THE EFFECT OF VELOPHARYNGEAL APERTURE ON ACOUSTIC MEASURES

Marissa Barlaz, Ryan Shosted, Brad Sutton

University of Illinois at Urbana-Champaign
{goldrch, rshosted, bsutton}@illinois.edu

ABSTRACT

Understanding the relationship between articulation and acoustics for nasal sounds requires the quantification of oral-nasal coupling, which is not trivial to directly observe. In this study we utilize ultra-fast MRI to directly image velopharyngeal opening in order to calculate the degree of coupling between the oral and nasal cavities. We utilize recent denoising algorithms in order to utilize the acoustic signal recorded simultaneously with MR images. Generalized additive models are used to model the non-linear, time-dynamic relationship between velopharyngeal aperture and acoustic measures. Results show a direct relationship between degree of oral-nasal coupling and first formant bandwidth, frequency, and amplitude, as well as $A1-P0$. Applications of non-linear models are promising for understanding the effects of articulatory changes on acoustic output.

Keywords: velopharyngeal opening, nasality, Brazilian Portuguese, MRI

1. INTRODUCTION

The production of nasal sounds occurs when the velopharyngeal port opens due to velar lowering. This effectively couples the oropharnygeal and nasal cavities, which increases the area of the sound filter. Overall sound pressure radiating from the vocal tract is reduced, and formant bandwidths tend to widen [16]. These effects are clearly seen on the first formant (around $200-800\ Hz$, depending on vowel quality and speaker), and generally result in F1 centralization. That is, F1 lowers for low nasal vowels and raises for high nasal vowels, with respect to their oral congeners [5, 6, 8, 15]. The effect of nasalization on F1 is determined based on the degree of velopharyneal opening [5], in addition to oropharyngeal articulatory differences [2].

Quantification of oral-nasal coupling is a difficult task, due to the many degrees of freedom in nasal sound production. Acoustic measures such as $A1-P0$ (the difference between the first formant and the first nasal formant's amplitudes), F1 bandwidth, and spectral tilt have been shown to most accurately distinguish oral and nasal vowel congeners [17]. However, the use of these measures is not trivial, as it is often difficult to tease apart oral and nasal formants, or to detect nasal anti-formants. Recent advances in ultra-fast magnetic resonance imaging (MRI) technology allow for non-invasive imaging of velopharyngeal opening [7], permitting direct measurements of oral-nasal coupling [4].

The current study applies MRI on velopharyngeal opening to a question of phonemic versus phonetic nasalization. The phonemic inventory of Brazilian Portuguese (BP) includes five phonemic oral-nasal vowel pairs /i∼ĩ e∼ẽ a∼ã o∼õ u∼ũ/ and five phonemic nasal vowels [1], as well as phonetically nasalized vowels (defined as an oral vowel that undergoes coarticulatory nasalization due to the presence of a heterosyllabic adjacent nasal consonant). Members of each oral/nasal vowel pair have been shown to manifest different oropharyngeal articulations. Comparatively, nasalized vowels show more variation in articulatory configuration, compared to their nasal counterparts [2].

The current study also applies recently-developed noise-cancelling technology to the acoustics recorded in the MR scanner, which are inherently noisy and therefore difficult to use in spectral analysis. The use of the simultaneously-acquired acoustics allows direct comparison between articulatory configuration and acoustic output.

The aim of this study is to analyze the temporal aspects of oral-nasal coupling and their effect on acoustic measures. This will give further insight into how the phonological difference between nasal and nasalized vowels is manifested in phonetic outputs. Furthermore, the use of generalized additive mixed models (GAMMs) allows for understanding time-dynamic differences in velopharyngeal opening, and how these differences affect acoustic outputs, particularly the properties of the first formant.

2. METHODS

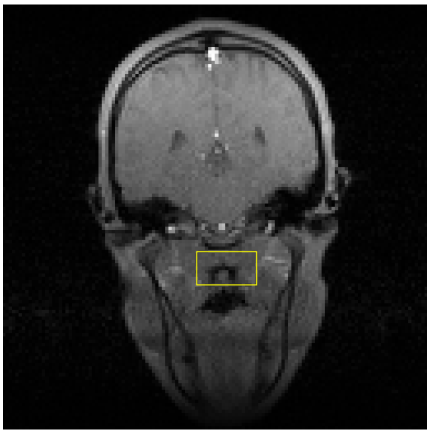
Data was collected from 8 male speakers and 5 female speakers of BP who were born and raised in the states of Minas Gerais and São Paulo in southeastern Brazil. The data used here are part of a larger study on nasalization [2]. Due to timing constraints, only oral, nasal, and nasalized variants of /a, i, u/ were recorded. Target vowels were in the second, stressed syllable of a trisyllabic word, surrounded by consonants in the phonetic environment [LABIAL]___[ALVEOLAR]. Target words were placed in the carrier phrase *digo X duas vezes* "I said X two times." Phrases were presented in a randomized order in the 3 T Siemens Trio MRI scanner at the Beckman Institute for Advanced Science and Technology, at the University of Illinois at Urbana-Champaign. Participants lay in supine position in the scanner and repeated each phrase at a normal rate, speaking until the noise of the scanner ceased.(Note that [11] find no significant differences in velar size, shape, or setting between supine and upright position.) Each recording was approximately 90 seconds long. The acoustic signal was recorded using a MR-compatible headset with an attached optical microphone worn by participants (Dual Channel-FOMRI, Optoacoustics, Or Yehuda, Israel). Due to different speaking rates, an unequal number of tokens was collected for each speaker and lexical item (range 23–48).

Images were reconstructed using the Partial Separability model [12, 22] in Matlab 2012a [10]. A single image taken in an oblique orientation, which captured the velopharyngeal opening, was used for this study. Temporal resolution is 25 frames per second. The image resolution is 128 × 128 voxels, and the resolution of each voxel is 2.2 mm × 2.2 mm × 6.5 mm (through-plane depth).

MR images were converted to black and white using the *im2bw* function in MATLAB [10], at a pixel intensity threshold of 0.2. A region of interest (ROI) was selected around the velopharyngeal opening. Figure 1 shows an example MR image and its respective ROI. The number of black pixels was divided by the total number of pixels in the ROI, to give an open proportion (OP) within the ROI. Range of OP is 0–1, where 0 is totally closed (i.e., the ROI is full of tissue) and 1 is totally open (i.e., no tissue in the ROI). OP was extracted for each image within the duration of the target vowels, based on manual segmentation of simultaneously-recorded acoustic files in Praat [3]. No difference was observed in the range of OP due to different speaking rates.

The acoustic recordings were post-processed through a noise-reducing algorithm, which utilized dictionary learning and wavelet analysis techniques for audio enhancement [19]. Acoustic values including $A1 - P0$ (the difference in amplitude between the F1 and the first nasal peak), F1 frequency, bandwidth, and amplitude were extracted from the enhanced acoustic recordings, as these measures have been shown to be highly sensitive to changes due to nasalization. Ten points were taken in time-normalized intervals from the vowel's duration in Praat [3] using a series of scripts [17]. For a full acoustic/articulatory model, F1 was downsampled to match the number of data points in the OP analysis.

Figure 1: Oblique slice taken through the velopharyngeal opening, with the ROI highlighted.



The time-dynamic trajectories of $A1 - P0$, F1 frequency, bandwidth, and amplitude, and OP were plotted using smoothing spline ANOVA (SSANOVA), implemented in the *gss* package [9] in R [14] and compared across vowel and nasality conditions. GAMMs were used to determine the effect of velopharyngeal OP as a predictor of change in F1. Separate GAMMs were made for F1 frequency, bandwidth, and amplitude. Vowel quality and nasality were predictor variables, included as tensor smooth interactions with OP and normalized time. Speaker ID was included as a random factor smooth. The GAMMs were implemented using the *mgcv* R package [21], and visualized using the *itsadug* package [18].

3. RESULTS

Inspection of SSANOVA plots reveals expected results. F1 frequency tended to be centralized for nasal vowels. Specifically, F1 of /ã/ was significantly lower than that of /a/, and F1 of /ĩ/ and /ũ/ were higher than F1 of /i/ and /u/, respectively. In regards to the nasalized vowels, F1 trajectories tended toward an intermediate position between the oral and nasal vowels. Nasalized /a/ and /u/ tended to be more similar to their phonemic nasal counterparts, whereas nasalized /i/ was more similar to its oral counterpart. This is in line with reports of nasalization in BP based on high-fidelity recordings [2].

F1 bandwidth of the nasal vowels was overall higher than that of oral vowels, which is an expected result of nasalization. In general, the nasalized vowels showed bandwidth values similar to those of nasal vowels, though there was considerable individual variation in the patterning of nasalized vowels.

F1 amplitude of the nasal vowels was overall lower than that of the oral vowels. This is another expected result of nasalization. Once again, the nasalized vowels tended to show similar amplitudes to the nasal vowels. The most robust difference was for the vowel category /a/, which showed differences of up to 50% of the amplitude range, between the oral and nasal vowels, with the nasalized vowels showing similar ranges to the nasal vowels.

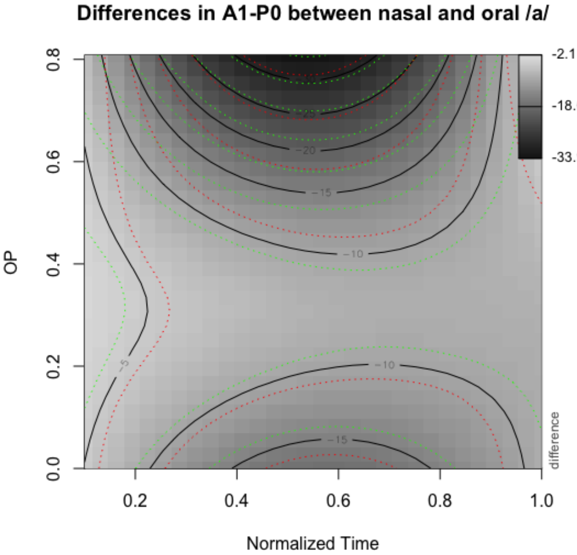
OP plots show that the oral vowels had the smallest amount of velopharyngeal opening across vowel categories. Unsurprisingly, nasal vowels showed the highest amount of velopharyngeal opening. Nasalized vowels showed intermediate amounts of velopharyngeal opening, though they tended to show similar OP values as nasal vowels, especially in the second half of their normalized durations.

Results of the GAMMs used to analyze differences in acoustic measures based on differences in vowel quality, nasality, and velopharyngeal opening are found in Table 1. For all GAMMs, there were main and interaction effects of vowel and nasality condition on the acoustic measures. For the tensor interactions between time and OP, effective degrees of freedom (EDF) were generally higher than 1, indicating a non-linear relationship between the variables and thus validating the use of GAMM for this analysis (see [20] for an extended discussion).

Visualizations of effects showed large differences for the acoustic measures. In regards to $A1-P0$, differences of up to 10 dB were observed between nasal and oral vowels across time. These differences were largest in magnitude for the vowel /a/. Differences between nasal and nasalized vowels were much lower in magnitude, and only reach a maximum difference of 5 dB (again, for the vowel /a/). For the nasalized-nasal difference, all three vowels showed significant differences after the first 25% of the vowels' durations, but not in the first quarter of their duration. Visualization of the dynamic effects of OP revealed that as OP increases, the distinctions between the nasality conditions became much greater, indicated by darker gray shading in contour plots. Figure 2 illustrates the relationship between time and OP on $A1-P0$, between nasal and oral /a/.

In regards to F1 frequency, differences mirror those seen in the SSANOVA. Differences between oral and nasal vowels are robust across time, especially for /a/, which shows average differences of 150 Hz between oral and nasal counterparts. This difference is seen across the entire vowel duration. The difference between nasal and nasalized /a/ is much smaller than that of the nasalized and oral /a/. For /i/ and /u/, the differences between nasality conditions are much smaller, with nasal vowels showing slightly higher F1 frequencies, and nasalized vowels patterning more similarly to oral vowels for /i/, and similar to nasal vowels for /u/. As OP increases, distinctions between the nasality conditions increases.

Figure 2: Relationship between oral and nasal vowels for $A1-P0$, across the range of OP and time.



F1 bandwidth was generally much higher for all vowels in the nasal condition, compared to the oral condition. Nasalized vowels tended to be more similar to nasal vowels. Increases in OP in the model led to much wider bandwidths, of up to 1000 Hz. For the majority of speakers, differences between oral and nasal vowel formant bandwidths were between $300-400$ Hz, and the difference between nasal and nasalized vowels was approximately 100 Hz.

F1 amplitude was generally much higher for oral vowels than for nasal and nasalized vowels. For some speakers, as OP increased, amplitude measures decreased across vowel and nasality qualities. For all speakers, variance in predicted values tended to increase (i.e., confidence intervals increased in size), indicating greater variability in amplitude measure prediction for higher OP values.

4. DISCUSSION

The results of this study show systematic differences between nasal and oral vowels. In general, F1 am-

	$A1-P0$	F1 Frequency	F1 Bandwidth	F1 Amplitude
(Intercept)	$-1.19(1.61)$	$556.68(27.87)^{***}$	$425.36(21.21)^{***}$	$24.19(2.36)^{***}$
/i/	$-6.94(0.23)^{***}$	$-29.67(8.28)^{***}$	$160.04(10.04)^{***}$	$-8.84(0.37)^{***}$
/u/	$-5.02(0.24)^{***}$	$-32.86(8.30)^{***}$	$52.83(10.20)^{***}$	$-6.64(0.38)^{***}$
nasalized	$2.24(0.22)^{***}$	$22.90(7.84)^{**}$	$-20.53(9.59)^{*}$	$2.30(0.35)^{***}$
oral	$7.49(0.22)^{***}$	$104.15(7.65)^{***}$	$-128.35(9.41)^{***}$	$6.56(0.35)^{***}$
/i/:nasalized	$-0.15(0.32)$	$-76.62(11.26)^{***}$	$60.65(13.64)^{***}$	$1.27(0.50)^{*}$
/u/:nasalized	$-0.70(0.32)^{*}$	$-24.28(11.33)^{*}$	$-7.77(13.88)$	$4.18(0.51)^{***}$
/i/:oral	$-4.48(0.32)^{***}$	$-199.65(11.23)^{***}$	$85.52(13.76)^{***}$	$1.57(0.51)^{**}$
/u/:oral	$-4.92(0.33)^{***}$	$-144.82(11.30)^{***}$	$105.90(14.02)^{***}$	$-0.57(0.52)$
EDF: (Time,OP): nasal	$4.46(4.50)^{***}$	$3.89(4.29)^{***}$	$2.50(2.50)^{**}$	$7.22(7.47)^{***}$
EDF: (Time,OP): nasalized	$4.20(4.45)^{***}$	$7.26(7.47)^{***}$	$2.50(2.50)^{*}$	$4.46(4.50)^{***}$
EDF: (Time,OP): oral	$7.47(7.50)^{***}$	$3.82(4.15)^{**}$	$7.35(7.49)^{***}$	$6.63(7.23)^{***}$
EDF: (Time,OP): /a/	$6.57(7.20)^{***}$	$3.47(3.72)^{*}$	$6.40(6.85)^{***}$	$4.50(4.50)^{***}$
EDF: (Time,OP): /i/	$2.50(2.50)$	$6.33(6.86)^{***}$	$4.23(4.36)^{**}$	$4.50(4.50)^{***}$
EDF: (Time,OP): /u/	$7.25(7.47)^{***}$	$4.05(4.28)^{***}$	$5.95(6.43)^{***}$	$3.83(4.64)$
EDF: s(RepNo)	$0.00(1.00)$	$0.80(1.00)^{*}$	$0.85(1.00)^{**}$	$0.49(1.00)$
EDF: s(Speaker)	$5.84(12.00)^{***}$	$5.93(12.00)^{***}$	$5.73(12.00)^{***}$	$5.97(12.00)^{***}$
EDF: s(Speaker,Time)	$100.89(108.00)^{**}$	$59.76(108.00)^{*}$	$81.40(108.00)^{***}$	$99.75(108.00)^{**}$
Deviance	977577.73	1220336440.86	1821205843.50	2434215.50
Deviance explained	0.42	0.19	0.25	0.54
Dispersion	50.72	63175.23	94387.05	126.29
R^2	0.41	0.18	0.24	0.54
Num. obs.	19421	19421	19421	19421
Num. smooth terms	9	9	9	9

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 1: Results of GAMMs with $A1-P0$, F1 frequency, bandwidth, and amplitude values as dependent variables, and nasality, vowel quality, velopharyngeal opening, and time as predictor variables. Effective degrees of freedom (EDF) are given for each smooth term, with EDF above 1 indicating a nonlinear relationship between variables.

plitude and $A1-P0$ is smaller for nasal vowels compared to oral vowels, whereas formant bandwidth is higher for nasal vowels compared to oral counterparts. F1 frequency of nasal vowels shows patterns of centralization of the vowel space—F1 raises for high nasal vowels and lowers for low nasal vowels, compared to their oral counterparts. These effects are in line with those expected of nasalization. Nasalized vowels tended to be more similar to oral vowels for /i/, and similar to nasal vowels for /a/ and /u/. Formant frequency data is in line with previous research on BP nasalization [2].

The study of acoustic parameters beyond formant frequency allows a deeper understanding of the direct impact of oral-nasal coupling on acoustics. Formant frequencies, especially F1, are modulated by oropharyngeal articulation in addition to velopharyngeal opening. Previous work shows that BP speakers modulate oropharyngeal position to enhance the acoustic effects of velopharyngeal coupling in nasal vowels. For example, speakers lower the tongue in production of /ũ/ and /ĩ/ and raise the tongue for /ã/, in comparison to their oral counterparts. These lingual movements have the effect of centralization on F1 frequency. It is therefore difficult to tease apart the effects on F1 that are due to velopharyngeal positioning from those effects of oropharyngeal modulation.

Analysis of formant bandwidth and amplitude, which are related to velopharyngeal coupling, allows for a direct comparison between articulation and acoustics. Results show formant bandwidth and OP are positively correlated, and as OP increases, there is more variability in bandwidth ranges. In addition, amplitude of nasal vowels is lower than that of oral vowels. The results show vowel-specific patterns for the relationship of nasalized vowels to their oral and nasal counterparts. $A1-P0$ is much higher for oral vowels than nasal vowels across the vowel qualities, and its magnitude increases as OP increases. In addition, the results regarding the measure $A1-P0$ confirm that this measure can be used as an acoustic parameter related to the degree of physiological nasalization.

The successful denoising of the acoustics recorded in the MR scanner will also allow for direct comparisons between acoustics and articulatory configuration. While the analysis of denoised acoustics produced some results with high variability, particularly for formant bandwidth, continued improvement of these signal processing methods will better facilitate direct comparisons of acoustic and image-based physiological signals. This will help researchers make progress towards solving the many-to-one problem of phonetics [13], and will open the doors for many further research questions to be explored.

5. REFERENCES

[1] Barbosa, P. A., Albano, E. C. 2004. Brazilian Portuguese. *Journal of the International Phonetic Association* 34(02), 227–232.

[2] Barlaz, M., Shosted, R., Fu, M., Sutton, B. 2018. Oropharygneal Articulation of Phonemic and Phonetic Nasalization in Brazilian Portuguese. *Journal of Phonetics* 71, 81 – 97.

[3] Boersma, P., Weenink, D. 2012. Praat, a system for doing phonetics by computer. http://www.praat.org/.

[4] Carignan, C., Shosted, R., Fu, M., Liang, Z.-P., 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. *Journal of Phonetics* 50(34–51).

[5] Diehl, R. L., Kluender, K. R., Walsh, M. A., Parker, E. M. 1991. Auditory enhancement in speech perception and phonology. *Cognition and the symbolic processes: Applied and ecological perspectives* 59–76.

[6] Feng, G., Castelli, E. 1996. Some acoustic features of nasal and nasalized vowels: A target for vowel nasalization. *The Journal of the Acoustical Society of America* 99(6), 3694–3706.

[7] Fu, M., Zhao, B., Carignan, C., Shosted, R. K., Perry, J. L., Kuehn, D. P., Liang, Z.-P., Sutton, B. P. 2015. High-resolution dynamic speech imaging with joint low-rank and sparsity constraints. *Magnetic Resonance in Medicine* 73(5), 1522–2594.

[8] Fujimura, O., Lindqvist, J. 1971. Sweep-tone measurements of vocal-tract characteristics. *The Journal of the Acoustical Society of America* 49(2B), 541–558.

[9] Gu, C. 2007. gss: General Smoothing Splines. *http://cran.r-project.org/web/packages/gss/gss.pdf* R package version (2007): 1-0.

[10] Inc., T. M. 2012. Version 7.14.0.039 (r2012a). http://www.mathworks.com/.

[11] Kollara, L., Perry, J. L. 2014. Effects of gravity on the velopharyngeal structures in children using upright magnetic resonance imaging. *The Cleft Palate-Craniofacial Journal* 51(6), 669–676.

[12] Liang, Z.-P. 2007. Spatiotemporal imaging with partially separable functions. *Noninvasive Functional Source Imaging of the Brain and Heart and the International Conference on Functional Biomedical Imaging* 181–182.

[13] Maeda, S. 1990. Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. In: *Speech production and speech modelling*. Springer 131–149.

[14] R Development Core Team, 2007. R: A language and environment for statistical computing. *http://www.R-project.org*.

[15] Serrurier, A., Badin, P. 2008. A three-dimensional articulatory model of the velum and nasopharyngeal wall based on MRI and CT data. *The Journal of the Acoustical Society of America* 123(4), 2335–2355.

[16] Stevens, K. 2000. *Acoustic Phonetics*. Cambridge: MIT Press.

[17] Styler, W. 2017. On the acoustical features of vowel nasality in English and French. *The Journal of the Acoustical Society of America* 142(4), 2469.

[18] van Rij, J., Wieling, M., Baayen, R. H., van Rijn, H. 2017. itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs.

[19] Vaz, C., Ramanarayanan, V., Narayanan, S. S. Aug. 2013. A two-step technique for MRI audio enhancement using dictionary learning and wavelet packet analysis. *Proceedings of InterSpeech*.

[20] Wieling, M. 2018. Analyzing dynamic phonetic data using generalized additive mixed modeling: a tutorial focusing on articulatory differences between l1 and l2 speakers of english. *Journal of Phonetics*.

[21] Wood, S., Wood, M. S. 2017. mgcv. *https://cran.r-project.org/web/packages/mgcv/mgcv.pdf* R Package version (2017) 1.8-22.

[22] Zhao, B., Haldar, J., Christodoulou, A., Liang, Z.-P. Sept 2012. Image reconstruction from highly undersampled (**k**, t)-space data with joint partial separability and sparsity constraints. *Medical Imaging, IEEE Transactions on* 31(9), 1809–1820.