INTERSPEECH 2013

Observations of perseverative coarticulation in lateral approximants using MRI

Nicole Wong[1]*, Maojing Fu*[2]*, Zhi-Pei Liang*[2]*, Ryan K. Shosted*[1]*, Bradley P. Sutton*[3]

[1]Department of Linguistics, [2]Department of Electrical and Computer Engineering,
[3]Department of Bioengineering, University of Illinois at Urbana-Champaign
{nwwong2, mfu2, z-liang, rshosted, bsutton}@illinois.edu

Abstract

Recent advances in real-time magnetic resonance imaging (rt-MRI) make possible the following study of post-dorsal coarticulatory effects on lateral approximants in running speech. As the MRI is particularly well-suited for investigating pharyngeal articulations, it was capable of capturing perseverative coarticulation from a low central vowel to the immediately following syllable-initial lateral approximant. The lowered and retracted tongue body position associated with a low central vowel is articulatorily equivalent to a pharyngeal constriction [1]. We show here that rt-MRI techniques can track coarticulation direction by calculating the average pixel intensity [2] within a region of interest in MR images as a function of time.

Index Terms: magnetic resonance, pharynx, lateral approximants, Tigrinya

1. Introduction

Multiple studies have investigated coarticulation and variation in laterals cross-linguistically, with particular attention paid to the perception and distribution of the allophonic variation of “dark” and “light” /l/ in English [3, 4] as well as Romance languages like Eastern Catalan and European Portuguese [5, 6]. Others have investigated the articulation of laterals using a variety of methods ranging from x-ray [7] to electropalatography [8], and Magnetic Resonance Imaging (MRI) [9]. While studies agree that a posterior constriction characterizes so-called “dark” laterals [3, 6], little research has been conducted on the specific articulatory configuration that contributes to the perception of “darkness”. Instead, “darkness” is typically interpreted with respect to acoustics, specifically the relative lowering of F2 [6]. Furthermore, articulatory descriptions of “dark” laterals can include tongue dorsum raising and retraction or even both simultaneously [3, 8, 9]. This suggests that “darkness” may simply be an umbrella term that covers a range of dorsal articulations from velarization to pharyngealization. Given that “dark” /l/ in English is characterized acoustically by both its relatively low F2 and high F1 with respect to “clear” /l/, it is reasonable to suggest that a combination of constrictions at the velar and pharyngeal area may be responsible for the perception of “darkness” [10, 11]. In comparison, Ladefoged [12] identifies a single velar constriction (i.e. arching of the tongue dorsum) as characteristic of a “dark” lateral.

This study seeks to address two concerns before examining the articulations responsible for the perception of “darkness” in a future study. First, we wish to establish whether it is possible to distinguish between two sounds that possess similar articulatory configurations but different manners of articulation, using rt-MRI. Given that “clear” /l/ lacks the dorsal gesture associated with “dark” /l/, “clear” /l/ should look quite similar to the alveolar plosive /t/ in mid-sagittal MR images. Assuming no coarticulation, both /t/ and /l/ should be articulated with a closure at the alveolar ridge. Secondly, we wish to establish a starting point for a more intensive study of the articulation of “darkness” in laterals.

We will address these two concerns by conducting a precursory analysis of a readily available dataset of real-time MR images which were obtained from a single male native speaker of Tigrinya (Ethio-Semitic; Eritrea) [13, 14, 15] for a different experiment. The phonological system of Tigrinya is well-documented [16, 17]. In addition, there is a good deal of literature on the obstruent system [18, 19, 20, 21]. However, it is not readily apparent whether the lateral /l/ in Tigrinya has “light” and “dark” allophones. Furthermore, we leave the effects of syllable position to future study, since all instances of /t/ and /l/ in the dataset used here happen to be syllable-initial (as the originally-intended experiment was not designed to study laterals at all). Therefore, this study will not directly address the perception of “darkness” in Tigrinya, but the extent to which coarticulation of /l/ is observable using rt-MRI.

rt-MRI allows us to safely obtain detailed and accurate images of the vocal tract, particularly the post-velar region [9]. While static MRI has been shown to successfully capture effects of syllable position in laterals [9, 6], this method requires sustained production and the results are thus most likely representative of hyper-articulated speech [22]. As such, rt-MRI is particularly well-suited to studying post-velar coarticulation in running speech. Additionally, improvements in the spatial and temporal resolution of rt-MRI are advancing rapidly.

2. Methods

rt-MR images of the vocal tract of a single male native speaker of Tigrinya were acquired using a MAGNETOM Trio 3T system. A Partial Separability model-based imaging method [23, 24] was used to achieve high frame-rate, single-slice images of the moving vocal tract. A single slice in the mid-sagittal region of the vocal tract was acquired at 102.25 frames per second with voxel size 2.2 mm × 2.2 mm × 8 mm (through-plane). The subject, in a supine position, produced nine test words embedded in the carrier phrase /təmali __ ʔilu/ ‘Yesterday __ he said’ [20]. The carrier phrase contained the relevant sequences /tə/, /ali/, and /ʔilu/. The nine test words were recorded with another experiment in mind (cf. [20]); material from the carrier phrase and one test word, which happened to contain a lateral, will be considered in this paper. Each test word was three syllables long. Stress in Tigrinya generally falls word-finally. The subject repeated each carrier phrase as many times as possible in an approximately 75-second window, during which MR data and audio were acquired. This resulted in around 18 repetitions of the phrase in each recording.

The participant's speech was simultaneously audio recorded at 8 kHz using a noise-cancelling microphone during the imaging procedure and was later synchronized with the MR signal. Segmentation of /t/ and /l/ was conducted manually using Praat [25] with visual guidance provided by spectrograms of the recorded audio; the boundaries of /t/ were determined by measuring from the start of the burst to the end of the aspiration, while the boundaries of /l/ were determined by measuring from the beginning to the end of the sudden drop in amplitude that characterizes laterals [26]. Since the closure of /t/ was not measurable due to its word initial position, an additional 0.045 s was subtracted from the start of the burst to estimate the closure duration [27]. Despite the use of a noise-cancelling microphone, the resulting acoustic signal still contained enough noise to obscure formant frequencies. It was therefore not possible to analyze the formants of the segmented laterals. The segmented time intervals for /t/ contained approximately 10 MR images per interval, while the segmented time intervals for /l/ contained approximately 6-9 MR images per interval. Using MATLAB R2011a, the MR images were then averaged pixel-by-pixel into a single image within each segmented time interval, such that one averaged MR image represented each uttered phone. In total, this elicitation procedure resulted in 160 instances of /tə/, 140 of /ali/, and 160 of /ilu/ (including 18 instances of /ilu/ from the test word /bəxʷʼilu/ 'it germinated').

There are several reasons for the uneven number of tokens. Because of the machine noise, it was occasionally difficult to identify the presence of a lateral in both the waveform and spectrogram. Secondly, the subject would occasionally grow tired midway through acquisitions and would begin to use less precise articulations when producing the carrier phrase. This created additional difficulties when attempting to segment the lateral. As a result, all ambiguous tokens were discarded, of which the majority were laterals from the sequence /ali/. Fortunately for the purpose of this study, most of the boundaries for /t/ and /l/ were reasonably unambiguous to identify.

After initial examination of these images for the purpose of determining the placement of potential regions of interest (ROIs), two observations were made: 1) the articulatory configuration of the consonants in /tə/ and /ilu/ were comparable, with a distinct constriction at the alveolar ridge and no observable coarticulation at any of the three ROIs; 2) the consonant in /ali/ seemed to show subtle differences from both /tə/ and /ilu/: widening of the lip aperture, a constriction at the lower pharynx, and a slight opening of the velopharyngeal port (see Figure 1).

From these observations, two post-hoc hypotheses were developed: 1) the consonants in the sequences /tə/ and /ilu/ will not be distinguishable mid-sagitally; 2) the lateral in /ali/ will manifest significant differences from the consonants in both /tə/ and /ilu/. To test these hypotheses, three ROIs were identified upon a precursory survey of the data set (see Figure 1): 1) lower lip (LL), 2) velopharyngeal port (VP), and 3) hypopharynx (hP). Since the subject's head remained in a fixed position throughout the entire MRI session, the pixel coordinates of these regions were fixed across acquisitions. The final analysis of the average pixel intensity (API) in a given ROI was conducted on these images, permitting statistical analysis. In a grayscale image, the intensity of a pixel is given by a numerical value ranging from 0 to 1, which represents a shade of gray from white (1) to black (0). A high pixel intensity close to 1 suggests the presence of tissue, while lower pixel intensities suggest the presence of bone or air, i.e., a relative absence of tissue. The API is calculated by summing the intensity of each pixel in the ROI and dividing the sum by the number of pixels in the ROI. Following Proctor and colleagues [2], API was used as a measure to determine whether articulators move in or out of a given ROI, as deduced from the increase or decrease of the API.

3. Discussion of results

The distribution of the data was first checked for normality. The Shapiro-Wilk test applied to each of three ROIs for the consonants in /tə/, /ali/, and /ilu/ resulted in a rejection of the null hypothesis ($p < 0.05$), suggesting that the APIs were not normally distributed. Therefore, the non-parametric Mann-Whitney-Wilcoxon test was applied to determine whether API differed within each of the three ROIs across the consonants in the sequences /tə/, /ali/, and /ilu/. Since the tests were applied multiple times, the significance level $\alpha = 0.01$ was Bonferroni-corrected to avoid Type I error.

Table 1: p-*values given for comparisons between APIs of ROIs at the lower lip (LL), velopharyngeal port (VP), and hypopharynx (hP) for the consonants in* /tə/, /ali/, *and* /ilu/. *Corrected significance level* $\alpha' = 0.001$, *with significant differences starred.*

ROI	/tə/ × /ali/	/tə/ × /ilu/	/ali/ × /ilu/
LL	2.2e-16*	0.002	3.87e-15*
VP	9.73 e-12*	0.031	2.416e-14*
hP	2.2e-16*	0.353	2.2e-16*

Given the corrected significance level, the statistical results suggest that the articulatory configurations of /t/ and the /l/ in the sequence /ilu/ are not significantly different with regards to the three ROIs (LL, VP, hP: $p > 0.001$), while the articulatory configuration of /l/ in the sequence /ali/ is significantly different from that of both the /l/ in /ilu/ (LL, VP, hP: $p < 0.001$) and /t/ (LL, VP, hP: $p < 0.001$) at the lower lip, velopharyngeal port, and hypopharynx. This difference suggests that the /l/ in the sequence /ali/ may be coarticulated, possibly with the low (pharyngeal) vowel /a/. The direction of coarticulation will be addressed below.

To determine the source of coarticulation observed above, further investigation was conducted on non-averaged MR images of the coarticulated lateral in /ali/. Figure 2 presents the selected sequence /təmali/, with a rendering of the time-varying position of the articulators using API measures. Rising API is interpreted as the movement of tissue into the ROI (most likely the tongue root or lower lip for the respective ROIs at the hypopharynx and lower lip), while API measures of 0.1 or less suggest a relative absence of tissue in the ROI (at the velum, most likely the opening of the velopharyngeal port). Figure 2 is complemented by Figure 3, which displays a spectrogram of the same sequence /təmali/ and includes the same phone boundaries.

In Figure 2, the termination of the aspiration for /t/ is identified at 49.52 s, with the estimated onset of the closure [27] for /t/ at 49.47 s. The following /ə/ is not visible, even in the spectrogram (see Figure 3). The /ə/ may have been voiceless, low-amplitude, or deleted by the speaker; in any case, its presence was difficult to identify. We use the sudden rise in API for the lower lip ROI at 49.55 s to estimate the end of the schwa / onset of the nasal /m/. The difficulty in segmenting the vowel /ə/ and the nasal /m/ is consistent across repetitions, however, it

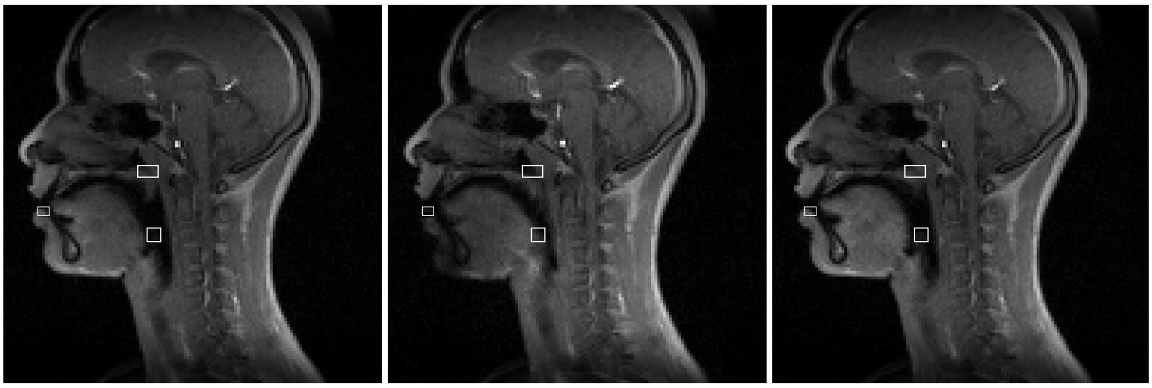

Figure 1: *Averaged MR images for the duration of [t] or [l] in the sequences* /tə/ *(left),* /ali/ *(center), and* /ilu/ *(right). ROIs are indicated in white.*

is not problematic for the research questions in this study as it did not affect the segmentation of /l/ and /t/. The lateral is easily identified by the drop in amplitude, while the termination of the plosive was marked by the end of the aspiration. Difficulty identifying the onset of the closure for word-initial voiceless plosives is ambiguous regardless of noise. Higher amplitude sound pressure at 49.62 s signals the onset of /a/ and the termination of /m/, while the subsequent sharp fall in amplitude at 49.77 s signifies the onset of the lateral and the termination of /a/. The lower amplitude continues until 49.83 s, after which we see the amplitude rise once more for the vowel /i/ until the end of the vowel at 49.95 s. The annotation of the phones /ə, m, a, i/ proceeded based on data from multiple channels (including API) and is intended to guide the exploration of the data in Figure 2, with further spectrographic detail provided in Figure 3. Statistical results are reported only for the segmentation of /t/ and /l/, which were relatively less ambiguous to identify in comparison with the sounds /ə/ and /m/ and could be identified using the audio alone.

Illustrated in Figure 2, the direction of hypopharyngeal coarticulation appears to be perseverative, rather than anticipatory. Given the landmarks in the acoustic signal, the constriction at the hypopharynx observed during the lateral seems to result from the preceding central low (or pharyngeal) vowel /a/. This can be inferred from the rising API measures for the hypopharyngeal ROI. API in the hypopharynx peaks midway through the low vowel /a/ and does not fully subside before overlapping with the onset of the lateral, suggesting that the direction of coarticulation is perseverative. This observation is perhaps unsurprising, as the low and somewhat retracted tongue position associated with the vowel /a/ produces an articulatory configuration that is consistent with a pharyngeal constriction [1]. Similarly, the movement of the lower lip out of the ROI is due to coarticulatory effects from the lowering of the jaw associated with the preceding low vowel. In contrast, the lateral in the sequence /ilu/ remains "clear" throughout (not shown in Figure 2); other than the constriction between the tongue tip and the alveolar ridge, we do not observe the presence of retraction or arching of the tongue dorsum. There is no suggestion of influence from the following high back rounded vowel /u/, such as tongue root retraction or movement of the lower lip.

The bilabial nasal /m/ appears to cause persistent nasalization: the velopharyngeal port remains open for the low central vowel /a/, as well as the lateral in /mali/. The effects appear subtle in Figure 2, however statistical results (see Table 1) support the hypothesis that perseverative nasalization caused significant differences at the velopharyngeal port between the laterals that occurred after /m/ and those that did not. In contrast, the velopharyngeal port remains closed for both the alveolar plosive and the lateral in /ilu/ as both sounds did not follow nasals in the materials used for this study.

Within the three regions of interest observed here, significant opening of the velopharyngeal port during the /l/ was particularly surprising. The MR images (see Figure 2) revealed two possible sources of coarticulation: the preceding vowel /a/ and the nasal /m/. The so-called velic opening hypothesis [28] posits that low vowels manifest increased velopharyngeal port opening. Therefore, it is possible that the low central vowel /a/ may be contributing to the degree of opening at the velopharyngeal port and by association, contributing to potential nasalization. However, further testing of simple /VlV/ disyllables is required in order to support this argument, as carryover nasalization from the nasal /m/ is more likely.

Given that the literature equates retraction of the tongue dorsum and/or lowering of the tongue body with "darkness" [3, 8, 9], this study finds limited evidence that syllable-initial laterals in Tigrinya can become "dark" via perseverative pharyngeal coarticulation from a preceding low central vowel. In addition to undertaking a controlled rt-MRI experiment on the subject, subsequent studies must include acoustic analysis of the formant frequencies to substantiate whether the degree of pharyngeal coarticulation found here is a sufficient cue for the perception of "darkness".

4. Conclusion

Through the comparison of alveolar approximant laterals and alveolar stops, this study demonstrated that rt-MRI is capable of capturing the direction of coarticulation and its effects, i.e. perseverative pharyngealization from the low vowel /a/ to the immediately following lateral approximant. Both concerns identified for investigation in this study are addressed: 1) mid-sagittal MRI is not suitable for distinguishing two sounds that possess similar articulatory configurations but lateral vs. plosive manner of articulation; 2) this technique provides a suitable foundation for a more extensive study of the articulation of "darkness" in laterals.

The data here suggests that syllable-initial laterals may, through perseverative coarticulation with preceding low vowels, become "dark" in articulatory terms. However, since the

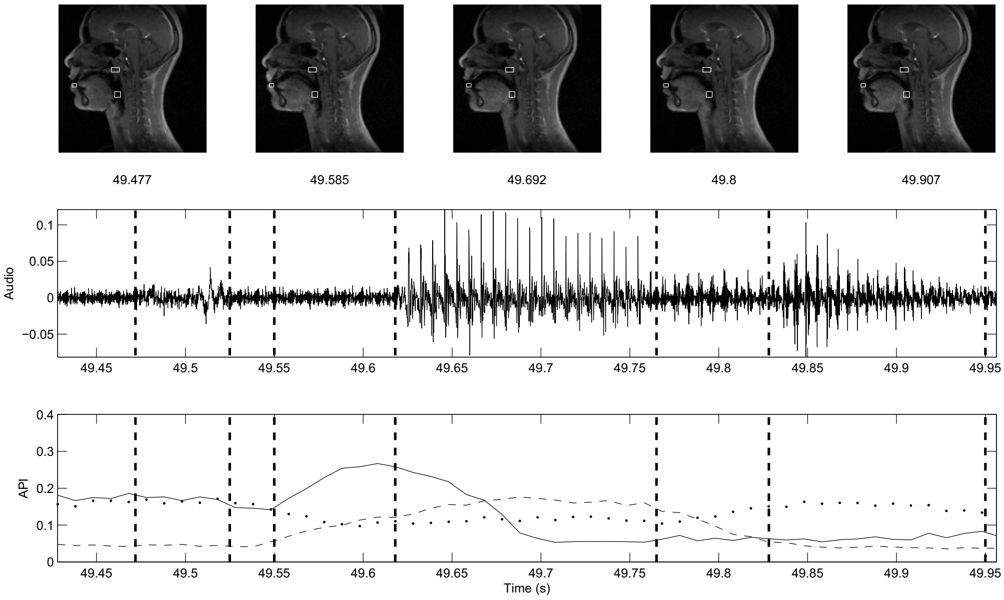

Figure 2: *The sequence* /təmali/. *First tier: Five evenly-spaced MR images represent 55 MR images acquired during the production of* /təmali/. *Second tier: Corresponding acoustic signal. Third tier: Time-varying API for each ROI. API measures for the ROI at the lower lip (LL) are indicated by a solid line; velopharyngeal port (VP) by a dotted line; and hypopharynx (hP) by a dashed line. Annotations for the sounds /ə, m, a, i / are provided solely for illustrative purposes.*

MR data studied here were not collected specifically for evaluation of lateral allophony, it is unknown whether syllable position and lateral allophony are associated with each other in Tigrinya as they are in American English [3]. Further investigation of coda laterals is necessary before we can comment on lateral allophony in Tigrinya.

Due to the previously mentioned constraints of using data collected for another experiment, effects from potential factors such as stress and word length could not be controlled. This is an acknowledged limitation of the study and more controlled experiments are required. In particular, analysis of the acoustics and the formant frequencies of the coarticulated laterals as well as an extended rt-MRI study of the effect of low vowels on laterals (in other positions besides syllable-initial) is necessary for confirming the presence of "dark" and ""light" /l/ allophony in Tigrinya (and other languages). More controlled studies to measure velopharyngeal aperture and to isolate spontaneous nasalization resulting from low vowels [29, 30, 31, 32] are also proposed.

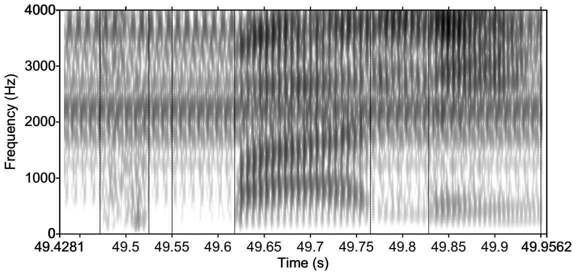

Figure 3: *Spectrogram accompanying audio in Figure 2, with corresponding phone boundaries.*

5. References

[1] Kokawa, T., Saigusa, H., Aino, I., Matsuoka, C., Nakamura, T., Tanuma, K. and Niimi, S., "Physiological studies of retrusive movements of the human tongue", Journal of Voice, 20(3):414-422, 2006.

[2] Proctor M., Lammert A., Katsamanis A., Goldstein L., Hagedorn C. and Narayanan S., "Direct estimation of articulatory kinematics from real-time magnetic resonance image sequences", in *INTERSPEECH-2011*, pp. 281-284.

[3] Sproat, R. and Fujimura, O., "Allophonic variation in English /l/ and its implications for phonetic implementation", Journal of Phonetics, 21:291-311, 1993.

[4] Yuan, J. and Liberman, M., "Investigating /l/ Variation in English through Forced Alignment", in *INTERSPEECH-2009*, pp. 2014-2018.

[5] Recasens, D., Fontdevila, J. and Pallares, M. D., "Linguopalatal coarticulation and alveolar-palatal correlations for velarized and non-velarized/l", Journal of Phonetics, 24(1):165-185, 1996.

[6] Martins, P., Oliveira, C., Ferreira, C. and Silva, A., "3D MRI and semi-automatic segmentation techniques applied to the study of European Portuguese lateral sounds", in Proc. 9th International Seminar on Speech Production (ISSP), 2003, pp. 235-242.

[7] Fant, G. Acoustical Theory of Speech Production. The Netherlands: Mouton, 1960.

[8] Scobbie, J. M. and Pouplier, M., "The role of syllable structure in external sandhi: An EPG study of vocalisation and retraction in word-final English /l/", Journal of Phonetics, 38:240-259, 2010.

[9] Narayanan, S. S., Alwan, A. A. and Haker, K., "Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part I. The laterals", The Journal of the Acoustical Society of America, 101:1064, 1997.

[10] Potter, R. K., Kopp, G. A. and Green, H. C. Visible Speech. D. Van Nostrand Company, 1947.

[11] Lehiste, I. "Acoustical characteristics of selected English consonants", Indiana University, Vol. 34, 1964.

[12] Ladefoged, P. A. Course in Phonetics. Boston: Thomson/Wadsworth, 2006.

[13] Maddieson, Ian and Kristin Precoda. 1991. UPSID-PC: The UCLA Phonological Segment Inventory Database. http://www.linguistics.ucla.edu/faciliti/sales/software.htm (retrieved 15 January 2013).

[14] Leslau,W., "Observations sur quelques dialectes du Tigrigna: Dialectes d'Akkele Gouzay, d'Adoua, et du Hamasen", Journal Asiatique, 231:61-115, 1939.

[15] Leslau, W. Documents Tigrigna (Éthiopien Septentrional): Grammaire et textes. Paris: Librairie C. Klincksieck, 1941.

[16] Pam, Martin David, "Tigrinya phonology." Ph.D. dissertation, City University of New York, New York, 1973.

[17] Fitzgerald, C. M., "More on phonological variation in Tigrinya", In 15th International Conference on Ethiopian Studies, 2006, pp. 763-768.

[18] Kingston, John, "The phonetics and phonology of the timing of oral and glottal events." Ph.D. dissertation, University of California at Berkeley, 1985.

[19] Kogan, Leonid. "Tigrinya", in The Semitic languages, Robert Hetzron, Ed. New York: Taylor & Francis, 1997, pp. 424-445.

[20] Shosted, R. K. and Rose, S., "Affricating ejective fricatives: The case of Tigrinya", Journal of the International Phonetic Association, 41:41-65, 2011.

[21] FreWoldu, Kiros, "The perception and production of Tigrinya stops", in Reports from Uppsala University Department of Linguistics 13, Uppsala: Uppsala University, 1985.

[22] Engwall, O., "Are static MRI measurements representative of dynamic speech? Results from a comparative study using MRI, EPG and EMA", in Proc. ICSLP, 2000, pp. 17-20.

[23] Liang, Z.-P., "Spatiotemporal imaging with partially separable functions", in IEEE. Intl. Symp. Biomed. Imaging., 2007, pp. 988-991.

[24] Fu, M., Christodoulou, A. G., Naber, A. T., Kuehn, D. P., Liang, Z. P. and Sutton, B. P., "High-frame-rate Multislice Speech Imaging with Sparse Samping of (k,t)-space", in Intl. Soc. Magn. Reson. Med., 2012, p. 12.

[25] Boersma, P., and Weenink, D., Praat: doing phonetics by computer [Computer program]. Version 4.3.31. Amsterdam: University of Amsterdam. Online: http://www.fon.hum.uva.nl/praat, 2005.

[26] Johnson, K. Acoustic and Auditory Phonetics. Wiley-Blackwell, 2011.

[27] Reetz, H. and Jongman, A. Phonetics: Transcription, Production, Acoustics, and Perception. Wiley-Blackwell, 2009.

[28] Hajek, J. Universals of Sound Change in Nasalization. Oxford: Blackwell, 1997.

[29] Brücke, E. W. Grundzüge der Physiologie und Systematik der Sprachlaute für Linguisten und Taubstummenlehrer. Vienna: Gerold, 1856.

[30] Czermak, J. N. Wesen und Bildung der Stimm- und Sprachlaute. In Gesammelte Schriften (vol. 2): Populäre Vorträge und Aufsätze [1879]. Leipzig: Wilhelm Engelman, 1869, pp. 76-104.

[31] Ohala, J. J. "Phonetic explanations for nasal sound patterns", in Nasálfest: Papers from a Symposium on Nasals and Nasalization, C. A. Ferguson, L. M. Hyman, and J. J. Ohala, Eds. Stanford: Language Universals Project, Stanford University, 1975, pp. 289-316.

[32] Shosted, Ryan K., "Vocalic context as a condition for nasal coda emergence: aerodynamic evidence", Journal of the International Phonetic Association, 36(1):39-58, 2006.