

# Using ultrasound and nasalance to separate oral and nasal contributions to formant frequencies of nasalized vowels

Christopher Carignan<sup>a)</sup>

Institut für Phonetik und Sprachverarbeitung, Ludwig-Maximilians-Universität München, Schellingstraße 3,  
80799 Munich, Germany

(Received 17 December 2017; revised 26 March 2018; accepted 11 April 2018; published online 2 May 2018)

The experimental method described in this manuscript offers a possible means to address a well known issue in research on the independent effects of nasalization on vowel acoustics: given that the separate transfer functions associated with the oral and nasal cavities are merged in the acoustic signal, the task of teasing apart the respective effects of the two cavities seems to be an intractable problem. The proposed method uses ultrasound and nasalance to predict the effect of lingual configuration on formant frequencies of nasalized vowels, thus accounting for acoustic variation due to changing lingual posture and excluding its contribution to the acoustic signal. The results reveal that the independent effect of nasalization on the acoustic vowel quadrilateral resembles a counter-clockwise chain shift of nasal compared to non-nasal vowels. The results from the productions of 11 vowels by six speakers of different language backgrounds are compared to predictions presented in previous modeling studies, as well as discussed in the light of sound change of nasal vowel systems. © 2018 Acoustical Society of America. <https://doi.org/10.1121/1.5034760>

[JFL]

Pages: 2588–2601

## I. INTRODUCTION

According to the source-filter theory of speech production, the acoustic output of a phonated vowel is considered to arise from the modulation of the voice source (the larynx) by the acoustic transfer function of the filter (the supralaryngeal vocal tract), resulting in a glottal spectrum that is shaped into a sequence of broadband peaks. The spectral consequences of this modulation are generally well understood for vowels that involve spectral filtering by a single transfer function—e.g., oral vowels, wherein the acoustic transfer function is associated with the singular, continuous cavity formed in the vocal tract from the glottis to the lips. However, for vowels that involve the combination of multiple transfer functions, the spectral modulation by the supralaryngeal vocal tract is arguably less well understood. Vowel nasalization, for example, is produced by lowering the velum, allowing air to resonate simultaneously through both the oral and nasal cavities. When these cavities are coupled via the velopharyngeal (VP) port—a process known as VP coupling or, more commonly, nasalization—their respective acoustic transfer functions are merged, resulting in a wide range of modifications to the acoustic spectrum in comparison to the non-VP-coupled vocal tract. While theoretical predictions for these acoustic modifications have been described in depth (Fujimura and Lindqvist, 1971; Maeda, 1993; Feng and Castelli, 1996; Chen, 1997, *inter alia*), comparatively less attention has been paid to observable effects in physiological data. The relative lack of production

evidence is, perhaps, due to the problematic nature of VP coupling itself:

“The difficulty in studying the acoustic behaviour of nasal vowels arises from the fact that we do not know what exactly characterizes a nasal sound. Thus it is very difficult, if not impossible, to make a comparison between a simulation result and a real nasal spectrum, even with a measured nasal transfer function” (Feng and Castelli, 1996, p. 3695).

To pose the problem more directly: how can researchers test the “real world” practical validity of VP-coupled acoustic models when the separate acoustic effects of the oral and nasal cavities are conjoined in production data? Once the separate oral and nasal tract transfer functions have merged in the acoustic signal, how can we tease apart their respective spectral effects? The current study represents an attempt to address this issue by using ultrasound and nasalance technologies to separate the oral and nasal contributions to formant frequencies of VP-coupled vowels, allowing observation of frequency modulations that are due to VP coupling itself, without the influence of variation in tongue posture. The method is used here to observe the acoustic effect of VP coupling in production data from six speakers with different language backgrounds, and the results are compared with predictions derived in a *post hoc* fashion from figures appearing in various modeling studies (Fujimura and Lindqvist, 1971; Maeda, 1993; Feng and Castelli, 1996; Serrurier and Badin, 2008).

### A. Vowel nasalization and formant frequency shifts

When VP coupling occurs, the increased surface area (the nasal cavity and sinuses, in addition to the pharynx and oral cavity) and soft, mucousy tissues along the walls of the nasal cavity absorb acoustic energy to a greater degree than

---

<sup>a)</sup>Also at: The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Sydney, Australia. Electronic mail: [c.carignan@phonetik.uni-muenchen.de](mailto:c.carignan@phonetik.uni-muenchen.de)

occurs in the oropharyngeal tract alone, resulting in an overall reduction in formant amplitude and widening of formant bandwidths for nasalized vowels compared to oral vowels (Stevens, 2000, p. 193). Additional transfer function poles (spectral resonances) and zeros (spectral anti-resonances) arising from the coupling of the nasopharyngeal side branch to the oral main branch are also introduced to the spectrum. Maeda (1993) suggests that the frequency of the first nasal pole ( $N1$ ) generally falls within the range of  $\approx 250$ – $400$  Hz (for males), while the frequency of the first nasal zero ( $Z1$ ) increases from  $\approx 400$  to  $\approx 900$  Hz as the aperture of the VP port (i.e., the degree of nasalization) increases. Although acoustic damping has been shown to generally flatten the spectrum in the region of  $F1$ – $F2$  (Maeda, 1982a), the asymmetry in the frequency ranges of constructive energy (from  $N1$ ) and destructive energy (from  $Z1$ ) in the  $F1$  region has been shown to result in increased spectral tilt for nasalized vowels compared to oral vowels (Styler, 2017).

While these acoustic consequences of VP coupling affect the spectrum in a broad manner, variation in formant frequencies is predicted to occur due to the interaction of the all-pole oral transfer function with the pole-zero nasal transfer function. This interaction is expected to affect formant frequencies of the combined transfer function in three primary ways (schematized examples appear in Fig. 1):

1. The addition of spectral energy from a nasal pole can shift acoustic energy surrounding an oral pole if the two poles are near to each other in frequency, as in Fig. 1(a).
2. The partial cancellation of the spectral energy of an oral pole from a nasal zero can result in the residual oral pole energy having a center frequency that is shifted relative to the non-VP-coupled oral pole, as in Fig. 1(b).
3. The complete cancellation of the spectral energy of an oral pole from a nasal zero can result in an annihilation of the oral pole energy, wherein the nasal pole becomes the exclusive pole in the combined transfer function, as in Fig. 1(c).

It is, thus, not surprising that the way in which the pole-zero pairs affect the spectrum of VP-coupled vowels is often obscure, since the predictions of formant frequency shifts differ among vowels, depending on the acoustic transfer function of the oral tract and on the degree of nasalization. Nevertheless, there have been a number of important studies which have predicted the effect of VP coupling on various

vowel qualities, primarily using tube models of the coupled naso-oro-pharyngeal vocal tract. The majority of the results from these studies, however, appear only in the form of figures displaying predicted acoustic transfer functions (sometimes hand-drawn), without quantified frequency measurements. This makes it not only difficult to predict the explicit magnitude of formant frequency shifts, but also to compare the results across studies in a meaningful way. In the following section, four of these studies—Fujimura and Lindqvist (1971), Maeda (1993), Feng and Castelli (1996), and Serrurier and Badin (2008)—are described and, in the interest of synthesizing and clarifying their predictions for formant frequency shifts of VP-coupled vowels, frequency measurements (in Hz) have been made directly from digitally scanned copies of figures that appear in each of the manuscripts.

### 1. Deriving formant measurements from previous studies

The earliest of the four studies, Fujimura and Lindqvist (1971), employed physical excitation of the vocal tract transcutaneously at a point just above the glottis, passing a sweep-tone signal through a moving-coil-type electromagnetic transducer pressed against the neck, and placing a microphone near the lips, in order to measure transfer function characteristics. Using this method, transfer functions were calculated for the participants' productions of sustained vocal tract configurations for a number of speech sounds, including nasalized vowels. In order to measure the difference in transfer functions between VP-coupled and non-VP-coupled vowels, the participants articulated alternately nasalized and non-nasalized versions of vowels "with approximately the same tongue articulation" (p. 553). This method was carried out for one speaker's articulation of "a neutral vowel [ʒ] [sic]" (p. 553), and for another speaker's articulation of [i], [a], and [u]. The resulting transfer functions appear in their manuscript in Fig. 14(a)–14(b) for the neutral vowel, and in Fig. 15 (first and third panels of each sub-figure) for the three corner vowels.

In order to obtain frequency measurements from these transfer functions for the purposes of the current study, the figures were digitally scanned using a Ricoh MP C3003 multifunction commercial printer and opened as portable document format (.PDF) documents at 300 dpi resolution in the GNU Image Manipulation Program ([the] GIMP; The GIMP

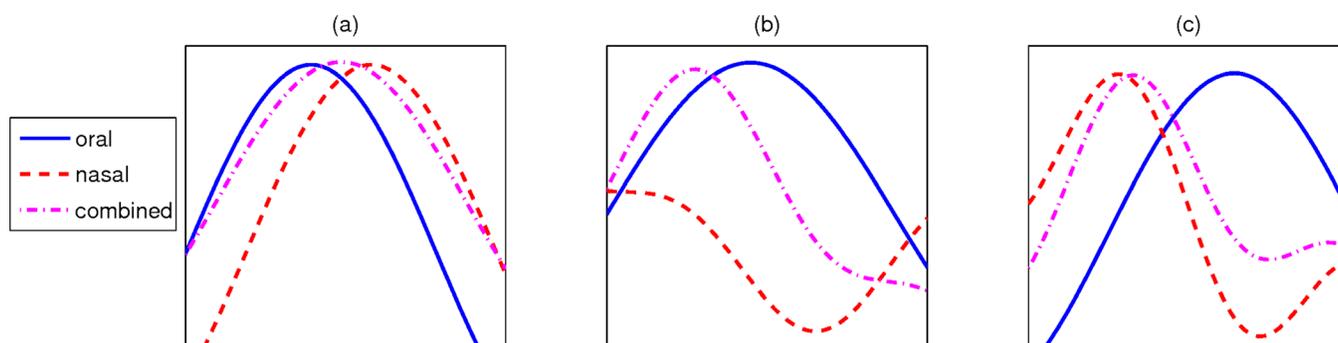


FIG. 1. (Color online) Schematic representations of three types of formant frequency shifts in nasalized vowels.

Development Team, 2017). The scanned figures were first rotated manually to be parallel to the  $x$ - and  $y$ -axis in the image program, using grid lines as a visual guide. Using a cross-hair indicator, the cursor was aligned precisely with the  $x$ -axis tick marks for 100, 200, 500, 1000, and 2000 Hz (the transfer functions were plotted on a log-transformed frequency scale in the manuscript). Once aligned, the pixel index in the  $x$ -dimension was logged for each tick mark. The same process was used to locate the  $x$ -dimension pixel value for the transfer function peaks in the corresponding figure.  $F1$  and  $F2$  were logged for oral transfer functions;  $N1$ ,  $F1$ , and  $F2$  were logged for combined oral-nasal transfer functions. For each figure, the frequency values and corresponding pixel values for the  $x$ -axis tick marks were imported to R (R Core Team, 2016), and the pixel values were inverse-log-transformed before creating a linear model of the relation between the two dimensions. The model was subsequently used to determine the frequency values for the transfer function peaks, using the inverse-log-transformed  $x$ -dimension pixel values of the peaks as predictor variables. The result is a precise frequency value (in Hz) for the peaks that appear in the acoustic transfer functions printed in the manuscript figures.

Maeda (1993) used a model originally proposed in Maeda (1982b) to calculate the oral and nasalized transfer functions for the vowels /i e o u/. The model incorporates a physiological change to the shape of the oral tract, in which an increase in the VP port aperture corresponds to a concomitant decrease in the oral tract aperture at the velum (due to velum being lowered toward the tongue dorsum). This model, therefore, necessarily results in a change to the oral tract transfer function that is separate from (and in addition to) any changes in the combined transfer function that are due to interaction with nasal pole-zero pairs. The resulting non-VP-coupled and VP-coupled acoustic transfer functions for the five vowels appear in the manuscript in Figs. 3(a)–3(e). Formant measurements were obtained from these figures in the same manner as described above, with  $x$ -axis tick references logged for 0, 500, 1000, 1500, and 2000 Hz. Since the transfer functions in these figures appear on a linear frequency scale, linear models were created using the raw pixel values without any preceding transformation. In the same manner as described above, the models were subsequently used to determine the frequency values (in Hz) that correspond to the transfer function peaks in each sub-figure.

In Feng and Castelli (1996), transfer functions were created for simulations of the nasalization of 11 French vowels; the figures for only three of these vowels ([i a u]) appear in the manuscript. An important contribution to the simulations in this study is the inclusion of a completely pharyngonasal tract, with a total closure in the oral cavity from the lowered velum (i.e., a [ŋ]-like target) for each of the vowels. This allowed the authors to create ten different transfer functions along a gradient scale from fully oral to fully nasal for each vowel. Moreover, the authors provide additional simulations of these three vowels, which incorporate the coupling of a sinus volume of 18 cm<sup>3</sup> given that, “from knowledge of the complexity of the nasal tract labyrinth, it seems unrealistic to neglect the sinus cavities (mainly the maxillaries)” (p.

3702). As such, the sinus-coupled figures were used here to derive formant frequency values, and the first transfer function (corresponding to fully oral) along with the fifth transfer function (corresponding to 50% along the gradient scale from fully oral to fully nasal) were chosen to represent the non-VP-coupled and VP-coupled vowels, respectively. These transfer functions appear in Fig. 11 from Feng and Castelli (1996). In the same manner as described for Maeda (1993),  $x$ -axis tick references were logged for 0, 500, 1000, 1500, and 2000 Hz. Linear models were created as above, and the resulting frequency values (in Hz) corresponding to the transfer function peaks in each sub-figure were logged.

In the most recent of the four studies, Serrurier and Badin (2008) used physiological data from midsagittal magnetic resonance (MR) images to calculate area functions of the vocal tract from the glottis to the lips. Because these images include movement of the velum within the midsagittal plane, their area functions include changes to the shape of the oral cavity from the lowering of the velum, like in Feng and Castelli (1996). Also, in a manner similar to Feng and Castelli (1996), Serrurier and Badin (2008) model ten transfer functions along a gradient scale from fully oral to fully nasal (i.e., a [ŋ]-like target). Although the transfer functions for only one vowel are provided ([a]; Fig. 22 of the manuscript), an additional figure (Fig. 23) displays the acoustic  $F1$ – $F2$  triangles formed by [i a u], juxtaposing the fully oral triangle with the fully nasal triangle. In comparing results for [a] in these two figures, it is clear that what the authors consider as  $F1$  and  $F2$  in the fully nasal representation of [a] actually arises from oral  $F1$  and nasal  $N1$ , respectively. With this understanding, estimates of nasalized versions of each of the three vowels can be obtained by averaging between their fully oral and fully nasal versions from Fig. 23. Such estimates were created for each of the three vowels in order to compare the model results from Serrurier and Badin (2008) with those from the other three studies. For  $F1$  estimates,  $y$ -axis tick references were logged for 200, 400, 600, and 800 Hz. For  $F2$  estimates,  $x$ -axis tick references were logged for 500, 1000, 1500, and 2000 Hz. The  $x$ - and  $y$ -dimension pixel values for the corners of the vowel triangle (corresponding to the vowels [i a u]) were logged for both the oral and nasal versions of the vowel space. Linear models were created as above for both  $F1$  and  $F2$  dimensions, and the resulting frequency values (in Hz) corresponding to these corners were logged.

## 2. Results of the derived formant measurements

The results from the process described above are shown both in Table I (tabular form) and in Fig. 2 (graphical form). The values displayed in Table I represent the change in formant frequency from the oral vowel to the nasalized vowel; increases are denoted by “+” and bold text, and decreases are denoted by “–” and italic text. For both of the plots in Fig. 2, the arrows indicate the direction of frequency shift from the oral vowel (beginning of the arrow) to the nasalized vowel (end of the arrow). The results are presented in two ways: oral formant shifts and what will be referred to as “first candidate selection.” The oral formant shifts represent

TABLE I. Formant frequency shifts measured from figures appearing in Fujimura and Lindqvist (1971), Maeda (1993), Feng and Castelli (1996), and Serrurier and Badin (2008). Oral formant shifts appear in the OF columns and first candidate selection shifts appear in the FC columns. Increases are displayed in bold text and decreases are displayed in italic text.

Study	Effect on:	/i/		/ɛ/		/ɜ/		/a~ɑ/		/o/		/ɔ/		/u/	
		OF	FC	OF	FC	OF	FC	OF	FC	OF	FC	OF	FC	OF	FC
F&L 1971	F1					<b>+54</b>	<i>-200</i>	<b>+82</b>	<i>-273</i>	<b>+145</b>	<i>-101</i>			<b>+175</b>	<b>+6</b>
	F2					<i>-219</i>	<i>-949</i>	<i>-40</i>	<i>-374</i>	0	<i>-234</i>			<b>+50</b>	<i>-237</i>
M 1993	F1	<b>+32</b>	<b>+32</b>	<b>+113</b>	<i>-48</i>					<i>-16</i>	<i>-16</i>	<b>+97</b>	<i>-73</i>	<b>+33</b>	<b>+33</b>
	F2	<b>+57</b>	<i>-1667</i>	<b>+16</b>	<i>-1309</i>					<b>+57</b>	<i>-234</i>	<b>+32</b>	<i>-407</i>	<i>-8</i>	<i>-236</i>
F&C 1996	F1	<b>+51</b>	<b>+51</b>					<i>+57</i>	<i>-235</i>					<b>+32</b>	<b>+32</b>
	F2	<i>-51</i>	<i>-1341</i>					0	<i>-477</i>					<b>+191</b>	<b>+19</b>
S&B 2008	F1	<b>+69</b>	<b>+69</b>					<i>-93</i>	<i>-93</i>					<b>+53</b>	<b>+53</b>
	F2	<i>-580</i>	<i>-580</i>					<i>-117</i>	<i>-117</i>					<b>+236</b>	<b>+236</b>

the modifications to the frequencies of oral  $F1$  and oral  $F2$  when VP coupling occurs, as predicted by the studies. However, for the majority of the transfer functions that appear in these studies—in fact, the only notable exceptions are /i o u/ from (Maeda, 1993)—the amplitude of the nasal pole  $N1$  is quite large, matching the amplitude of one or both of the oral formants, or even exceeding their amplitude when partial or complete oral pole cancellation from  $Z1$  occurs. Because of this, it is unreasonable to assume that formant estimation (be it automatic or manual) of spectra corresponding to these transfer functions would always be able to discern between  $N1$  and an oral pole. Therefore, the first candidate selection represents the predicted formant frequency shifts if each “first candidate” in the VP-coupled transfer function is selected as the corresponding formant. For example, if the pole frequency sequence is  $F1-N1-F2$ , the first two candidates ( $F1$  and  $N1$ ) are chosen as  $F1$  and  $F2$ , respectively. Since the  $F1-F2$  results for Serrurier and Badin (2008) include the possibility of  $N1$  being selected as a formant candidate, as described previously, the results for their study are simply duplicated in both plots.

With reference to the oral formant shifts, the results among the four studies are not entirely consistent. While there is a general trend towards  $F1$ -raising throughout the vowel space, the magnitude of this raising is not consistent across the studies. Moreover, Serrurier and Badin (2008) predict  $F1$ -

lowering for low vowels, while Fujimura and Lindqvist (1971) and Feng and Castelli (1996) predict  $F1$ -raising for the same. Predictions for  $F2$  shifts are even less straightforward, as most of the results indicate either no change or no substantial change in  $F2$ , while three items suggest  $F2$ -lowering (the neutral vowel from Fujimura and Lindqvist, 1971, and [i a] from Serrurier and Badin, 2008), but others suggest  $F2$ -raising ([u] from both Feng and Castelli, 1996; Serrurier and Badin, 2008). The overall trend of  $F1$ -raising throughout the vowel space is consistent with the claim from Fujimura and Lindqvist (1971) that “all formants of a nasalized vowel shift monotonically upwards” with increased velopharyngeal opening (p. 552), yet inconsistent with the prediction from Serrurier and Badin (2008) of  $F1$ -lowering for [a]. Furthermore, other researchers posit that  $F1$ -lowering should result from nasalization of low vowels when the degree of nasalization is sufficient to introduce a high-amplitude  $N1$  and/or cancellation of oral  $F1$  by  $Z1$  occurs (Diehl *et al.*, 1990; Stevens, 2000), an effect which has been observed in, e.g., the production of phonetically nasalized [ã] when tongue height is held constant (Carignan *et al.*, 2011).

With reference to first candidate formant selection, the results are more consistent among the three studies: there is a tendency for formants to shift towards the high-back corner of the vowel space. Of course, this is not at all surprising, given that first candidate selection will necessarily result in

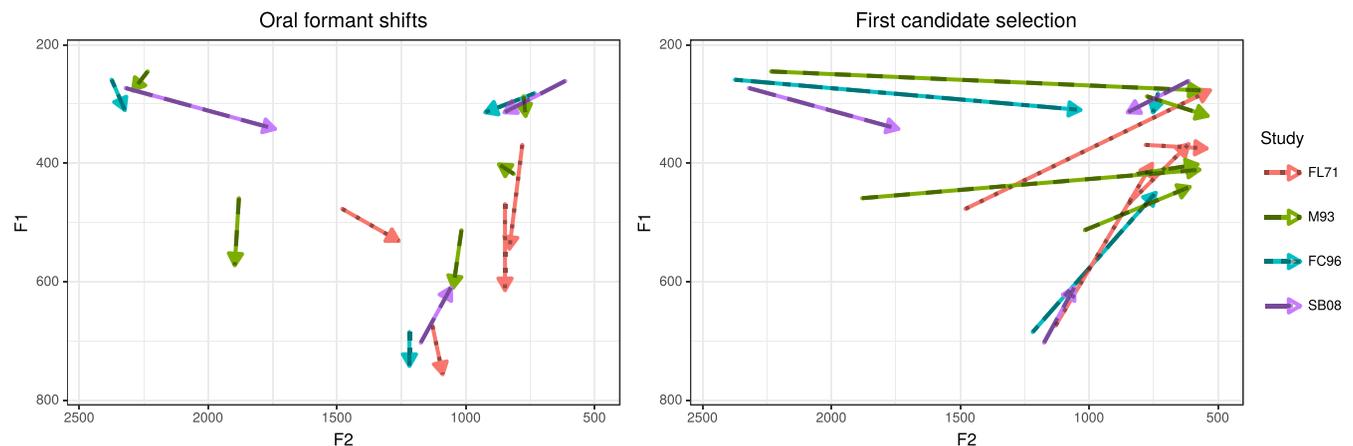


FIG. 2. (Color online) Formant frequency shifts measured from figures appearing in Fujimura and Lindqvist (1971), Maeda (1993), Feng and Castelli (1996), and Serrurier and Badin (2008).

the possibility of either  $F1$ - or  $F2$ -lowering (or both) for all vowels, since  $N1$  is never expected to have a higher frequency than oral  $F2$ , thus making it a candidate for formant selection in every vowel. In high vowels, however,  $F1$ -raising is still observed, since the upward-shifted oral  $F1$  is lower in frequency than  $N1$ . Nevertheless, the  $F1$ -lowering for low vowels predicted by [Serrurier and Badin \(2008\)](#), posited by other researchers, and observed in production data does, indeed, manifest in the results from the three studies containing low vowels when first candidate selection is enforced.

## B. Vowel nasalization and oral articulatory shifts

The difficulty in understanding and predicting formant frequency shifts in VP-coupled vowels arises not only from the somewhat inconsistent predictions from previous studies, but also from the fact that VP coupling itself is not the only possible source of formant frequency modulations in VP-coupled vowels. In addition to velum lowering, phonemic nasal vowels have been observed to be produced with modifications to the *shape of the oral tract* in Northern French ([Straka, 1965](#); [Brichler-Labaeye, 1970](#); [Zerling, 1984](#); [Bothorel et al., 1986](#); [Delvaux et al., 2002](#); [Demolin et al., 2003](#); [Delvaux, 2012](#); [Carignan, 2014a](#); [Carignan et al., 2015](#)), Southern French ([Carignan, 2017](#)), Laurentian French ([Carignan, 2013](#)), Brazilian Portuguese ([Barlaz et al., 2015](#); [da Matta Machado, 1993](#); [Shosted, 2015](#); [Shosted et al., 2015](#)), European Portuguese ([Martins et al., 2012](#); [Oliveira et al., 2012](#); [Teixeira et al., 2012](#)), and Hindi ([Shosted et al., 2012](#)). The degree of these modifications has also been shown to differ between male and female speakers in, e.g., French ([Engwall et al., 2006](#)). Differences in oral tract configuration are not unique to phonemic nasal vowels, but have also been observed in contextually nasalized vowels. In American English, for example, evidence has shown that speakers may raise the tongue body during the production of phonetically nasalized oral /i/ ([Carignan et al., 2011](#)), and/or lower the tongue body during the production of phonetically nasalized oral /a/ ([Arai, 2004](#)), perhaps as a way of actively compensating for  $F1$  shifts due to VP coupling and, thus, helping to prevent a phonemic oral-nasal split ([Carignan et al., 2011](#), pp. 678–679; [Arai, 2004](#), pp. 46–47).

These findings introduce a further level of complexity to the already intricate problem of determining the acoustic effects of VP coupling in vowels, since any changes to the oral cavity would be assumed to modify formant frequencies in ways that are independent of VP coupling. The fundamental problem posed in Sec. I thus remains: once the separate oral and nasal tract transfer functions (which have now been shown to vary in independent ways) have merged in the acoustic signal, how can we tease apart their respective spectral effects? One possible option is to take measurements from phonetically nasalized vowels in order to compare them to the same oral vowels in non-nasalized contexts (e.g., [Beddor, 1982](#); [Stevens, 2000](#), pp. 319–320). However, the findings from [Arai \(2004\)](#) and [Carignan et al. \(2011\)](#) for American English highlight the possibility of oral articulatory shifts even in a phonetically nasalized context, negating

the purpose of using contextual vowel nasalization to study the independent effects of VP coupling.

Another option is to devise a more contrived condition, wherein speakers lower their velum while attempting to maintain oral configuration, as in [Fujimura and Lindqvist \(1971\)](#). There are two primary issues from that study, however, which would likely need to be addressed in order to obtain a more complete understanding of formant frequency shifts arising from VP coupling. The first issue regards the generalizability of the results for the entire vowel quadrilateral. The transfer functions measured from the sweep-tone signals represent productions for only four vowels: one vowel from one speaker, three vowels from another. Thus, it is questionable whether the results can be generalized across the vowel quadrilateral more globally. The second issue regards the fact that, although the authors state that the oral and nasalized vowels were produced with “approximately the same tongue articulation,” there is simply no way to be certain of this since no kinematic lingual measurements were made during the experiment. Although the VP-coupled transfer functions were “matched with a correction function derived from the matchings of the non-nasalized vowels samples” ([Fujimura and Lindqvist, 1971](#), p. 553), this correction represents a modified version of the actual production data and, thus, arguably distances the results from the reality of the speech signal. A possible solution, therefore, is to build on the work of [Fujimura and Lindqvist \(1971\)](#) but to accept that these potential articulatory shifts may happen and *account for them* in some manner, thus controlling for their acoustic effects. The rest of the current manuscript is devoted to describing in detail a method that aims to address these concerns through the use of ultrasound and nasalance technologies, along with the results obtain from the method.

## II. METHODS

### A. Data collection

Native speakers of six different languages/dialects participated in the study (American English, Australian English, Mandarin, Cantonese, French, and Hungarian): four males and two females, with a mean age of 31.3 [standard deviation (SD) 7.5]. Speaker background information is given in Table III in the [Appendix](#). All speakers were either graduate students or professional academics in phonetics and/or phonology. The speakers were instructed to produce 20 sustained repetitions of each of the 11 vowels /i ɪ e ε æ a ɑ ɔ o ʊ u/; the repetitions were carried out in individual blocks for each vowel, proceeding in the order indicated above. The target vowel set included a wide range of qualities around the periphery of the vowel quadrilateral in order to induce a large amount of phonetic variation; only monophthongs were included since the task involved sustaining vowel productions. For each repetition, the speaker was instructed to sustain phonation of an oral quality of the vowel, then subsequently lower the velum during the sustained phonation while attempting to maintain tongue posture; International Phonetic Alphabet (IPA) symbol prompts of the alternations were displayed on a laptop screen in order to aid the speakers in the task, e.g., “[u] ~ [ũ].” Due to the contrived

and marked nature of the experimental task, the targeted realizations of the sounds are likely more physiological than linguistic. However, it is nevertheless possible that speaker- and language-specific differences could arise in the experiment. Thus, in order to control for these possible effects and help ensure that the results from the study are generalizable, it was desirable to include multiple speakers (instead of a single speaker) from multiple language backgrounds (instead of a homogeneous language background).

During the vowel productions, the experimenter (who was also one of the speakers) monitored tongue posture on a GE LOGIQ *e* portable ultrasound system. If the experimenter judged the tongue posture to have changed substantially, the item was repeated; this process continued until 20 repetitions of each vowel were obtained that each displayed minimal change in tongue posture.<sup>1</sup> Due to the somewhat difficult nature of the experiment, speakers were sometimes unable to achieve the task for a particular vowel or were unable to obtain 20 repetitions; in these cases, the speaker was instructed to advance to the next target vowel in the set.<sup>2</sup> During the sustained vowel productions, nasalance data and ultrasound data related to tongue posture were collected simultaneously. Nasalance data were captured using a Glottal Enterprises H-SEP-MU nasalance plate, which consists of two directional microphones located on either side of an acoustic baffle that surrounds the speaker's upper lip. Ultrasound images of midsagittal tongue shape were generated using a GE 8C-RS transducer held in place by a non-metallic elastic headset (Derrick *et al.*, 2015). An example of the experimental setup is shown in Fig. 3.

Before data collection began, the speaker was instructed to produce a range of vowels while the experimenter adjusted the position of the transducer underneath the speaker's jaw in order to ensure that the transducer was oriented to align with the midsagittal plane of the tongue. During data collection, ultrasound video was captured in real time from the GE LOGIQ *e* VGA video output using an Epiphan VGA2USB Pro video grabber. The nasalance audio and



FIG. 3. (Color online) An example of the experimental setup used in the study, including a hand-held Glottal Enterprises nasalance device and an ultrasound probe holder headset (Derrick *et al.*, 2015).

ultrasound video data were co-registered on a dedicated computer, using Fast Forward Moving Pictures Expert Group software (FFmpeg; FFmpeg Development Team, 2016) to record a continuous Audio Video Interleaved (.AVI) file at 30 fps with embedded audio sampled at a rate of 44.1 kHz. During preliminary investigations, it was found that frame rates higher than 30 fps consistently caused FFmpeg to drop frames, which would be problematic for synchronizing the nasalance and ultrasound data. As such, 30 fps was chosen for this study; given the nature of the elicited stimuli (sustained vowels), this frame rate was high enough to capture the necessary data.

Synchronization of the nasalance and ultrasound data was obtained by segmenting both data streams according to comparable acoustic and articulatory landmarks. At the beginning and end of each data collection session, the speaker was instructed to say “TA,TA,TA.” Since the sound sequence [ta] provides both transient acoustic information (a burst in the spectrogram and corresponding waveform) and transient articulatory information (a rapid transition of the tongue blade from alveolar contact to a low position within the mouth), it is relatively straightforward to identify corresponding acoustic and articulatory landmarks for this sequence. After the respective landmarks were located at the beginning and end of the recordings, the nasalance data and ultrasound frames between the two landmarks were extracted and, hence, synchronized.

## B. Data analysis

### 1. Nasalance data analysis

Analysis of the synchronized nasalance data was carried out using PRAAT (Boersma and Weenink, 2015). The sustained vowel productions were segmented manually according to the broadband spectrogram and corresponding waveform. Changes in voice quality (e.g., creaky voice at the end of vowel utterances<sup>3</sup>) as well as any occurrences of nasal codas were excluded from the sustained vowel segmentations. The average duration for the segmented vowels was 1.77 s (SD 0.57 s), ranging from a minimum average of 1.59 s (S03) to a maximum average of 2.33 s (S01). Separate amplitude tracks for the oral and nasal signals were created, and nasalance was derived by calculating the proportional nasal amplitude, i.e.,  $A_{nasal} / (A_{oral} + A_{nasal})$ . While this measurement controls for differences in overall amplitude (i.e., differences due to variation in sub-glottal pressure), it is nevertheless sensitive to vowel-specific differences due to variation in oral impedance (i.e., differences due to variation in intra-oral pressure). However, since nasalance is used in this study only to locate within-token time points, and not to make generalizations about the degree of nasalization across vowel qualities, this particular sensitivity of the measure is not a concern for the methodology presented here.

The time points associated with the minimum and maximum nasalance in each token were located automatically; these time points correspond to the most oral and most nasal parts of the token, and will be referred to throughout the manuscript as the “oral point” and “nasal point,” respectively. An example of one of speaker S01's utterances of /a/

is shown in Fig. 4. The respective waveforms of the two channels from the nasalance plate are plotted in the top panel, with the nasal channel shown in the top-most portion of the panel and the oral channel underneath it. The broadband spectrogram (0–3000 Hz) is plotted in the middle panel, with superimposed trajectories of  $F1$  and  $F2$  (estimated according to the methods described below in Sec. II B 2). The segmentation derived according to the protocol described above is denoted in the bottom panel; the oral point and nasal point for the utterance are shown in the bottom-most tier.

## 2. Acoustic data analysis

Formant estimation was performed on the combined audio from the stereo nasalance channels, created using the “Combine to mono” function in PRAAT. Two-formant estimation at the oral and nasal time points was carried out using the Burg linear predictive coding (LPC) method,<sup>4</sup> with optimized parameters for each speaker and vowel, derived from a semi-automated procedure similar to Escudero *et al.* (2009). First,  $F1$  and  $F2$  measurements were taken at the oral and nasal point of each token in an iterative manner: at each iteration in the range from 1000 to 4000 Hz, the  $F2$  ceiling was increased by 50 Hz and formant measurements were logged. Second, the set of formant measurements was imported into R and the cumulative variance for  $F1$  and  $F2$  was measured separately for the oral and nasal points at each 50 Hz ceiling step and for each vowel. Third, the variance for  $F1$  and  $F2$  was summed and plotted separately for the oral and nasal time points for each vowel; the plots were visually inspected and the ceiling which yielded the lowest variance for both the oral and nasal points of each vowel was logged (i.e., the ceiling parameter which produced the most consistent  $F1$  and  $F2$  measurements for the given vowel). Finally, the suitability of these optimized ceilings was verified manually in PRAAT by inspecting the resulting formant tracks against a broadband spectrogram. In some cases (especially for front vowels), using the same  $F2$  ceiling

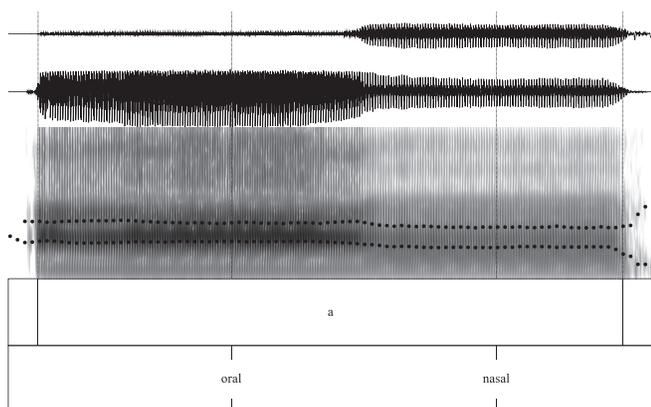


FIG. 4. Example of one of S01’s utterances of /a/. Top panel: nasalance stereo waveforms. Middle panel: broadband spectrogram with  $F1/F2$  trajectories superimposed. Bottom panel: segmentation of the utterance, including the point of minimum nasalance (the “oral” point) and maximum nasalance (the “nasal” point).

for both the oral and nasal points yielded inaccurate formant tracking for one of the two time points. In these cases, the two  $F2$  ceilings that were the closest in frequency but which yielded accurate formant estimation for both time points were chosen for the final analysis. After each of the optimized  $F2$  ceilings was verified for each vowel and for each speaker, the final logged parameters were used to measure  $F1$  and  $F2$  at the oral and nasal points for each token. The  $F2$  ceilings used in the final analysis are given in Table IV in the Appendix.

## 3. Ultrasound data analysis

The indices of all of the ultrasound frames located between the oral and nasal points of the vowel tokens in each recording were logged, and these ultrasound images were subsequently filtered and processed separately for each speaker in MATLAB (The Mathworks Inc., 2015) using Temporally Resolved Articulatory Configuration Tracking of Ultrasound software (TRACTUS; Carignan, 2014b): the images were downsized via bicubic interpolation to 20% of their original resolution, a region of interest (RoI) around the bounds of the movement of the tongue surface was applied, and the down-sampled pixels in the RoI were used as dimensions in principal components analysis (PCA) modeling (additional information about treatment of the ultrasound frames in the software can be found in Mielke *et al.*, 2017). Since different speakers may realize the lingual articulation of a set of speech sounds with idiosyncratic degrees of freedom, the PCA modeling was carried out on a by-speaker basis, rather than applying a single PCA model to all of the speaker data. The exclusion of ultrasound images that were not located between the oral and nasal points of the vowel tokens ensured that the PCA model only accounted for image variance corresponding to tongue postures that are associated with the productions of the 11 target vowels. Principal components (PCs) that independently explained at least 1% of the total image variance were retained, yielding between 12 and 15 total PCs for each speaker. Due to the orthogonal nature of the components, the PC scores are able to be used as independent variables in regression models, as explained in the following section.

## 4. Articulatory-to-acoustic mapping

In order to map lingual articulation to the acoustics, two separate regression models (for  $F1$  and  $F2$ ) were created for each speaker. Each model included formant values as the dependent variable and the ultrasound PC scores related to the oral time point<sup>5</sup> of each token as independent variables. Linear models were tested along with second-, third-, and fourth-order polynomial models. The goodness of fit for each model was assessed using adjusted  $R^2$  values. In every case, a linear model provided the best fit, yielding between 0.75  $R^2$  and 0.97  $R^2$  across all models. These linear models were subsequently used to predict formant values for the corresponding nasal point of each token, using the ultrasound PC scores from the nasal acoustic time points as predictor variables. The result represents formant values (in Hz) that are predicted by tongue posture alone, without any acoustic

Oral time point



Predicted F1: 486 Hz, Measured F1: 476 Hz, Difference: -10 Hz  
 Predicted F2: 901 Hz, Measured F2: 832 Hz, Difference: -69 Hz

Nasal time point



Predicted F1: 486 Hz, Measured F1: 277 Hz, Difference: -209 Hz  
 Predicted F2: 901 Hz, Measured F2: 749 Hz, Difference: -152 Hz

FIG. 5. Composite heat maps generated by a linear combination of the ultrasound PC scores and coefficients for the oral and nasal time points of one of S04's utterances of /ɔ/. In these images, the tongue root is to the left, and the tongue tip is to the right. The predicted and measured formant values, as well as their differences, are provided for the respective time points.

influence of nasalization. Observable differences between predicted and measured formant values at the nasal point of each token can, therefore, be considered to be due to the effect of VP coupling and not to any changes in tongue configuration. In order to combine the vowel spaces of the different speakers, formant values were normalized for each speaker via *z*-score transformation (i.e., Lobanov transformation) before translation back to Hz using the grand mean and average SD, in order to preserve the within-speaker normalized structure while retaining interpretability of the results.

### 5. Oral model error correction

An example of the results obtained from this method are provided in Fig. 5. The heat maps shown in this figure were created via a linear combination of the 13 PC scores and coefficients for a single production of /ɔ/ by speaker S04. Thus, they are a spatial representation of the PC scores for this token that were used as independent variables in the regression model for this speaker. In each heat map, the tongue root is to the left and the tongue tip is to the right, revealing that the speaker produced this vowel with a lowered blade and bunched dorsum. There are no discernible spatial differences between the two images, revealing that the speaker successfully maintained tongue posture while lowering the velum during this utterance. However, the differences between the predicted and measured formant values for the oral point highlight that the linear model does not perfectly predict formant values. In other words, while relatively large formant differences are observed for the nasal time point, smaller differences are nonetheless also observed for the oral time point. Thus, it is possible that a portion of the difference between predicted and measured formant values at the nasal time point might be due to model error or to formant frequency modifications that arise from non-lingual oral articulation (e.g., labial configuration).

To control for these possible sources of error, formant predictions and measurements were also made at the oral

time point of each token in order to obtain baselines of error for the oral models. Figure 6 displays the formant values taken from the oral time points, grouped by target vowel category. The opaque colored dots are formant values predicted by tongue shape, while the transparent colored triangles are formant values that were actually measured. The arrows connect the means of the predicted values (start of the arrow) and measured values (end of the arrow). In other words, the arrows display the direction and magnitude of average acoustic change from the formant values predicted by tongue shape to the actual measured formant values for each of the 11 vowel categories. For the majority of the vowel space, there are no differences between the predicted and observed formant values, indicating that the linear model predictions are accurate. However, there are small but non-negligible differences in the high-front and high-back corners of the vowel space. It is likely that these differences are due to error in the linear mapping (i.e., the extreme corners of the vowel space do

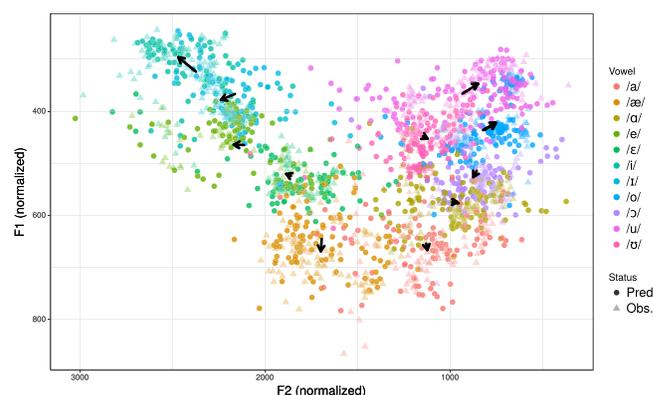


FIG. 6. (Color online) Acoustic vowel space of speaker-normalized formant values, measured at the oral point of each token. Opaque colored dots represent formant values predicted by lingual ultrasound images; transparent triangles represent actual measured values. Arrows connect the means of the predicted and measured categories.

not match well with the overall linear trends) and/or to difference in labial configuration (e.g., lip spreading for /i ɪ e/ and lip rounding for /u o/). Therefore, in addition to the “raw” observed-predicted differences, the final analysis will also include a corrected version of these measurements: the oral model errors were averaged for each vowel category, and these error baselines were subtracted from the measured formant values at the nasal point of the corresponding vowel tokens. In this way, each data point was corrected for vowel-specific errors in the linear mapping, yielding a substantially more conservative estimate of acoustic differences observed for the nasal time point.

### III. RESULTS

Figure 7 displays the uncorrected formant values taken from the nasal time points and Fig. 8 displays the corresponding corrected measurements. The opaque colored dots are formant values predicted by tongue shape, while the transparent colored triangles are formant values that were actually measured. The arrows connect the means of the predicted values (start of the arrow) and measured values (end of the arrow). The overall pattern suggests that vowels with  $F2$  values less than  $\approx 2000$  Hz all undergo both  $F1$ - and  $F2$ -lowering to various degrees. The global effect for this part of the vowel space is similar to a counter-clockwise chain shift: low vowels raise and retract in the vowel space, encroaching on the acoustic space of the mid-back vowels, which also raise and retract, encroaching on the acoustic space of the high-back vowels, which also raise and retract.

Figure 9 contains a more detailed visualization of the acoustic change for each vowel category, displaying the individual distributions for oral-corrected  $F1$  (left plot) and oral-corrected  $F2$  (right plot) of each vowel in the form of violin plots and corresponding box plots. The violin plots display the respective data distribution in the form of mirrored, vertically oriented probability density; densities have been scaled across the separate distributions to aid visualization. For each distribution, box plots have been superimposed on the corresponding violin plot. For any given box plot, the

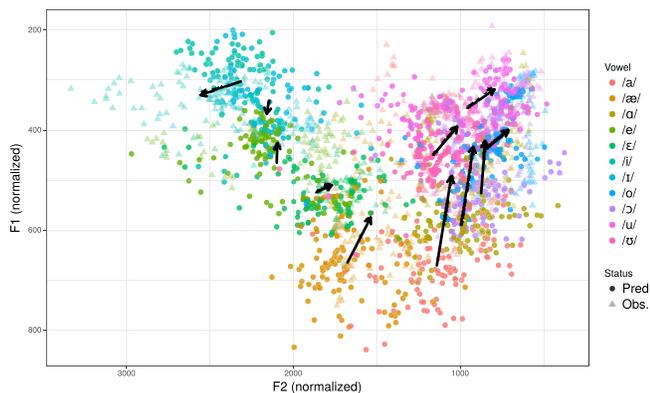


FIG. 7. (Color online) Acoustic vowel space of speaker-normalized formant values, measured at the nasal point of each token, before correction for oral model error. Opaque colored dots represent formant values predicted by lingual ultrasound images; transparent triangles represent actual measured values. Arrows connect the means of the predicted and measured categories.

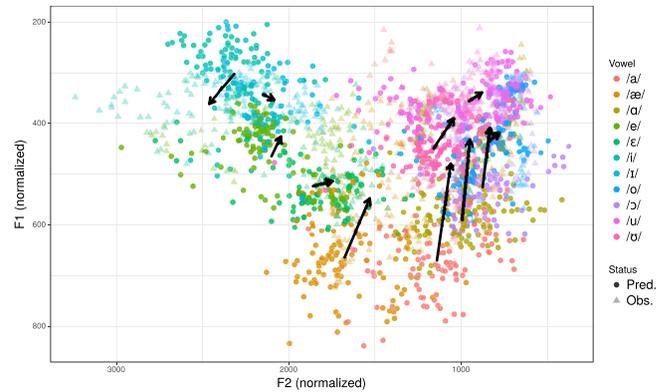


FIG. 8. (Color online) Acoustic vowel space of speaker-normalized formant values, measured at the nasal point of each token, after correction for oral model error. Opaque colored dots represent formant values predicted by lingual ultrasound images; transparent triangles represent actual measured values. Arrows connect the means of the predicted and measured categories.

horizontal line is the median of the distribution, the center notch pointing inward displays the 95% confidence interval around the median, the box displays the interquartile range (i.e., the middle 50% of the data), and the vertical lines extending from the box demarcate  $1.5 \times$  the interquartile range. Outliers have been removed to aid visualization. Box plots that display “flipped” notches represent distributions where the 95% confidence interval around the median extends beyond the interquartile range. In comparing any two box plots, the respective distributions can be assumed to differ significantly at  $\alpha = 0.05$  if their notches do not overlap along the y-axis. For both the  $F1$  and  $F2$  plots in the figure, the opaque colored violin and box plots (left plots in each category pair) represent distributions of formant values predicted by tongue shape, while the transparent colored violin and box plots (right plots in each category pair) represent distributions of formant values there were actually measured. The vowel categories in each plot have been arranged from left to right in ascending value of the median frequency of the formant values predicted by tongue shape. Welch Two Sample  $t$ -tests were performed on the oral-corrected data from each vowel category; the results are provided in Table II, with cells highlighted in gray for tests that reached significance at  $\alpha = 0.05$ .

Globally, the most striking result is that the general  $F1$ -raising throughout the vowel space predicted by Fujimura and Lindqvist (1971), Maeda (1993), and Feng and Castelli (1996) is not evidenced in these data. In fact,  $F1$ -raising is only observed for the high-front vowels /i ɪ/ (although only /i/ reaches significance in the  $t$ -tests), while the high-back vowel /u/ undergoes slight lowering (16 Hz), rather than raising. On the other hand,  $F1$ -lowering can be observed for all vowels other than /i ɪ/, although the effects for /ɛ o/ are not significant. For some vowels (e.g., the low vowels /a/ and /ɑ/), the effect is relatively large: a difference  $>150$  Hz. With regard to  $F2$ , lowering can be observed for all vowels except for /i/, for which  $F2$  is raised. Although the  $F2$ -lowering effect is consistent across 10 of the 11 vowels, the effect does not reach significance in the  $t$ -tests for four of these ten vowels (/a e ɪ ɔ/).

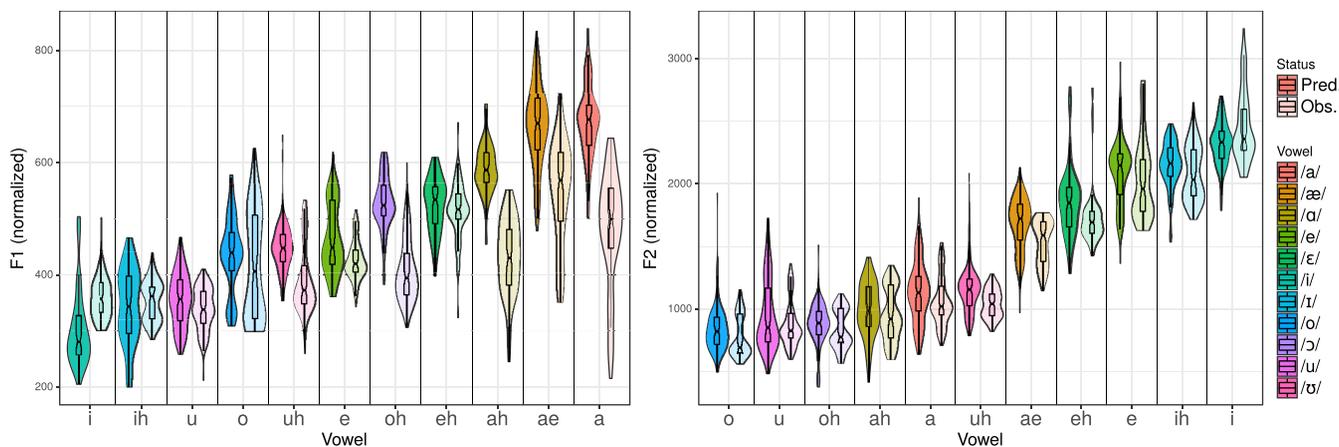


FIG. 9. (Color online) Violin and box plots of speaker-normalized formant values, measured at the nasal point of each token, after correction for oral model error. Opaque colors represent formant values predicted by lingual ultrasound images; transparent colors represent actual measured values. Predicted and measured values for  $F1$  are shown in the left plot and for  $F2$  in the right plot. Categories (denoted by color) are the original target vowel categories.

## IV. DISCUSSION

### A. Summary of results and comparison with predictions

The results for the formant frequency shifts observed in the current study suggest that  $F1$ -lowering, and to a lesser extent  $F2$ -lowering, are general properties of the effect of VP coupling on the acoustic quadrilateral, with a caveat:  $F1$ - and  $F2$ -raising are both observed, but only for /i/. The  $F1$ -raising observed for /i/ did not reach significance, and  $F2$ -raising is not observed for any vowels other than /i/. The observation of  $F1$ -raising for the high vowel /i/ is consistent with previous modeling literature, since all of the studies summarized in Sec. IA 2 predicted  $F1$ -raising for this vowel. However, the other high vowel included in those studies (i.e., /u/) was also predicted to exhibit  $F1$ -raising due to nasalization. The results from the current study are not consistent with this prediction, since  $F1$  of /u/ was observed here to be slightly, yet significantly lower under the influence of velum lowering. The more pervasive  $F1$ -related consequence in the current study is the lowering of  $F1$  for VP-coupled non-high vowels. This result is inconsistent with three of the four previous studies, yet is consistent with the only study to predict formant frequency shifts directly from physiological speech data—i.e., Serrurier and Badin (2008), who predicted  $F1$ -lowering for the nasalization of /a/. However, if  $N1$  is taken into account for these studies as a

possible candidate for formant selection, then all four of the studies display evidence of  $F1$ -lowering for non-high vowels. Moreover, in the current study, the  $F1$  range for /i/ (the only vowel to exhibit  $F1$ -raising) is in the frequency range of 250–400 Hz posited by Maeda (1993) as the location for  $N1$  (see Sec. IA), which suggests that the  $F1$ -raising observed for this vowel may be due to  $N1$  selection. However, the  $F1$  ranges for /i u/ are also within this frequency range, and these vowels do not exhibit  $F1$ -raising. Nonetheless, the results from this study suggest that both oral pole shifts and the possibility of nasal pole selection (two scenarios described in Sec. IA) may affect the measurement of  $F1$  frequency in a real nasal vowel spectrum.

The finding of a relatively minor, but consistent lowering of  $F2$  throughout the majority of the vowel space has generally not been considered in the previous literature. While most of the predicted transfer functions for VP-coupled vowels indicate a relatively stable  $F2$  frequency,  $F2$ -lowering was predicted for /i a/ by Serrurier and Badin (2008) and for the “neutral vowel” by Fujimura and Lindqvist (1971); although  $F2$ -lowering was not observed for [i] in the current study,  $F2$ -lowering for the rest of the vowel space was observed. Once again, if  $N1$  is taken into account as a possible candidate for formant selection, then the majority the vowels in the previous studies discussed in Sec. IA 2 display  $F2$ -lowering. However, the magnitudes of the shifts in that case are quite large, which was not observed

TABLE II. Welch Two Sample  $t$ -test results for oral-corrected values measured at nasal time points. Cells highlighted in gray represent models that were significant at  $\alpha = 0.05$ .

Formant	Status	/a/	/æ/	/ɑ/	/e/	/ɛ/	/i/	/ɪ/	/o/	/ɔ/	/u/	/ʊ/
F1	Predicted	670	664	588	465	524	303	343	436	526	355	447
	Observed	480	548	432	426	514	361	353	418	409	339	390
	Difference	-190	-116	-156	-40	-10	+58	+10	-18	-117	-16	-57
F2	Predicted	1140	1676	996	2101	1860	2317	2148	845	875	954	1157
	Observed	1060	1530	950	2046	1745	2461	2085	780	834	879	1042
	Difference	-80	-146	-46	-55	-115	+144	-63	-65	-41	-75	-115

in the results from the current study. This suggests that the possibility of nasal pole selection is likely not the cause of the  $F2$ -lowering observed here; otherwise, the shifts in  $F2$  frequency are expected to be much larger than has been observed in the current study. In Hindi, Shosted *et al.* (2012) observed  $F2$ -lowering in all non-front nasal vowels compared to their oral congeners, which in many cases could not be accounted for by tongue configuration. These results from Hindi are strikingly similar to the results observed here for  $F2$  shifts in non-front vowels. Shosted *et al.* (2012) posit that the  $F2$ -lowering in Hindi may be a result of the lowering of the velum toward the tongue dorsum, which may also explain the similar result observed here. Moreover, Delvaux (2009) has shown that  $F2$ -lowering alone is sufficient to trigger the percept of nasality on synthesized vowels in French, and Beddor (1993) suggests that the increased  $F1$ - $F2$  proximity of non-front nasal vowels observed for Hindi, Turkish, Igbo, and English (Beddor, 1982) should result in perceptual retraction compared to their oral counterparts—she notes, however, that this retraction is not well supported in the perceptual vowel spaces of Wright (1986). These results suggest that the  $F2$ -lowering effect observed in the current study—argued here to be a physiological consequence of nasalization—may very well play a role in the production of vowel nasality, even if its role in the perception of vowel nasality is as yet undetermined. Although much of the previous literature on vowel nasalization has focused on the acoustic effects in the spectral region of  $F1$ , the results observed here suggest that more attention should be paid in future research to the effect on  $F2$ , as well.

Finally, the global pattern of the VP-coupled vowel space compared to the non-VP-coupled vowel space resembles a counter-clockwise acoustic chain shift: low vowels retract and raise, mid-low back vowels also retract and raise, and mid-high back vowels also retract and raise. This pattern is consistent with counter-clockwise chain shifts of nasal compared to oral realizations that have occurred naturally in, e.g., Northern Metropolitan French (Fónagy, 1989; Hansen, 2001; Malderez, 1991; Carignan, 2014a). Beddor *et al.* (1986) and Krakow *et al.* (1988) observed that the  $F1$  variation due to nasalization can be attributed by listeners to changes in tongue height: an increase in  $F1$  for high vowels may be attributed to either a lower tongue position or an increase in degree of nasalization, and a decrease in  $F1$  for low vowels may be attributed to either a higher tongue position or an increase in degree of nasalization. Similarly, Wright (1975, 1986) found that listeners perceived nasalized [ĩ] as lower and more retracted than oral [i] and nasalized [ã] as higher than oral [a]. Combined with the finding from Delvaux (2009) that  $F2$ -lowering (most typically associated with retraction of the tongue body; Stevens, 2000, pp. 276–277) can trigger percept of nasality (at least in French), these results suggest that the  $F1$ / $F2$  shifts due to nasalization may be perceived by listeners as changes in lingual configuration that emulate a counter-clockwise chain shift. Although diachronic chain shifts in natural languages are the result of vowel quality displacements that occur in a sequential manner over time, it is possible that the psycho-acoustic pressure on the vowel space that arises due to VP coupling

may act as a catalyst for the initial vowel category shift and/or may influence the direction of subsequent shifts in the vowel sub-system. However, while cross-linguistic evidence of nasal vowel category shifts (summarized in Beddor, 1993) suggests that vowel height centralization is typologically common, vowel retraction is not generally prevalent. Nevertheless, given their similarity to the perceptual effects discussed above, the results observed in the current study pose an interesting set of questions for future research on vowel nasality and sound change.

## B. Limitations of the proposed method

By virtue of using ultrasound video of tongue posture to account for changes in oral tract configuration, the method that has been proposed here is, admittedly, entirely “linguo-centric.” Although much of the previous research on articulatory co-variation in vowel nasality has similarly focused on lingual configuration, nasal and nasalized vowels have also been shown to be produced with differences in lip posture (Zerling, 1984; Bothorel *et al.*, 1986; Montagu, 2002; Delvaux *et al.*, 2002; Delvaux, 2012; Carignan, 2013, 2014a), lower pharyngeal aperture (Carignan *et al.*, 2015), and even voice quality (Garellek *et al.*, 2016; Carignan, 2017), in comparison with their oral vowel counterparts. Given that ultrasound captures only an image of the tongue (thus neglecting variation in labial configuration and voice quality), as well as the fact that this image is often restricted to the polar field between the mandible and the hyoid bone (often obscuring the tongue root in the lower pharynx), the articulatory modifications listed above cannot be accounted for in the proposed method. Nevertheless, the measurements made using this method involve token-wise correction for oral model error (see Sec. IIB 5), helping to mitigate possible formant frequency modulation due to non-lingual adjustments. Moreover, it is reasonable to assume that the articulatory dimension that is most likely to be affected (if any) in the controlled experimental conditions described here (conscious, active lowering of the velum while sustaining a vowel production) is lingual configuration, due to proprioceptive stimulation of the velum lowering toward the tongue dorsum and the intrinsic muscular connection between the soft palate and the tongue via the *palatoglossus* muscle (Kuehn and Azzam, 1978; Zemlin, 1998).

Alternatively, magnetic resonance imaging (MRI) video of the entire vocal tract could be used in place of both ultrasound and nasalance in the proposed method since changes in tongue posture, lip configuration, and pharyngeal aperture could be accounted for, and since the degree of nasalization could be determined directly from velic height. However, due to the acoustic reality of MRI scanning (i.e., high amplitude noise in the recording environment), the complexity of the interaction between the oral and nasal transfer functions would likely be exacerbated, and the (often small) magnitudes of acoustic change observed in the results presented here would likely be masked. Even with noise cancellation techniques in place, MRI scanning is arguably not the most appropriate tool for capturing the fine-grained acoustic

effects observed in the current study, which may be the reason why acoustic modeling was used in place of acoustic measurements in [Serrurier and Badin \(2008\)](#).

## V. CONCLUSION

The experimental method that has been presented in this manuscript is an attempt to address a well-known issue in research on the acoustics of vowel nasalization: given that the separate transfer functions related to the oral and nasal cavities are merged in the acoustic signal, how can we observe the independent effect of velopharyngeal coupling on the spectrum without the influence of changes to the shape of the oral cavity? By using ultrasound and nasalance to isolate the effect of lingual configuration on formant frequencies of 11 nasalized vowels produced by six speakers of different language backgrounds, the current study reveals that the independent effect of nasalization on the acoustic vowel quadrilateral is  $F1$ - and  $F2$ -raising of /i/,  $F1$ -lowering of non-high vowels, and  $F2$ -lowering of non-front vowels. The cumulative effect of these formant frequency modifications generally resembles a counter-clockwise chain shift of nasal compared to non-nasal vowels. These results are consistent with modeling literature with regard to  $F1$ -raising of /i/, but generally inconsistent with regard to  $F1$ -lowering of non-high vowels and  $F2$ -lowering of non-front vowels. Although it is helpful for researchers to understand the predicted effects of VP coupling in a theoretical system, it is arguably equally helpful for researchers to understand the effects of VP coupling that they are likely to encounter in real production data. The results of the study presented here help to enrich this knowledge of vowel nasalization, as well as challenge traditional modeling assumptions for VP-coupled vowels.

## ACKNOWLEDGMENTS

The author is grateful to Ryan K. Shosted for his comments and suggestions on an earlier version of this manuscript, as well as to the two anonymous reviewers, whose thoughtful examination has greatly improved the article. The author would also like to thank staff and students at Macquarie University, Sydney, Australia, where preliminary results were presented and additional data were recorded.

## APPENDIX: ADDITIONAL TABLES

TABLE III. Speaker background information.

Speaker	Sex	Age	Native language	Time in Australia
S01	Male	35	American English	2.5 years
S02	Female	27	Hungarian	2 years
S03	Male	27	Mandarin	2 years
S04	Male	29	French	1.5 years
S05	Female	25	Australian English	25 years
S06	Male	45	Cantonese	7 years

TABLE IV. Ceiling parameter values used for speaker-normalized formant estimation in PRAAT.

Vowel	Nasality	$F2$ ceiling (by speaker)					
		S01	S02	S03	S04	S05	S06
/i/	oral	3200	3500	2750	2700	3350	2800
	nasal	3500	4000	3600	3200	3350	2800
/ɪ/	oral	2800	3200	*	2800	3400	2800
	nasal	2800	3700	*	2800	3400	3000
/e/	oral	2800	3100	3100	2700	2800	2600
	nasal	2800	4000	3100	2700	3000	2600
/ɛ/	oral	2650	3050	*	2400	2650	2300
	nasal	2650	3700	*	2400	2650	2300
/æ/	oral	2650	3000	2400	2300	2650	2200
	nasal	2650	3000	2400	2300	2650	2200
/a/	oral	1800	2600	2000	2000	2450	1900
	nasal	1800	2600	2000	2000	2450	1900
/ɑ/	oral	1600	2200	2000	1600	1600	1900
	nasal	1850	2200	2000	1600	1600	1900
/ɔ/	oral	1400	2000	1800	1400	1600	1900
	nasal	1400	2000	1800	1400	1600	1900
/o/	oral	1200	1600	1500	1000	1300	1800
	nasal	1200	1600	1500	1000	1300	1800
/ʊ/	oral	1750	1600	1400	1600	2300	1850
	nasal	1750	1600	1400	1600	2300	1850
/u/	oral	1650	1000	1000	1100	2300	1800
	nasal	1650	1000	1000	1100	2300	1800

<sup>1</sup>It is important to note that the method described in this manuscript does not necessarily require the tongue posture to remain constant throughout the vowel phonation, since the resulting metric accounts for variation in formant values that is due to tongue shape. However, maintaining the tongue posture helps ensure that the ultrasound image variance that is used to predict formant values in the final analysis falls within the range of image variance that is used to map the articulation to the acoustics (see Sec. II B 4).

<sup>2</sup>Between the cases of speakers being unable to achieve the task for certain vowels (in particular, speaker S03, who was unable to produce /ɪ/ and /ɛ/ while controlling velum movement) and the cases of speakers being unable to produce 20 repetitions of a vowel, 1012 of the intended 1320 target items (77%) were ultimately included for analysis.

<sup>3</sup>As noted by a reviewer, the presence of creaky voice at the end of some utterances suggests that larynx lowering might have sometimes occurred at the end of the vowel productions (i.e., during the nasalized portion). Larynx lowering would yield lower pitch and, crucially, a depression of all formants for reasons unrelated to nasalization (i.e., lengthening the vocal tract). In order to test this possibility, a separate analysis was carried out on  $F0$  measurements taken in PRAAT at the oral and nasal time points of each token. These measurements were speaker-normalized before translation back to Hz, in the same way as for the formant values. A Welch Two Sample  $t$ -test revealed a slight, yet significant difference between average oral  $F0$  (137 Hz) and average nasal  $F0$  (139 Hz). However, given the negligible magnitude of the difference (2 Hz), it is unlikely that larynx height has any bearing on observed formant frequency shifts in the data.

<sup>4</sup>The scenario exemplified in Fig. 1(a) of Sec. IA merits particular comment at this juncture. When a nasal pole (e.g.,  $N1$ ) and an oral pole (e.g.,  $F1$ ) are near to each other in frequency, the additional spectral energy from  $N1$  not only widens the bandwidth of  $F1$ , but also shifts the center of gravity around  $F1$ , either raising the center of gravity (when  $N1 > F1$ ) or lowering the center of gravity (when  $N1 < F1$ ). This shift in  $F1$  center of gravity can, naturally, also influence the measurement of  $F1$ : since “formants are [...] simply mountain ranges formed of many harmonic hills” (Styler, 2015, p. 28), any changes to the spectral structure have the capacity to affect formant measurement. Thus, an all-pole formant estimation technique such as LPC may potentially be influenced by the presence of  $N1$ , even if the  $F1$  associated with the oral transfer function alone remains stable.

- <sup>5</sup>Before inclusion of the PC scores in the regression models, due to the difference in sampling rates between the audio and ultrasound data, the ultrasound data were up-sampled via linear interpolation between the PC scores for the ultrasound frames immediately preceding and immediately following the oral acoustic time points. The same interpolation was carried out between the PC scores for the ultrasound frames immediately preceding and immediately following the nasal acoustic time point prior to their inclusion in the predictor models.
- Arai, T. (2004). "Comparing tongue positions of vowels in oral and nasal contexts," in *Proceedings of Interspeech 2005*, September 4–8, Lisbon, Portugal, pp. 33–49.
- Barlaz, M., Fu, M., Dubin, J., Shosted, R., Liang, Z.-P., and Sutton, B. (2015). "Lingual differences in Brazilian Portuguese: An rt-MRI study," in *Proceedings of the 18th Meeting of the International Congress of the Phonetic Sciences*, August 10–14, Glasgow, Scotland, Paper No. 0819.1-5 retrieved from <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0819.pdf>.
- Beddor, P. S. (1982). "Phonological and phonetic effects of nasalization on vowel height," Ph.D. thesis, University of Minnesota, Minneapolis, MN.
- Beddor, P. S. (1993). "The perception of nasal vowels," in *Nasals, Nasalization, and the Velum*, edited by M. K. Huffman and R. A. Krakow (Academic Press, New York), pp. 171–196.
- Beddor, P. S., Krakow, R. A., and Goldstein, L. M. (1986). "Perceptual constraints and phonological change: A study of nasal vowel height," *Phonol. Yearbook* 3, 197–217.
- Boersma, P., and Weenink, D. (2015). "'PRAAT: Doing phonetics by computer' [computer software program]," <http://www.praat.org/> (Last viewed October 15, 2015).
- Bothorel, A., Simon, P., Wioland, F., and Zerling, J.-P. (1986). *Cinéradiographie des Voyelles et Consonnes du Français (Cineradiography of French Vowels and Consonants)*, Travaux de l'Institut de Phonétique de Strasbourg Vol. 18 (Institut de Phonétique de Strasbourg, Strasbourg).
- Brichler-Labaeye, C. (1970). *Les Voyelles Françaises. Mouvements et Positions Articulatoires à la Lumière de la Radiocinématographie (coll. Bibliothèque Française et Romane, Série A, No. 18) [French Vowels. Articulatory Movements and Positions in the Light of Radiocinematography (coll. French and Romance Collection, Series A, No. 18)]* (Klincksieck, Paris).
- Carignan, C. (2013). "When nasal is more than nasal: The oral articulation of nasal vowels in two dialects of French," Ph.D. thesis, University of Illinois at Urbana-Champaign, Champaign, IL.
- Carignan, C. (2014a). "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.
- Carignan, C. (2014b). "TRACTUS (Temporally Resolved Articulatory Configuration Tracking of UltraSound) software suite [computer software program]," <http://christophercarignan.github.io/TRACTUS> (Last viewed August 17, 2015).
- Carignan, C. (2017). "Covariation of nasalization, tongue height, and breathiness in the realization of F1 of Southern French nasal vowels," *J. Phon.* 63, 87–105.
- Carignan, C., Shosted, R., Fu, M., Liang, Z.-P., and Sutton, B. (2015). "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.
- Carignan, C., Shosted, R., Shih, C., and Rong, P. (2011). "Compensatory articulation in American English nasalized vowels," *J. Phon.* 39, 668–682.
- Chen, M. Y. (1997). "Acoustic correlates of English and French nasalized vowels," *J. Acoust. Soc. Am.* 102, 2360–2370.
- da Matta Machado, M. (1993). "Fenômenos de nasalização vocálica em português: Estudo cine-radiográfico" ("Vowel nasalization phenomena in Portuguese: A cine-radiographic study"), *Cadernos de estudos lingüísticos* 25, 113–127.
- Delvaux, V. (2009). "Perception du contraste de nasalité vocalique en français" ("Perception of the vowel nasality contrast in French"), *French Lang. Studies* 19, 25–29.
- Delvaux, V. (2012). *Les Voyelles Nasales du Français: Aérodynamique, Articulation, Acoustique et Perception (French Nasal Vowels: Aerodynamics, Articulation, Acoustics and Perception)* (Peter Lang, Bruxelles).
- Delvaux, V., Metens, T., and Soquet, A. (2002). "French nasal vowels: Acoustic and articulatory properties," in *Proceedings of the 7th International Conference on Spoken Language Processing*, September 16–20, Denver, CO, Vol. 1, pp. 53–56.
- Demolin, D., Delvaux, V., Metens, T., and Soquet, A. (2003). "Determination of velum opening for French nasal vowels by magnetic resonance imaging," *J. Voice* 17(4), 454–467.
- Derrick, D., Best, C. T., and Fiasson, R. (2015). "Non-metallic ultrasound probe holder for co-collection and co-registration with EMA," in *Proceedings of 18th International Congress of Phonetic Sciences (ICPhS)*, August 10–14, Glasgow, Scotland, pp. 1–5.
- Diehl, R. L., Kluender, K. R., and Walsh, M. A. (1990). "Some auditory bases of speech perception and production," in *Advances in Speech, Hearing and Language Processing*, edited by W. A. Ainsworth (JAI Press, London), pp. 243–268.
- Engwall, O., Delvaux, V., and Metens, T. (2006). "Interspeaker variation in the articulation of nasal vowels," <http://www.speech.kth.se/prod/publications/files/1926.pdf> (Last viewed April 2, 2017).
- Escudero, P., Boersma, P., Rauber, A. S., and Bion, R. A. H. (2009). "A cross-dialect acoustic description of vowels: Brazilian and European Portuguese," *J. Acoust. Soc. Am.* 126(3), 1379–1393.
- Feng, G., and Castelli, E. (1996). "Some acoustic features of nasal and nasalized vowels: A target for vowel nasalization," *J. Acoust. Soc. Am.* 99(6), 3694–3706.
- FFmpeg Development Team (2016). *FFmpeg* [computer software program], <https://ffmpeg.org> (Last viewed November 5, 2016).
- Fónagy, I. (1989). "Le français change de visage?" ("Is French getting a face-lift?"), *Revue Romane* 24(2), 225–254.
- Fujimura, O., and Lindqvist, J. (1971). "Sweep-tone measurements of vocal-tract characteristics," *J. Acoust. Soc. Am.* 49, 541–558.
- Garellek, M., Ritchart, A., and Kuang, J. (2016). "Breathy voice during nasality: A cross-linguistic study," *J. Phon.* 59, 110–121.
- Hansen, A. B. (2001). "Lexical diffusion as a factor of phonetic change: The case of modern French nasal vowels," *Lang. Var. Change* 13(2), 209–252.
- Krakow, R. A., Beddor, P. S., and Goldstein, L. M. (1988). "Coarticulatory influences on the perceived height of nasal vowels," *J. Acoust. Soc. Am.* 83(3), 1146–1158.
- Kuehn, D. P., and Azzam, N. A. (1978). "Anatomical characteristics of palatoglossus and the anterior faucial pillar," *Cleft Palate J.* 15(4), 349–359.
- Maeda, S. (1982a). "Acoustic correlates of vowel nasalization: A simulation study," *J. Acoust. Soc. Am.* 72, S102.
- Maeda, S. (1982b). "The role of the sinus cavities in the production of nasal vowels," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, May 3–5, Paris, France, pp. 911–914.
- Maeda, S. (1993). "Acoustics of vowel nasalization and articulatory shifts in French nasal vowels," in *Nasals, Nasalization, and the Velum*, edited by M. K. Huffman and R. A. Krakow (Academic Press, New York), pp. 147–170.
- Malderez, I. (1991). "Tendance de neutralisation des oppositions entre voyelles nasales dans la parole des jeunes gens d'Ile-de-France" ("Tendency toward neutralization of nasal vowel contrasts in the speech of Ile-de-France youth"), in *Actes du XIIème Congrès International des Sciences Phonétiques*, August 19–24, Aix-en-Provence, France, Vol. 2 (Université de Provence, France), pp. 174–177.
- Martins, P., Oliveira, C., Silva, S., and Teixeira, A. (2012). "Velar movement in European Portuguese nasal vowels," in *Proceedings of IberSpeech—VII Jornadas en Tecnologia del Habla and II Iberian SLTech Workshop*, November 21–23, Madrid, Spain, pp. 231–240.
- Mielke, J., Carignan, C., and Thomas, E. R. (2017). "The articulatory dynamics of pre-velar and pre-nasal /æ/-raising in English: An ultrasound study," *J. Acoust. Soc. Am.* 142(1), 332–349.
- Montagu, J. (2002). "L'articulation labiale des voyelles nasales postérieures du français: Comparaison entre locuteurs français et anglo-américains" ("Labial articulation of posterior nasal vowels in French: Comparison between French and Anglo-American speakers"), in *XXIVèmes Journées d'Étude sur la Parole*, Nancy, France, pp. 253–256.
- Oliveira, C., Martins, P., Silva, S., and Teixeira, A. (2012). "An MRI study of the oral articulation of European Portuguese nasal vowels," in *Proceedings of INTERSPEECH 2012*, September 9–13, Portland, OR, pp. 2690–2693.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria [computer software program], <http://www.R-project.org> (Last viewed January 28, 2017).
- 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(4), 2335–2355.
- Shosted, R. (2015). "Nasal vowels are not [+nasal] oral vowels," in *Proceedings of the 42nd Linguistic Symposium on Romance Languages*

- 2012, edited by J. Smith and T. Ihsane (John Benjamins, Amsterdam), pp. 63–76.
- Shosted, R., Carignan, C., and Rong, P. (2012). “Managing the distinctiveness of phonemic nasal vowels: Articulatory evidence from Hindi,” *J. Acoust. Soc. Am.* **131**(1), 455–465.
- Shosted, R., Meneses, F., Pozzani, D., Wong, N., Hermes, Z., and Loucks, T. (2015). “Articulatory characterization of oral/nasal vowel contrast in Brazilian Portuguese (São Paulo state): Evidence from EMA,” in *Proceedings of the 5th EICEFALA—International Meeting on Speech Sciences*, September 28–30, Belo Horizonte, Brazil.
- Stevens, K. N. (2000). *Acoustic Phonetics* (MIT Press, Cambridge, MA).
- Straka, G. (1965). *Album Phonétique* (L’Université de Laval, Québec, Canada).
- Styler, W. (2015). “On the acoustical and perceptual features of vowel nasality,” Ph.D. thesis, University of Colorado, Denver, CO.
- Styler, W. (2017). “On the acoustical features of vowel nasality in English and French,” *J. Acoust. Soc. Am.* **142**(4), 2469–2482.
- Teixeira, A., Martins, P., Oliveira, C., Ferreira, C., Silva, A., and Shosted, R. (2012). “Real-time MRI for Portuguese,” in *Computational Processing of the Portuguese Language* (Springer, New York), pp. 306–317.
- The GIMP Development Team (2017). *GNU Image Manipulation Program 2.9.7* [computer software program], <https://www.gimp.org> (Last viewed November 15, 2017).
- The Mathworks Inc. (2015). “MATLAB 2015a” Natick.
- Wright, J. T. (1975). “Effects of vowel nasalization on the perception of vowel height,” in *Nasálfest: Papers From a Symposium on Nasals and Nasalization*, edited by C. A. Ferguson, L. M. Hyman, and J. J. Ohala (Stanford University Language Universals Project, Palo Alto, CA), pp. 373–388.
- Wright, J. T. (1986). “The behavior of nasalized vowels in perceptual vowel space,” in *Experimental Phonology*, edited by J. J. Ohala and J. J. Jaeger (Academic Press, New York), pp. 45–67.
- Zemlin, W. R. (1998). *Speech and Hearing Science: Anatomy and Physiology*, 4th ed. (Allyn and Bacon, Boston, MA).
- Zerling, J. P. (1984). “Phénomènes de Nasalité et de Nasalization Vocaliques: Étude Cinéradiographique pour Deux Locuteurs” (“Phenomena of Vowel nasality and nasalization: A cineradiographic study of two speakers”), *Travaux de l’Institut de Phonétique de Strasbourg* **16**, 241–266.