The Emergence of Nasal Velar Codas in Brazilian Portuguese: An rt-MRI Study

Marissa Barlaz[1], *Maojing Fu*[2,4], *Zhi-Pei Liang*[2,3], *Ryan Shosted*[1,4], *Brad Sutton*[3,4]

[1]Department of Linguistics,[2]Department of Electrical and Computer Engineering,
[3]Department of Bioengineering, [4]Beckman Institute, University of Illinois at Urbana-Champaign
{goldrch2, mfu2, z-liang, rshosted, bsutton}@illinois.edu

Abstract

Real-time MRI has become an increasingly useful tool in articulatory studies, especially in examining posterior regions of the vocal tract. Previous work on Brazilian Portugese has shown the emergence of coda consonants following word-final nasal vowels, though this has been limited to the discussion of front vowels and coda emergence in anterior regions of the vocal tract. We present rt-MRI evidence showing a narrow constriction between the tongue dorsum and velum emerging at the end of back nasal vowels, patterning similarly with velar stop consonants. Though our present subject does not show evidence of complete occlusion, we believe the small distance between the tongue dorsum and velum makes the emergence of a velar consonant highly likely. The emergence of these consonants is a result of sound change that further distinguishes oral and nasal vowel pairs. We also show the utility of advanced image processing methodologies to give a more accurate and computationally economical way to do speech articulation research.

Index Terms: articulatory phonetics, rt-MRI, Brazilian Portuguese, nasal vowels, epiphenomenal contact

1. Introduction

Articulatory phonetic investigations are routinely concerned with place of articulation (POA), a concept associated with constriction or occlusion of the vocal tract in a specific region or at a specific point. Accurate identification of POA is still a considerable challenge, requiring speaker-customized instrumentation such as electropalatography (EPG) or advanced image processing, in the cases of ultrasound and real time magnetic resonance imaging (rt-MRI). Limitations of these methods are well known and it is often necessary to use a variety of techniques to best characterize the articulation of a speech sound. It is particularly difficult to use images to conclusively identify a full occlusion in the vocal tract.

In this study, we consider how best to identify constrictions and complete occlusions using rt-MRI in nasal vowels of Brazilian Portuguese (BP). BP has an inventory containing five nasal vowels: /ĩ/, /ẽ/, /ɐ̃/, /õ/, and /ũ/ [1, 2]. Various researchers have claimed that these vowels terminate in a nasal consonant [3, 4, 5], whose POA is dependent on the quality of the preceding vowel. The emergence of the nasal coda has implications for understanding sound change in Romance [6] and more generally, for the perception of oral/nasal distinctions among vowels [7, 8, 9]. Because reports suggest the tendency to produce an occluded nasal vowel is variable in POA and frequency, it is important to develop a technique for reliably identifying complete or incomplete occlusion across many repetitions, using an algorithmic approach. We show that it is possible to use rt-MRI to support claims suggesting the vowels /ũ/ and /õ/ frequently experience velar constriction approaching full closure. EPG evidence [10] has demonstrated that the complete occlusion after /ĩ/ and /ẽ/ is post-palatal, but more posterior occlusions (suggested by aerodynamic work [11]) cannot be registered with EPG.

rt-MRI has been used to successfully demonstrate articulatory differences in the posterior regions of the vocal tract in French [12, 13], showing its utility in regions that are otherwise difficult to visualize and analyze. Here, we use rt-MRI to image a more posterior region of the tract, though this does not provide absolute certainty regarding full occlusion. Working within the constraints of rt-MRI, we propose an automatic method for measuring distance between two soft-tissue structures using an application of a demon's algorithm to compare images by means of deformations. (Unfortunately, this excludes the alveopalatal region, at least in our scans, where the relatively thin mucous membrane layer covering the palatine bone cannot be registered in the scanning sequence).

2. Methods

2.1. Data

The subject is a female, native BP speaker from Fortaleza, Brazil, in her mid-20s. The data used here constitute part of a larger project on vocal tract imaging. A word list consisting of BP lexical items was used, with the target vowels in word final position and preceded by obstruents. Test items include: *ba**fum*** [bafũ] 'bad smell', *tup**ã*** [tupɐ̃] 'Tupi-Guarani divinity', *ca**pim*** [capĩ] 'species of grass', *da**bom*** [dabõ] 'species of tree known in Portuguese India', and *re**fém*** [xefẽ] 'hostage'. We acknowledge that these contexts do not control for coarticulatory effects on the nasal vowels. However, as the POA for all of the prevocalic stops is anterior, any velar constriction can reasonably be ascribed to the emergence of a velar coda consonant.

Target words were placed in the carrier phrase *diga X agora* 'say X now'. Phrases were presented in randomized order in the 3T Siemens Trio MRI scanner at the Beckman Institute for Advanced Science and Technology, at the University of Illinois. The speaker was instructed to repeat the phrase at her normal speaking rate until the noise of the scanner ceased (about 2 minutes). A single mid-sagittal slice was used for this study, with a frame rate of 100 frames per second. The image resolution of each slice is 128x128 voxels, and the resolution of each voxel is 2.2 mm x 2.2 mm x 6.5 mm (through-plane depth). Due to different speaking rates between trials, an unequal number of tokens was collected for each lexical item: *bafum* (48), *tupã* (48), *capim* (52), *dabom* (48), and *refém* (49).

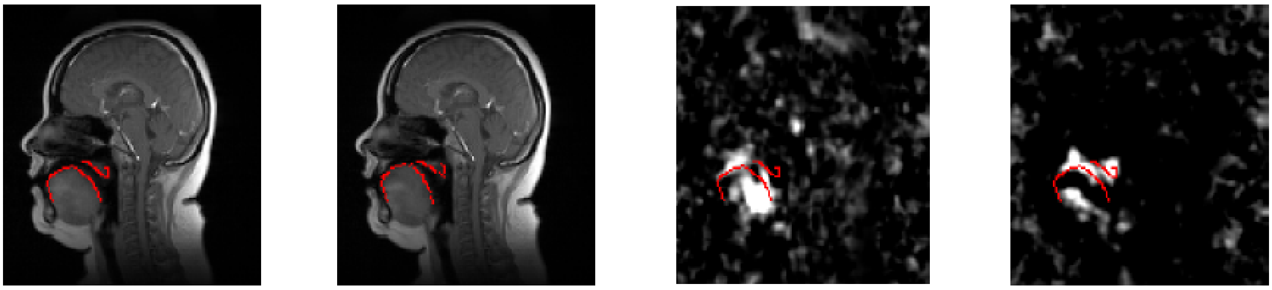

Figure 1: *Example of the derivation of the tongue and velum contours of the midpoint of /i/ in 'capim'. Left to right: reference image with reference tongue and velum contours, randomly selected test image with contours calculated with the deformation fields, horizontal deformation field (*$\mathbf{Dx}$*) with reference contours, vertical deformation field (*$\mathbf{Dy}$*) with reference contours.*

2.2. MR Image Reconstruction

In this work, we applied a Partial Separability (PS) model-based method with cone navigation [14] to recover high-quality spatiotemporal images from rt-MRI experiments. Specifically, the PS model assumes L^{th} order partial separability of the desired image, $I(\mathbf{k}, t) = \sum_{l=1}^{L} \Psi_l(\mathbf{r})\Phi_l(t)$, where $\Psi_l(\mathbf{r})$ and $\Phi_l(t)$ represent the l^{th} spatial and temporal basis function, respectively. In a matrix-vector form, the PS model can be expressed as **I = UV**, where **U** and **V** represent the spatial and temporal subspaces of the desired image function **I**, respectively. **I** can be recovered when **U** and **V** are determined.

We choose to determine **U** and **V** using an approach initially proposed in [15]. Specifically, singular value decomposition is first performed to determine **V**. With **V** determined, **U** can be estimated through a least-squares fitting procedure. However, previous research has shown that such procedure may be susceptible to ill-conditioning [14], especially when a high model order is used but limited measurements are available. Extending upon a strategy proposed in [14], we propose to estimate **U** by solving a regularized least-squares problem,

$$\widehat{\mathbf{U}} = \underset{\mathbf{U}}{\operatorname{argmin}} \|\mathbf{d} - \mathbf{E}\{\mathbf{F}_S\mathbf{UV}\}\|_2^2 + \lambda\|\mathbf{D}_{ref}\{\mathbf{UV}\}\|_1, \quad (1)$$

where $\mathbf{E}$ is an imaging operator consisting of sparse sampling and sensitivity encoding, $\mathbf{F}_S$ is a spatial Fourier transform matrix, $\mathbf{D}_{ref}$ is a composite operator consisting of two operations: 1) a non-rigid spatiotemporal warping transform from $\mathbf{I}$ to a reference image $\mathbf{I}_{ref}$; 2) a temporal Fourier transform applied to the warped image. In particular, the spatiotemporal warping transform is performed based on a dense deformation field $\mathbf{\Theta}$ that represents pixel-wise coordinate displacement of $\mathbf{I}$ with respect to $\mathbf{I}_{ref}$ in both horizontal and vertical directions.

It should be noted that the proposed method aims to determine **U**, but also obtains $\mathbf{\Theta}$ as an important by-product via the alternating minimization procedures: With a determined $\mathbf{\Theta}$, **U** is estimated using an algorithm based on additive half-quadratic regularization [16]; With **U** determined, $\mathbf{\Theta}$ is updated using the *imregdemons* function in MATLAB 2014b [17] to efficiently estimate the corresponding deformation between $\mathbf{I}$ and $\mathbf{I}_{ref}$. The latter step is similar to an approach taken in [18].

2.3. Data Processing

The reconstructed reference image $I_{ref}(\mathbf{r}, t)$ for each vowel was converted to a black and white edge image in Matlab 2014b [17] using the Canny method of edge detection with the *edge* function at a 0.05 threshold. The tongue contour and velum on the reference image were extracted using the *bwboundaries* function on the black and white images, and manually corrected as needed. The points corresponding to the tongue back and lower surface of the velum were manually selected for analysis of distance.

The spectrograms were derived from synchronized, noise-cancelled audio recorded with a MR-compatible headset, with attached optical microphone worn by the subject (Dual Channel-FOMRI, Optoacoustics, Or Yehuda, Israel). The start and end point of each vowel were manually segmented in Praat [19]. For each image of interest I_{n} within the vowel's duration, the position of the tongue and velum were calculated by subtracting the deformation values from the reference image's coordinates; that is, each point in the tongue and velum contour of I_{n} is calculated as

$$(x_{\mathrm{n}}, y_{\mathrm{n}}) = (\mathbf{Dx}_{ref}\{x_{\mathrm{ref}}\}, \mathbf{Dy}_{ref}\{y_{\mathrm{ref}}\}) \quad (2)$$

where $\mathbf{Dx}$ is horizontal deformation, and $\mathbf{Dy}$ is vertical deformation. The tongue contour of the reference image and derived tongue contour of a test image, with the accompanying deformation fields, can be seen in Figure 1.

Cartesian distance was calculated between points in the tongue dorsum and in the lower surface of the velum. We employed this dorso-velic distance (DVD) measure to determine amount of constriction, and whether a complete occlusion occurs. We assume that if DVD is equal to zero, a complete occlusion has occurred. Statistical analyses were done in R [20]. Smoothing Spline ANOVA (SSANOVA) [21], a useful—though underutilized—tool for capturing dynamic characteristics of speech, was used to compare DVD across the normalized duration of the vowels, using the R package *gss* [22]. This was done in order to determine the timing of nasal coda emergence within the vowel.

In order to determine whether any velar closures were complete, the velar stop /k/ segmented from the word *capim* was analyzed in the same manner described above for comparison. Because velar stops display a complete occlusion, if a vowel terminates in a constriction with the same DVD as that of /k/, we can conclude that a velar consonant has emerged at the end of these nasal vowels.

It is possible to derive these distance measures by means of calculating all tongue and velum contours within the sounds of interest individually. However, this is computationally very expensive. Furthermore, current methods are subject to much error, which then requires a large amount of manual correction. The use of deformation fields in calculating the contours is computationally much faster, and is more accurate than previous methods.

3. Results

Results show that there is a significant difference in DVD between front vowels (/ĩ/ and /ẽ/) and back vowels (/ɐ̃/, /õ/, and /ũ/) at all points throughout the vowel durations (t = 36.1600, p <0.0001). Comparing individual segments, the front nasal vowels /ĩ/ and /ẽ/ have rather large DVD values, even at the point in time when minimum distance occurs. We see that /ẽ/ manifests a larger DVD than /ĩ/. The back vowels manifest a much smaller DVD, indicating a much higher degree of constriction (7.4 mm on average, compared to 14 mm for front vowels). However, these distances are still positive, suggesting that a full closure has not occurred.

To study timing of the constriction, we compared the normalized durations of the vowels. For back vowels, the degree of constriction grows steadily throughout the vowel, so that maximum constriction (i.e., minimum DVD measure) is found approaching the end of the vowel. Figure 2 shows the minimum DVD values for each of the vowels found within the last 25% of each vowel. In all cases except /ɐ̃/, the minimum values occur within the last 25% of the vowel's duration. Within the back vowels, /ɐ̃/ has the highest average DVD value, while /ũ/ and /õ/ have smaller aperture sizes. While /ɐ̃/ has a steadily increasing DVD throughout the vowel, the increase in velar opening is not very large (on average, no more than a 2 mm difference).

A SSANOVA model shows that across the duration of the back vowels, the DVD measure decreases throughout the duration of the vowel, as seen in Figure 3. This is indicative of a constriction in the posterior section of the vocal tract manifesting throughout the duration of the vowel. This constriction is very narrow, averaging an opening of 7 mm at the end of both /ũ/ and /õ/.

Finally, we compared /k/ with the back nasal vowels to determine whether the constriction made at the end of the vowels approached that of a full closure. A SSANOVA model shows that /k/ and /ũ/ exhibit the same amount of closure in the second half of their respective durations, while /k/ and /õ/ distances converge in the last 90% of the respective sounds, as also seen in Figure 3. Interestingly, /k/ exhibited a very low, but positive, DVD measure throughout the sound (mean = 6.82 mm, SD = 0.68 mm). This could imply that the /k/ does not undergo complete occlusion in this setting. However, this is an onset /k/, and is therefore unlikely to undergo lenition. We consider this distance to be the result of error, and therefore to be the threshold for whether a complete occlusion occurs; that is, if a vowel displays a DVD of 6.82 mm or less, it is considered a complete closure similar to that of the stop /k/.

4. Discussion

The results of an rt-MRI study indicate a difference in posterior oral constrictions between front and back nasal vowels. Specifically, front vowels have a wider aperture between the soft palate and the tongue dorsum than back vowels. This is undoubtedly because the tongue is being stretched forward towards the front of the hard palate for front vowels. Front nasal vowels in BP have been shown to centralize towards the center of the oral cavity throughout the vowel [23], causing the tongue back to recede towards the velum, creating a smaller aperture. However, given

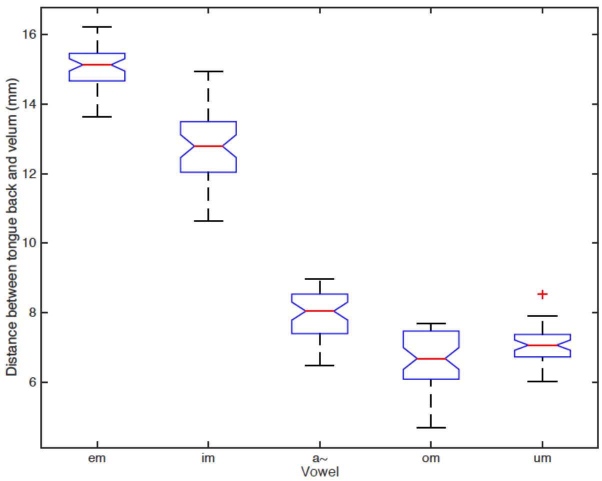

Figure 2: *Minimum dorso-velic distance (DVD) during the last 25% of each nasal vowel, in millimeters ('em' =* [ẽ]*, 'im' =* [ĩ]*, 'a∼' =* [ɐ̃]*, 'om' =* [õ]*, 'um' =* [ũ]*).*

the nature of the front vowels, this aperture still remains quite high (on average, 14 mm).

Non-front vowels show a much smaller DVD in this region, due to the POA of the vowels. The low vowel /ɐ̃/ maintains a higher value for DVD than the higher vowels /ũ/ and /õ/, due to the intrinsic lingual height differences in these vowels.This aperture increases throughout the duration of the vowel /ɐ̃/. As seen in [23] the vowel /ɐ̃/ has an associated fronting of the tongue throughout the duration of the vowel, which allows for the increase of the velar aperture; however, this difference is very subtle.

The motion of the tongue and velum indicate that back nasal vowels in BP manifest increased linguo-velic constriction at the end of the vowel. This suggests the emergence of a coda consonant for these vowels. Previous work [23] on the oral articulation of nasal/oral vowel congeners in BP shows that for the high vowel /ũ/, the tongue back lowers throughout the duration of the vowel to permit space for the velum to lower. While this movement can be considered a strategy to make space for the velum, this movement matched with velar lowering together create an occlusion in the velar region, i.e., a velar consonant.

As shown in Figure 2, these back vowels do not necessarily form a complete occlusion between the tongue back and velum at any point. In other words, DVD is never equal to zero and the central tendency for each vowel is always above the threshold for complete closure associated with /k/ (DVD=6.82 mm). Therefore, unlike the complete post-palatal closure shown in [10] using EPG, it appears post-vocalic velar constrictions for this speaker are not full stops. However, the coda constriction evident after nasal back vowels is suggestive of an environment where a coda could easily appear. As the emergence of such coda consonants is known to be variable in frequency, it is likely that this is currently a speaker-specific phenomenon so further study is needed with a larger number of speakers. Though the post-vocalic constriction we observed only tends toward full contact, it is possible that the gesture still represents an articulatory strategy that further differentiates oral and nasal vowel congeners.

Furthermore, comparison with the velar stop /k/ shows a convergence of constriction patterns at the end of the nasal

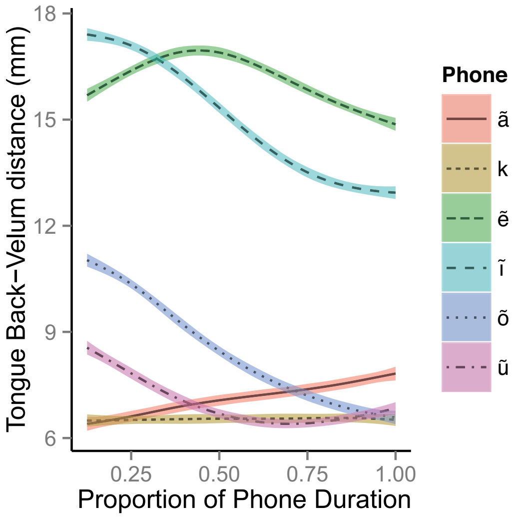

Figure 3: *SSANOVA plot displaying the average distance between velum and tongue back throughout the duration of the nasal vowels and the velar stop /k/.*

vowels with that of the stop, indicating that these nasal vowels are patterning similar to velar stops. This gives promising evidence that, even if this speaker does not show full occlusion in nasal codas, these consonants are approaching full velar contact. Considering the velar stop /k/ as a full occlusion and normalizing as such shows that the vowels /ũ/ and /õ/ very closely approach occlusion in the posterior region of the vocal tract.

Durational cues [10] suggest that post-vocalic occlusions have a similar phonological status to BP nasal consonants. Therefore, the emergence of a velar constriction following back vowels, paired with the emergence of the post-palatal occlusion for front vowels, is indicative of sound change in BP as these coda consonants emerge. Further work is needed to determine whether this change is motivated by the articulatory configuration of the tongue and velum, or by the need to further differentiate oral/nasal vowel congeners by way of acoustics.

It must be noted that while the canonical description of stops includes full occlusion at the POA, the control case of the velar stop /k/ suggested an incomplete closure between the tongue dorsum and the velum. It is possible, though rare, for voiceless stops in BP to be deleted [1], therefore manifesting an incomplete closure in the onset phoneme. It is also possible that gestural undershoot is occurring due to the subject's fast speaking rate, as seen in studies of spirantization in Spanish [24]. However, it is most likely that, like all other instruments used in phonetics research, there is a small amount of error that needs to be taken into account. Normalizing the data so that /k/ is considered a full closure allows for more realistic comparisons within the data, and normalizing all of the data with respect to /k/ further shows the emergence of close to complete occlusion at the end of the back vowels.

Finally, this study highlights the need for advanced imaging techniques to study posterior regions of the vocal tract during speech. We also highlight the utility of using advanced image processing techniques to study the differences between multiple images. The integration of curve fitting techniques with deformation-based calculations allows for analyses to be done without much computational expense, and with much lower need for manual correction.

5. Conclusion

We present rt-MRI evidence showing the emergence of an incomplete, though narrow constriction between the tongue dorsum and velum within the back vowels /ũ/ and /õ/. We argue that this is related to the emergence of post-palatal coda consonants following /ĩ/ and /ẽ/ in BP. These two coda emergences are the potential effects of a sound change in progress, the end result of which augments the articulatory distinction between oral and nasal vowel congeners in BP.

6. References

[1] P. A. Barbosa and E. C. Albano, "Brazilian Portuguese," *Journal of the International Phonetic Association*, vol. 34, no. 2, pp. 227–232, 2004.

[2] O. R. Kelm, "Acoustic characteristics of oral vs. nasalized /a/ in Brazilian Portuguese: Variation in vowel timbre and duration," *Hispania*, vol. 72, no. 4, pp. 853–861, 1989. [Online]. Available: http://www.jstor.org/stable/343563

[3] A. Almeida, *The Portuguese nasal vowels: phonetics and phonemics.* North-Holland Publishing Company, 1976, pp. 349–396.

[4] J. M. Lipski, "Brazilian Portuguese vowel nasalization: Secondary aspects," *Canadian Journal of Linguistics*, vol. 20, pp. 50–77, 1975.

[5] S. Parkinson, "Portuguese nasal vowels as phonological diphthongs," *Lingua*, vol. 61, pp. 157–177, 1983.

[6] R. Sampson, *Nasal vowel evolution in Romance*. Oxford: Oxford University Press, 1999.

[7] R. L. Diehl, K. R. Kluender, M. A. Walsh, and E. M. Parker, "Auditory enhancement in speech perception and phonology," in *Cognition and the symbolic processes, Vol 3: Applied and Ecological perspectives*, R. R. Hoffman and D. Palermo, Eds. Hillsdale, NJ: Erlbaum, 1991, pp. 59–76.

[8] P. S. Beddor, "The perception of nasal vowels," in *Phonetics and phonology: Nasals, nasalization and the velum*, M. Huffman and R. Krakow, Eds. New York: Academic Press, 1993, pp. 171–196.

[9] V. Delvaux, "Perception du contraste de nasalité vocalique en français," *Journal of French Language Studies*, vol. 19, pp. 25–59, 3 2009.

[10] R. Shosted, "Excrescent nasal codas in Brazilian Portuguese: An electropalatographic study," in *Proceedings of the XVIIth International Congress of Phonetic Sciences*, W.-S. Lee and E. Zee, Eds. Hong Kong: City University of Hong Kong, 2011, pp. 1834–1837.

[11] R. K. Shosted, "Vocalic context as a condition for nasal coda emergence: aerodynamic evidence," *Journal of the International Phonetic Association*, vol. 36, pp. 39–58, 6 2006.

[12] C. Carignan, R. Shosted, M. Fu, Z.-P. Liang, and B. P. Sutton, "The role of the pharynx and tongue in enhancement of vowel nasalization: a real-time MRI investigation of French nasal vowels," in *INTERSPEECH*, 2013, pp. 3042–3046.

[13] C. Carignan, R. Shosted, M. Fu, Z.-P. Liang, and B. Sutton, "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*, vol. 50, pp. 34–51, 2015.

[14] M. Fu, B. Zhao, C. Carignan, R. K. Shosted, J. L. Perry, D. P. Kuehn, Z.-P. Liang, and B. P. Sutton, "High-resolution dynamic speech imaging with joint low-rank and sparsity constraints," *Magnetic Resonance in Medicine*, vol. 73, no. 5, pp. 1820–1832, 2015. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1002/mrm.25302/abstract

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

[16] B. Zhao, J. Haldar, A. Christodoulou, and Z.-P. Liang, "Image reconstruction from highly undersampled ($\mathbf{k}$,t)-space data with joint partial separability and sparsity constraints," *IEEE Transactions on Medical Imaging*, vol. 31, no. 9, pp. 1809–1820, Sept 2012.

[17] MATLAB, *version 8.4.0.15021 (R2014b)*. Natick, Massachusetts: The MathWorks Inc., 2014.

[18] S. Lingala, E. DiBella, and M. Jacob, "Deformation corrected compressed sensing (DC-CS): A novel framework for accelerated dynamic MRI," *IEEE Transactions on Medical Imaging*, vol. 34, no. 1, pp. 72–85, Jan 2015.

[19] P. Boersma and D. Weenink, *Praat: doing phonetics by computer [Computer program], Version 5.3.32*, 2012. [Online]. Available: http://www.praat.org/

[20] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2007. [Online]. Available: http://www.R-project.org/

[21] C. Gu, *Smoothing Spline ANOVA Models*, ser. Springer Series in Statistics. New York: Springer, 2002.

[22] ——, *gss: General Smoothing Splines*, 2007, r package version 1-0. [Online]. Available: http://cran.r-project.org/web/packages/gss/gss.pdf

[23] M. Barlaz, M. Fu, J. Dubin, R. Shosted, Z.-P. Liang, and B. Sutton, "Lingual differences in Brazilian Portuguese oral and nasal vowels: An MRI study," in *Proceedings of the XVIIIth International Congress of Phonetic Sciences*, To Appear.

[24] B. Parrell, "Dynamical account of how /b, d, g/ differ from /p, t, k/ in Spanish: Evidence from labials," *Laboratory Phonology*, vol. 2, no. 2, pp. 423–449, 2011.