were identified as the primary acoustic characteristics of nasal vowels that distinguish them from oral vowels.^{36–38,41,42} In this study, we analyzed a number of synthetic nasal vowels and identified various acoustic characteristics of nasalization that were consistent with prior findings. As demonstrated by the examples in Fig. 5, these characteristics included reduced amplitude of F1 and/or F2 (e.g., /a/, /u/), increased bandwidth of F1 and F2 (e.g., /a/), shifted frequency of F2 (e.g., /i/), and an additional peak between F1 and F2 (e.g., /i/). Statistical analysis focusing on the two lowest formant frequencies indicated that the effects of nasalization on /a/, /i/, and /u/ were characterized by (1) an overall increase of F2

across all vowels and (2) an increase of F1 for high vowels (/i/ and /u/) (see Figs. 8 and 9).

These acoustic characteristics of nasal vowels were attributable to both VPO and oropharyngeal articulatory placement. As shown in Fig. 4, oral and nasal vowels demonstrated a variety of articulatory differences in addition to VPO. In these examples, nasal /a/ showed a more anterior tongue position relative to its oral counterpart; nasal /i/ and /u/ showed larger lip opening with respect to their oral counterparts (Fig. 4). Across all samples, as indicated by the orthogonal modes in Fig. 7 and Table I, nasal vowels were characterized by more anterior tongue positions (e.g., /a/, /i/) and larger lip openings (e.g., /a/, /i/, /u/) compared to the oral vowels. These oropharyngeal articulatory differences between nasal and oral vowels led to a variety of acoustic consequences that enhanced the effect of nasalization.

For back vowels /a/ and /u/, F2 (i.e., typically within the range of 1200 and 1500 Hz) was lower than but proximate to the resonant frequency of the nasal cavity (i.e., about 1500 Hz, as shown in Fig. 5), which resulted in a weakening of F2 due to the bypass of acoustic energy to the nasal cavity. Through comparison of the oropharyngeal articulation between nasal and oral vowels as shown in Fig. 7, the fronting of the tongue in nasal /a/ resulted in an upward shift of F2 toward the nasal resonance, which led to a further weakening of F2 and thus, enhanced the effect of nasalization. For round vowel /u/, the extent of lip rounding was reduced as a result of increased lip opening (Fig. 7) such that F2 was shifted upward toward the nasal resonance, which led to a further weakening of F2 and, in turn, an enhancement of nasalization. For front vowel /i/, F2 (i.e., above 2000 Hz) was above the resonant frequency of the nasal cavity such that the nasal resonance resulted in a prominent peak in a relatively flat region between F1 and F2 in the spectrum as shown in Fig. 5. Through comparison of the oropharyngeal articulation between nasal and oral /i/ as shown in Fig. 7, the fronting of the tongue and the increase of lip opening in nasal /i/ led to an upward shift of F2 away from the nasal resonance, which preserved the amplitude and bandwidth of F2. The fronting of the tongue, however, led to reduced volume and increased impedance of the oral cavity such that a larger proportion of the acoustic energy was bypassed to the nasal cavity, resulting in an enhancement of the nasal peak between F1 and F2 (Fig. 5).

Consistent with the findings of this study, prior studies of nasal vowels of different languages (e.g., French, Hindi, Portuguese) also identified a variety of oropharyngeal articulatory strategies to enhance the acoustic effects of nasalization, which facilitated the phonologization of vowel nasalization.^{12–15,17,43} On the contrary, a study of English vowel /i/ provided evidence of compensatory articulation to resist the phonologization of vowel nasalization.¹¹ These findings along with the findings of this study indicated an active role of oropharyngeal articulation for achieving the acoustic goal (e.g., either to enhance or to attenuate nasalization), which lays the groundwork for using oropharyngeal articulatory strategies to compensate for the acoustic disturbances caused by inappropriate VPO.

B. Oropharyngeal articulatory adaptation to compensate for the acoustic effects of VPO

As shown in Fig. 6, the articulatory strategies generated by the speaker-adaptive model varied across different vowels. Yet, all of the strategies led to compensations for the acoustic effects of nasalization on F2. Specifically, the tongue was more posterior for adjusted nasal /a/ compared to nasal /a/ [Figs. 6(a) vs 7(a)], which resulted in a shift of F2 toward lower frequency [Fig. 8(b)] and thus, counteracted the effect of nasalization. For adjusted nasal /i/ and /u/, lip openings were smaller than that of nasal /i/ and /u/ [Figs. 6(c), 6(e) vs 7(c), 7(e)], which lowered the frequency of F2 [Figs. 8(d), 8(f)] and, in turn, counteracted the effect of nasalization. In addition, the examples in Fig. 5 suggested that the peak between F1 and F2 related to the nasal resonance was attenuated by the articulatory adjustments for adjusted nasal /i/.

The effects of nasalization on F1 for /i/ and /u/ were, however, failed to be compensated by the oropharyngeal articulatory adjustments generated by the model. These effects were related to a low-frequency peak (e.g., around 250 Hz; see Fig. 5), which was related to the resonance of the paranasal sinuses⁴¹ and was identified as one of the most prominent acoustic cues of nasalization by previous studies.^{37,44} This low-frequency peak was in proximity to the F1 for high vowels /i/ and /u/, which, in turn, shifted F1 upward toward higher frequencies (Figs. 5 and 8), whereas the F1 for low vowel /a/ was at a much higher frequency than this low-frequency peak and was thus, less affected.

The failure of compensation for the upward shifting of F1 for high vowels might be attributed to two factors: (1) the lack of data on the velum, pharynx, and nasal tract and (2) the relatively conservative estimates of articulatory movement range of the model. Because the articulography did not provide data on the velopharynx, VPO was set at a constant area of 200 mm² for all nasal and adjusted nasal vowels, which might be an over-estimation for high vowels, considering that high vowels typically required less velopharyngeal coupling to be perceived as nasal as low vowels.⁴⁵ This potential over-estimation of VPO resulted in restricted space in the posterior part of the vocal tract and thus, limited the extent to which the tongue can be raised to lower F1 for /i/ and /u/. Due to the lack of data on the pharynx, the pharyngeal configuration of the model was estimated based on the biomechanical associations between the posterior part of the tongue and the pharynx, which limited the degrees of freedom for pharyngeal movement. Because F1 is known to be affected by not only the vertical position of the tongue but also the configuration of the pharynx,⁴⁶ the restrictions on the degrees of freedom for pharyngeal movement could have limited the extent to which F1 can be manipulated to counteract the effect of nasalization. Due to the lack of geometrical data on the nasal tract, the size and configuration of the nasal cavity and paranasal sinuses were estimated based on the measurements from the speaker in Pruthi *et al.*^{41,42} Because the acoustic features of nasal vowels are dependent on VPO, oropharyngeal configuration, as well as nasal tract geometry,⁴¹ the discrepancy of the geometry of nasal tract

between Pruthi *et al.*^{41,42} and the speaker in this study could lead to inaccurate estimates of nasal resonant frequencies, which might affect the effectiveness of the oropharyngeal articulatory adjustments. In addition, during model adaptation, we constrained the range of articulatory movements within the limits determined by the samples in the database, which ensured that the articulatory adjustments generated by the model were physically producible, but as a trade-off, these constraints resulted in a relatively conservative estimate of articulatory movement space and might have limited the extent to which the tongue can be raised to lower F1.

Across all vowels (/a/, /i/, /u/), oropharyngeal articulatory adjustments partially counteracted the acoustic effects of nasalization. Specifically, following articulatory adjustments, the vowel space was shifted backward (i.e., toward lower F2) and the size of the vowel space was increased relative to that of nasal vowels such that the adjusted nasal vowel space overlapped to a larger extent with the oral vowel space than did the nasal vowel space (Fig. 9). With further improvements of the model by incorporating measurements of VPO, pharyngeal configuration, and nasal tract, we expect the downward shift of /i/ and /u/ in the vowel space could also be compensated through oropharyngeal articulatory adjustments.

C. Pattern of oropharyngeal articulatory adaptation

Despite the difference of oropharyngeal articulatory adaptation strategies across different vowels, the pattern of articulatory adjustments shared some common features as indicated by the orthogonal modes in Fig. 6. Specifically, the first orthogonal modes corresponded to global articulatory adjustments such as changing tongue frontness to enhance phoneme distinctiveness. Because a general effect of nasalization on vowels is to make the nasal vowels acoustically less distinct than their oral vowel counterparts (e.g., distance between different vowels in the vowel space becomes smaller; and an overall reduction of vowel space area),⁴⁷⁻⁵¹ the global articulatory adjustments represented by the first orthogonal mode counteracted this effect of nasalization by enhancing the distinctiveness of nasal vowels. For example, the tongue was more posterior for the back vowel /a/ [Fig. 6(a)] and more anterior for the front vowel /i/ relative to their oral counterparts [Fig. 6(c)] to counteract the reduction of front-back vowel distinctiveness due to nasalization (Fig. 9). The second orthogonal modes corresponded to refined articulatory adjustments to enhance or attenuate the regional adjustments involved in the first orthogonal mode, aiming at tuning the acoustic outcome of the global articulatory adjustments. These refined adjustments were achieved through positioning of lips and jaw and coordinating different parts of the tongue. For example, the posterior part of the tongue was raised in /a/ and /u/ [Figs. 6(b), 6(f)] and the anterior part of the tongue as well as the lower lip and jaw were lowered in /i/ [Fig. 6(d)].

D. Effect of model adaptation

To assess whether the speaker adaptation of the model led to more effective articulatory strategies for counteracting the acoustic effects of VPO, we compared the spectra of

adjusted nasal vowels generated by the adapted model with those generated by the standard Mermelstein model. The examples in Fig. 5 suggested that the two types of adjusted nasal vowels presented a variety of differences in both low-frequency (e.g., F1, F2) and high-frequency (e.g., F3 and higher formants) spectral regions. Specifically, the adjusted nasal /a/ generated by the standard model showed larger discrepancies in F2 (e.g., reduced amplitude), F3 (e.g., lower frequency), and F4 (e.g., lower frequency) relative to the oral /a/ than did the adjusted nasal /a/ generated by the adapted model. The adjusted nasal /i/ generated by the standard model had a lower F2 frequency and an additional peak between F2 and F3 compared to the oral /i/, whereas the adjusted nasal /i/ generated by the adapted model presented a closer match in the spectral region above 2000 Hz with the oral /i/. As for /u/, both types of adjusted nasal vowels presented some discrepancies in the spectrum relative to that of the oral /u/, but the adjusted nasal /u/ generated by the adapted model showed overall smaller discrepancies in the spectral region of F2 and higher formants.

The comparisons of the two types of adjusted nasal vowels as shown in Fig. 5, although descriptive, indicated that speaker-specific articulatory adjustments might be more efficacious for counteracting the acoustic effects of VPO, particularly in the spectral region of F2 and higher formants. The high-frequency spectral components, although have not been assessed in the current study due to a trade-off of computational efficiency, contribute to the overall spectral balance, which was found as a secondary cue of nasality in addition to the low-frequency formants.³⁷ In follow-up studies, the effects of articulatory adjustments on the overall spectral balance need to be quantified and compared between the adjusted nasal vowels generated by the adapted model and those generated by the standard model. More importantly, perceptual ratings of nasality for these two types of adjusted nasal vowels are needed to determine whether the articulatory adjustments generated by the speaker-adaptive model lead to perceivable reductions of nasality compared to that of the standard model.

E. Implications and future directions

According to motor equivalence,^{16,52} the deviation from the acoustic goal as a result of disturbances in one articulator (e.g., inappropriate VPO) can be compensated by adjusting and coordinating other articulators. In this study, we demonstrated that the oropharyngeal articulatory adjustments generated by the speaker-adaptive model partially compensated for the acoustic effects of nasalization on /a/, /i/, and /u/, which shed light on the use of oropharyngeal articulatory adaptation to reduce hypernasality, but further investigations on other speech sounds and connected speech are warranted.

Due to the exploratory nature of this study, we made a number of assumptions during the modeling process to simplify the mechanisms of speech production and to improve computational efficiency. For example, because of the lack of data on the pharynx, we assumed the configuration of the pharyngeal cavity was determined by the shape and position of the posterior tongue body and tongue root without taking

into account the active movements of the pharynx. Considering that one type of velopharyngeal opening-closing pattern is achieved through the sphincteric strategy,⁵³ which involves active movements of the lateral and posterior pharyngeal walls, restricting these active pharyngeal movements limits the versatility of behavioral strategies that can be developed to maximally compensate for the acoustic effects of inappropriate VPO. Due to the lack of data on the velopharynx, we assumed the velopharyngeal port stayed open at a constant area of 200 mm² for all nasal and adjusted nasal vowels. However, in real speech, VPO is known to differ across vowels and moreover, vary during the time course of vowel nasalization. As discussed above, the simplification of VPO might have restricted the articulatory adjustments to compensate for the nasalization effect on F1 for high vowels /i/ and /u/. In future studies, we will focus on improving the articulatory model by integrating measurements of pharyngeal configuration, nasal tract, and VPO. Non-invasive techniques such as real-time MRI allow us to record the static structure of the nasal tract and dynamic movements of the entire vocal tract, which was used by Carignan *et al.* to study the oropharyngeal articulation of nasal vowels.¹³ We will explore the possibility of integrating this technique into our modeling approach in follow-up studies.

Another focus for future studies is to improve the inverse mapping of the model. In the current study, inverse mapping was based on the three lowest formant frequencies, which were derived directly from the transfer function. This apparent simplification of the acoustic representation of vowel nasalization improved computational efficiency but, as a trade-off, other acoustic cues of nasalization apart from the shifting of formant frequencies were not taken into account during inverse mapping. These acoustic cues, including reduction of formant amplitudes, increase of nasal resonances, and change of spectral balance across low- and high-frequency components,^{36–38} need to be quantified and included in the cost function of inverse mapping in the improved model. Meanwhile, to accommodate the additional computational cost related to the computation of these acoustic cues, more efficient algorithms than the SAA need to be developed accordingly.

In follow-up studies, we will also conduct perceptual tests to assess the effect of oropharyngeal articulatory adaptation on the percept of nasality. The perceptual ratings of

nasality will allow us to determine (1) whether the articulatory adaptation strategies are effective for reducing nasality and (2) whether the speaker-adaptive model provides more effective articulatory strategies for reducing nasality than does the standard model.

ACKNOWLEDGMENTS

This study was supported by the Campus Research Board at the Graduate College of the University of Illinois, Urbana-Champaign. We thank Torrey Loucks, Mark Hasegawa-Johnson, and Chilin Shih for their support and feedback to this study. We also thank Christopher Carignan, who assisted with data collection.

APPENDIX A: CONVERSION OF MIDSAGITTAL PROJECTIONS TO CROSS-SECTIONAL AREAS

In AR1 and AR9,

$$A = \frac{\pi}{2} * (slp1 * (gc/gw) + int1) * d * cf, \quad (A1)$$

where A is the cross-sectional area of the “tubes” within AR1 or AR9, gc and gw are two constants related to the geometry of the pharynx, d is the midsagittal projection, and cf is the cosine of the angle between the cross-section of the tube and the midline of the vocal tract.

In AR2,

$$A = slp2 * d^{1.5} * cf, \quad (A2)$$

where A is the cross-sectional area of the “tubes” within AR2, d is the midsagittal projection, and cf is the cosine of the angle between the cross-section of the tube and the midline of the vocal tract.

In AR23,

$$A = slp23 * d^{1.5} * cf, \quad (A3)$$

where A is the cross-sectional area of the “tubes” within AR23, d is the midsagittal projection, and cf is the cosine of the angle between the cross-section of the tube and the midline of the vocal tract.

In AR4,

$$A = cf * \begin{cases} slp4a * d, & d < 0.5, \\ slp4b * (d - 0.5) + 0.5 * slp4a, & 0.5 \leq d \leq 2, \\ slp4c * (d - 2) + (1.5 * slp4b + 0.5 * slp4a), & d > 2, \end{cases} \quad (A4)$$

where A is the cross-sectional area of the “tubes” within AR4, d is the midsagittal projection, and cf is the cosine of the angle between the cross-section of the tube and the midline of the vocal tract.

In AR5,

$$A = \frac{\pi}{2} * [d * (slp5 * (sl - pl) + int5)] * cf, \quad (A5)$$

where A is the cross-sectional area of the “tubes” within AR5, sl and pl are two parameters related to lip protrusion and opening, d is the midsagittal projection, and cf is the cosine of the angle between the cross-section of the tube and the midline of the vocal tract.

APPENDIX B: FACTOR ANALYSIS FOR DECOMPOSITION OF ORTHOGONAL MODES

The decomposition of orthogonal modes followed a procedure similar to the one used by Story and Titze,⁴⁰ which was illustrated by the following equations:

$$\alpha(v, s) = A(v, s) - A_0(v), \quad (\text{B1})$$

where $v = 1, 2, 3$ corresponding to the three vowels /a/, /i/, and /u/, respectively, $A(v, s)$ is the area function of the s th sample of N or AN , and $A_0(v) = (1/M) \sum_{s=1}^M A(v, s)$ is the average of area functions of all of the M samples of O :

$$R_{ij} = \frac{1}{M-1} \sum_{s=1}^M \alpha(v_i, s) \alpha(v_j, s), \quad (\text{B2})$$

where $i, j = 1, 2, \dots, N$ ($N = 60$) corresponds to the index of the vocal tract section,

$$R\phi = \phi I \lambda, \quad (\text{B3})$$

where R is the covariance matrix as determined in Eq. (B2), I is the identity matrix, ϕ and λ are the orthogonal modes (i.e., eigenvectors) and the corresponding eigenvalues, respectively,

$$pp(i, s) = \sum_{j=1}^N \alpha(v_j, s) \phi_i(v_j), \quad (\text{B4})$$

where $i = 1, 2, \dots, N$ corresponds to the i th mode; and $pp(i, s)$ is the coefficient, which is derived as the projection of the s th α vector on the i th mode.

¹D. P. Kuehn and L. J. Henne, “Speech evaluation and treatment for patients with cleft palate,” *Am. J. Speech Lang. Pathol.* **12**(1), 103–109 (2003).

²W. R. Zemlin, *Speech and Hearing Science: Anatomy and Physiology*, 4th ed. (Pearson, Boston, MA, 1997), p. 451.

³A. W. Kummer, *Cleft Palate and Craniofacial Anomalies: Effects on Speech and Resonance*, 2nd ed. (Thomson Delmar Learning, Clifton Park, NY, 2011), pp. 180–191.

⁴B. J. McWilliams, “The long-term speech results of primary and secondary surgical correction of palatal clefts,” in *Multidisciplinary Management of Cleft Lip and Palate*, edited by J. Bardach and H. L. Morris (Saunders, Philadelphia, 1990), pp. 815–819.

⁵R. D. Kent, G. Weismer, J. F. Kent, and J. C. Rosenbek, “Toward phonetic intelligibility testing in dysarthria,” *J. Speech Hear. Disorders* **54**(4), 482–499 (1989).

⁶R. D. Kent, J. F. Kent, G. Weismer, R. L. Sufit, J. C. Rosenbek, R. E. Martin, and B. R. Brooks, “Impairment of speech intelligibility in men with amyotrophic lateral sclerosis,” *J. Speech Hear. Disorders* **55**(4), 721–728 (1990).

⁷M. A. McHenry, “The effect of increased vocal effort on estimated velopharyngeal orifice area,” *Am. J. Speech-Lang. Pathol.* **6**(4), 55–61 (1997).

⁸A. M. Goberman, J. C. Selby, and H. R. Gilbert, “The effects of changes in speaking rate on nasal airflow and the perception of nasality,” *Folia Phoniatr. Logop.* **53**, 222–230 (2001).

⁹M. A. McHenry, “Aerodynamic, acoustic, and perceptual measures of nasality following traumatic brain injury,” *Brain Injury* **13**(4), 281–290 (1999).

¹⁰T. Watterson, S. L. York, and S. C. McFarlane, “Effects of vocal loudness on nasalance measures,” *J. Commun. Disorders* **27**(3), 257–262 (1994).

¹¹C. Carignan, R. Shosted, C. Shih, and P. Rong, “Compensatory articulation in American English nasalized vowels,” *J. Phon.* **39**(4), 668–682 (2011).

¹²C. Carignan, “An acoustic and articulatory examination of the ‘oral’ in ‘nasal’: The oral articulations of French nasal vowels are not arbitrary,” *J. Phon.* **46**, 23–33 (2014).

¹³C. Carignan, R. K. Shosted, M. Fu, Z.-P. Liang, and B. P. Sutton, “A real-time MRI investigation of the role of lingual and pharyngeal articulation in the production of the nasal vowel system of French,” *J. Phon.* **50**, 34–51 (2015).

¹⁴O. Engwall, V. Delvaux, and T. Metens, “Interspeaker variation in the articulation of nasal vowels,” in *Proceedings of the 7th ISSP* (2006), pp. 3–10.

¹⁵R. Shosted, C. Carignan, and P. Rong, “Managing the distinctiveness of phonemic nasal vowels: Articulatory evidence from Hindi,” *J. Acoust. Soc. Am.* **131**(1), 455–465 (2012).

¹⁶S. Maeda, “Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model,” in *Speech Production and Speech Modelling*, edited by William J. Hardcastle and Alain Marchal (Springer, Netherlands, 1990), Vol. 55, pp. 131–149.

¹⁷R. K. Shosted, “Nasal vowels are not [+nasal] oral vowels,” in *Romance Linguistics 2012: Selected papers from the 42nd Linguistic Symposium on Romance Languages (LSRL)*, edited by J. Smith and T. Ihsane (Jon Benjamins, Amsterdam, 2015), pp. 63–76.

¹⁸R. K. Shosted, D. Pozzani, F. Meneses, N. Wong, Z. Hermes, and T. Loucks, “The articulatory space of oral and nasal vowels in Brazilian Portuguese,” poster presented at the *15th Conference on Laboratory Phonology*, Ithaca, NY, 2016.

¹⁹P. Rong and D. P. Kuehn, “The effect of articulatory adjustment on reducing hypernasality,” *J. Speech Lang. Hear. Res.* **55**(5), 1438–1448 (2012).

²⁰P. Rong and D. P. Kuehn, “The effect of oral articulation on the acoustic characteristics of nasalized vowels,” *J. Acoust. Soc. Am.* **127**(4), 2543–2553 (2010).

²¹D. Beaufemps, P. Badin, and G. Bailly, “Linear degrees of freedom in speech production: Analysis of cineradio- and labio-film data and articulatory-acoustic modeling,” *J. Acoust. Soc. Am.* **109**(5), 2165–2180 (2001).

²²P. Mermelstein, “Articulatory model for the study of speech production,” *J. Acoust. Soc. Am.* **53**(4), 1070–1082 (1973).

²³Z. Al Bawab, L. Turicchia, R. M. Stern, and B. Raj, “Deriving vocal tract shapes from electromagnetic articulograph data via geometric adaptation and matching,” in *INTERSPEECH’09*, 2009, pp. 2051–2054.

²⁴B. Mathieu and T. Laprie, “Adaptation of Maeda’s model for acoustic to articulatory inversion,” in *Eurospeech* (1997), pp. 2015–2018.

²⁵R. S. McGowan and S. Cushing, “Vocal tract normalization for midsagittal articulatory recovery with analysis-by-synthesis,” *J. Acoust. Soc. Am.* **106**(2), 1090–1105 (1999).

²⁶M. Naito, L. Deng, and Y. Sagisaka, “Model-based speaker normalization methods for speech recognition,” *Electron. Commun. Jpn. Part II: Electron.* **86**, 45–56 (2003).

²⁷S. Panchapagesan and A. Alwan, “A study of acoustic-to-articulatory inversion of speech by analysis-by-synthesis using chain matrices and the Maeda articulatory model,” *J. Acoust. Soc. Am.* **129**(4), 2144–2162 (2011).

²⁸A. Toutios, S. Ouni, and Y. Laprie, “Estimating the control parameters of an articulatory model from electromagnetic articulograph data,” *J. Acoust. Soc. Am.* **129**(5), 3245–3257 (2011).

²⁹M. Wilson, “Fitting an articulatory model with EMA data: Toward an evaluation of speech inversion,” Master’s thesis, Universität des Saarlandes, Saarbrücken, Germany, 2008.

³⁰R. Winkler, S. Fuchs, P. Perrier, and M. Tiede, “Biomechanical tongue models: An approach to studying inter-speaker variability,” in *INTERSPEECH’11* (2011), pp. 273–276.

³¹Y. Yunusova, J. R. Green, and A. Mefferd, “Accuracy assessment for AG500, electromagnetic articulograph,” *J. Speech Lang. Hear. Res.* **52**(2), 547–555 (2009).

³²M. Rothenberg, “Measurement of air flow in speech,” *J. Speech Hear. Res.* **20**, 155–176 (1977).

- ³³L. Deng, X. Cui, R. Pruvencok, J. Huang, S. Momen, Y. Chen, and A. Alwan, "A database of vocal tract resonance trajectories for research in speech processing," in *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings* (2006), Vol. 1.
- ³⁴D. G. Childers, *Speech Processing and Synthesis Toolboxes* (Wiley, New York, 2000), pp. 211–246.
- ³⁵S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, "Optimization by simulated annealing," *Science* **220**(4598), 671–680 (1983).
- ³⁶M. Chen, "Acoustic correlates of English and French nasalized vowels," *J. Acoust. Soc. Am.* **102**(4), 2360–2370 (1997).
- ³⁷S. Hawkins and K. N. Stevens, "Acoustic and perceptual correlates of the non-nasal-nasal distinction for vowels," *J. Acoust. Soc. Am.* **77**, 1560–1575 (1985).
- ³⁸R. Kataoka, D. Warren, R. Mayo, D. J. Zajaz, and R. W. Lutz, "The relationship between spectral characteristics and perceived hypernasality in children," *J. Acoust. Soc. Am.* **109**(5), 2181–2189 (2001).
- ³⁹G. Fant, *The Acoustic Theory of Speech Production* (Mouton, the Hague, 1960), pp. 15–20.
- ⁴⁰B. H. Story and I. R. Titze, "Parameterization of vocal tract area functions by empirical orthogonal modes," *J. Phon.* **26**, 223–260 (1998).
- ⁴¹T. Pruthi, "Analysis, Vocal-tract modeling, and automatic detection of vowel nasalization," Ph.D. thesis, University of Maryland, College Park, 2007.
- ⁴²T. Pruthi, C. Y. Espy-Wilson, and B. H. Story, "Simulation and analysis of nasalized vowels based on magnetic resonance imaging data," *J. Acoust. Soc. Am.* **121**(6), 3858–3873 (2007).
- ⁴³J. P. Zerling, "Phénomènes de nasalité et de nasalization vocaliques: Étude cinéradiographique pour deux locuteurs" ("Phenomena of birth and nasalization vowel: Cineradiographic study for two speakers"), *Trav. l'Inst. Phon. Strasbourg* **16**, 241–266 (1984).
- ⁴⁴S. Maeda, "The role of the sinus cavities in the production of nasal vowels," in *Proceeding of ICASSP* (1982), pp. 911–914.
- ⁴⁵F. Bell-Berti, "Understanding velic motor control: Studies of segmental context," in *Nasals, Nasalization, and the Velum*, edited by M. K. Huffman and R. A. Krakow (Academic Press, New York, 1993), pp. 63–85.
- ⁴⁶J. S. Perkell and W. L. Nelson, "Variability in production of the vowels /i/ and /a/," *J. Acoust. Soc. Am.* **77**, 1889–1895 (1985).
- ⁴⁷C. A. Ferguson, "Assumptions about nasals: A sample study in phonological universals," in *Universals of Language*, edited by J. Greenberg (MIT Press, Cambridge, 1963), pp. 53–60.
- ⁴⁸M. Ruhlen, "Nasal vowels," in *Universals of Human Language*, edited by C. A. Ferguson and E. A. Moravcsik (Stanford University Press, Stanford, 1978), Vol. 2, pp. 203–241.
- ⁴⁹J. T. Wright, "The behavior of nasalized vowels in the perceptual vowel space," in *Experimental Phonology*, edited by J. J. Ohala and J. J. Jaegerpages (Academic Press, Orlando, FL, 1986), pp. 45–67.
- ⁵⁰Z. S. Bond, "Identification of vowels excerpted from neutral and nasal contexts," *J. Acoust. Soc. Am.* **59**(5), 1229–1232 (1976).
- ⁵¹B. Mohr and W. S.-Y. Wang, "Perceptual distance and the specification of phonological features," *Phonetica* **18**, 31–45 (1968).
- ⁵²O. M. Hughes and J. H. Abbs, "Labial-mandibular coordination in the production of speech: Implications for the operation of motor equivalence," *Phonetica* **33**(3), 199–221 (1976).
- ⁵³A. Amelot, L. Crevier-Buchman, and S. Maeda, "Observations of the velopharyngeal closure mechanism in horizontal and lateral directions from fiberoptic data," in *Proceedings of 15th ICPHS* (2003), pp. 3021–3024.