MULTIMODAL OBSERVATION AND MEASUREMENT OF LARYNX HEIGHT AND STATE DURING PHARYNGEAL SOUNDS

John H. Esling & Scott R. Moisik

University of Victoria, Canada
esling@uvic.ca; srmoisik@uvic.ca

ABSTRACT

Multiple imaging modalities are used to obtain qualitative and quantitative data on laryngeal height and state during pharyngeal consonant production. Videofluoroscopy provides insight into the relative changes in pharyngeal and epilaryngeal lumina, allowing us to compare the roles these structures play in epilaryngeal constriction. Simultaneous laryngoscopy and laryngeal ultrasound allow us to compare the image data of larynx/glottal state directly with quantification of larynx height using optical flow. The data reveal timing relationships between tongue retraction, aryepiglottic constriction, and larynx raising in the formation of pharyngeal stops and fricative trills.

Keywords: pharyngeals, laryngeal measurement, videofluoroscopy, laryngoscopy, ultrasound

1. INTRODUCTION

Imaging and measuring larynx movements during articulation continues to be a challenge. We are generally less informed about the production of pharyngeal consonants than any other category of speech sound. Research on pharyngeals has relied heavily upon x-ray traces, e.g. [2, 8, 9], which has placed considerable emphasis on the role of the 'tongue root' in the production of these sounds. The impact of these investigations can be observed in many phonological accounts of pharyngeal sounds, which attribute agency to the tongue root largely based on the x-ray evidence [1, 14, 16].

More recently, research using laryngoscopy [3, 4, 13] has provided strong evidence that pharyngeal articulation is a function of synergistic constriction of the epilaryngeal tube via contraction of intrinsic laryngeal muscles (TA, aryepiglottic), larynx raising, tongue retraction, and epilaryngeal tube constriction in general [6].

Quantifying larynx height during speech has remained one of the more challenging vocal tract variables to measure [10]. Attempts have been made using thyroumbrometry [7, 17], but this requires participants to have visible laryngeal prominences. MRI can be used successfully [10], but it provides the best imaging only when the participant maintains a static speech posture.

In this research, we have applied a battery of imaging modalities to gain a deeper insight into pharyngeal/laryngeal articulation. We first use videofluoroscopy to illustrate the importance of epilaryngeal tube constriction in pharyngeal consonant production. Then, we apply a novel imaging technique – simultaneous laryngoscopy and laryngeal ultrasound – to obtain information about larynx state using the laryngoscopic data and information about larynx height using optical flow analysis of the laryngeal ultrasound data. The integration of the optical flow velocity data is particularly revealing in cross-correlating laryngeal postures with vertical height of larynx structures.

2. METHODOLOGY

The methodological components of our study are: (i) videofluoroscopy (VFS) and (ii) simultaneous laryngoscopy and laryngeal ultrasound (SLLUS).

2.1. Speech data

Our data consist of canonical productions of pharyngeal consonants by a 60-year-old male phonetician. The sounds are: aryepiglotto-epiglottal stop [ʡ] and fricative trills [ʜ] and [ʢ]. In (i) the vocalic context is /aCa/, and in (ii) it is /iCi/. The difference in vocalic context reflects the differing exigencies of the imaging modalities.

2.2. (i) Videofluoroscopy

The videofluoroscopic data were obtained from the Instituto de Ciências Biomédicas, Departamento de Anatomia, Universidade Federal do Rio de Janeiro with the assistance of Dr. Milton Melciades Barbosa Costa. Although a videofluoroscopic system was used, no barium-impregnated material was ingested by our participant, because the objective was to image speech sound production rather than swallowing. The participant was screened radiographically in the lateral position (from the left). Beam intensity was set to provide optimal imaging of the epilarynx and pharynx. Our participant produced three tokens of each pharyngeal consonant.

2.3. (ii) Simultaneous laryngoscopy and laryngeal ultrasound

The laryngoscopy equipment is an Olympus ENF-P3 flexible fiberoptic nasal laryngoscope fitted with a 28mm wide-angle lens to a Panasonic KS152 camera. The laryngoscopy video was recorded using a Sony DCR-T4V17 digital camcorder. The ultrasound equipment is a portable LOGIQe R5.0.1 system with an 8C-RS probe (General Electric Corporation). Two audio signals were also recorded (at 44100 Hz, 16 bit) for the purposes of signal synchronization. One was captured along with the ultrasound video using a computer running Sony Vegas, the other was recorded using the camcorder. All signals were integrated and carefully aligned using Sony Vegas.

The laryngeal ultrasound provides a coronal view of the larynx by placing the ultrasound probe over the participant's right thyroid lamina, approximately 1cm behind the laryngeal prominence. The probe was held manually with the examiner's thumb placed firmly on the participant's neck so as to stabilize the probe. The ultrasound ruler (cm) was used to determine the pixel-to-mm scale, which is necessary for measuring the velocity of the larynx (in the optical flow analysis).

2.4. Data analysis

All data analysis was carried out using MATLAB (R2009a). For (i), we compared the relative areas of the visible parts of the epilarynx and pharynx. To do this we defined two regions of interest (ROIs) over these areas and converted each to black and white (from grayscale) by choosing suitable luminance threshold values. We then measured the pixel area of each lumen using the MATLAB function 'bwarea'. The areas were made comparable by converting each into a percentage of their respective area maxima (for a given production).

For (ii), separate analytical techniques were required for the two data types. The laryngoscopy was qualitatively analyzed with attention paid to the position of the epiglottis (and tongue), aryepiglottic folds, and visual impression of larynx height (following [5]). The laryngeal ultrasound was used to quantify larynx movements by means of optical flow analysis [11]. A custom optical flow algorithm was used, which is simply a block-wise (15 by 15 pixels) cross-correlation of pixel data in contiguous frame pairs [15]. Global velocity in the image is obtained by taking the mean of the vertical components in the resulting velocity field, ignoring outliers (more than three standard deviations above the mean) and null vectors. Independent validation of the optical flow algorithm was performed to ensure accuracy of the results.

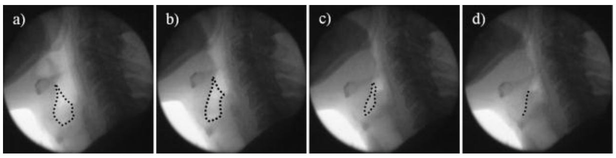
3. RESULTS

Exemplary tokens from the data were selected to illustrate the data we collected.

3.1. (i) Videofluoroscopy results

Selected still frames from the VFS in Figure 1 illustrate the continuum of laryngeal constriction manifest in the reduction in epilarynx area from (a) to (d). The pharynx in comparison does not show the same extent of constriction, never reaching a fully closed state. Figures 2-4 illustrate this relationship by comparing epilarynx area (solid line) with pharynx area (dashed line). Estimated change in larynx height is also shown (dotted line). In each case, the epilarynx always undergoes a greater degree of constrictive reduction and often independently of changes in pharynx area. The larynx rises during the pharyngeal and reaches its peak at the mid point of the second vowel.

Figure 1: Selected frames from videofluoroscopy data showing a continuum of increasing epilarynx tube constriction (dotted outline): (a) deep inhalation; (b) neutral; (c) [aˤ]; (d) [ʡ].



3.2. (ii) Simultaneous laryngoscopy and laryngeal ultrasound results

Results shown in Figures 5-7 for the SLLUS data consist of composite plots of audio, F0 trace, and larynx height (vertical larynx displacement) plots, with frames selected at an equal interval as the laryngoscopy video. There are two F0 traces, one is from the STRAIGHT algorithm [12] and the other is from a custom algorithm based on peak detection.

Figure 2: VFS data for [aʔa]: audio (top); epilarynx area (solid line), pharynx area (dashed line), and larynx height (dotted line), expressed as percentages of maximum.

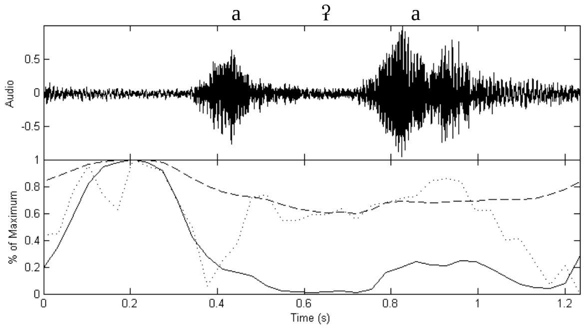


Figure 3: VFS data for [aʜa]. See Figure 2 caption for description.

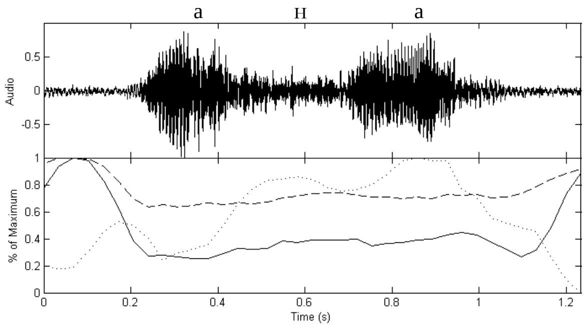
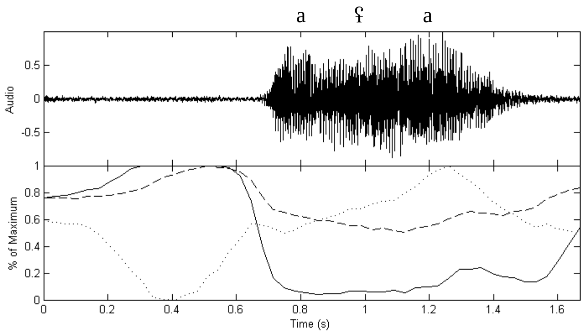


Figure 4: VFS data for [aʕa]. See Figure 2 caption for description.



4. DISCUSSION

The VFS data emphasize the role of the epilarynx in producing pharyngeal consonants. Tongue retraction controls global pharyngeal volume required for the [a] context and serves to displace the epiglottis towards the aryepiglottic folds, but there is clearly an intrinsic laryngeal mechanism which generates ultimate closure of the epilarynx.

Larynx raising is always a concomitant, but it is often delayed towards the end of the consonant and continues rising during the second vowel, possibly serving a secondary role in pitch control.

Figure 5: SLLUS data for [iʔi]: audio (top plot); F0 (middle plot); larynx height (bottom plot); arrows indicate temporal location of selected frames (indexed by number).

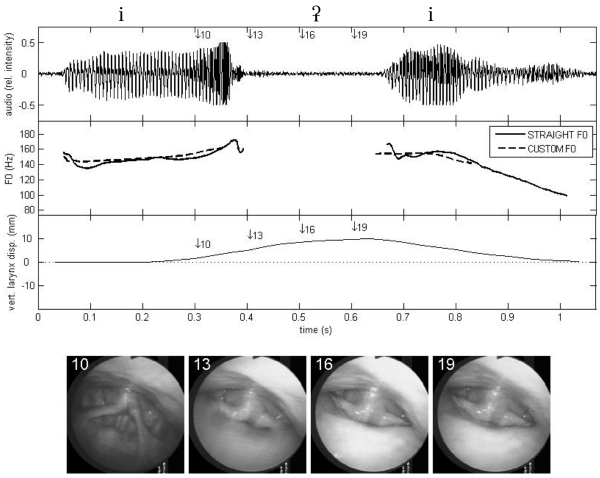


Figure 6: SLLUS data for [iʜi]. See Figure 5 caption for description.

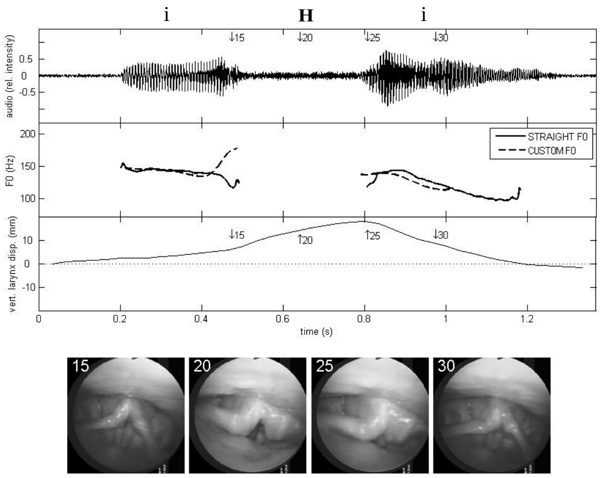
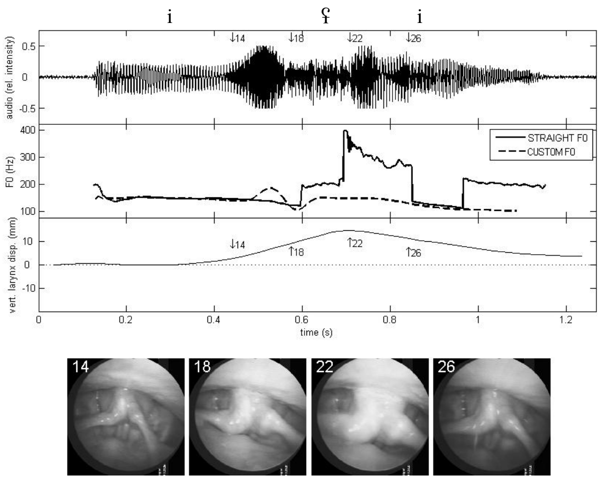


Figure 7: SLLUS data for [iʕi]. See Figure 5 caption for description.



The SLLUS data lead to similar conclusions. The aryepiglottic folds form epilarynx constriction at the onset of the consonant, complemented by tongue retraction (as evidenced by epiglottis position), which reaches its maximum before larynx height peaks towards the offset of the consonant. For our participant, larynx height was usually changed by 10 to 15 mm and shows (visually) some correlation with pitch during the modal voiced context vowels. It begins rapid ascent during the offset of the preceding vowel, which is anticipatory in nature. It is important to note that F0 does not always rise going into the consonant, indicating that the larynx raising is associated purely with the articulation of the pharyngeal at this point. As a mechanical system, the larynx raising/lowering mechanism appears to either be more inertive or lack quick-acting muscle control in contrast to the lingual and intrinsic laryngeal control systems.

The larynx height data obtained from optical flow analysis matched very closely the visual impression of larynx height in the laryngoscopic videos, indicating that the technique is suitable for studying dynamic changes in larynx height.

5. CONCLUSIONS

These data reveal important facts about the temporal sequencing of larynx raising and tongue retraction required to produce the aryepiglotto-epiglottal constriction characteristic of pharyngeals [3, 5]. We observed that there is a cascade of gestures for all of the pharyngeal sounds we analyze. It starts with full aryepiglotto-epiglottal constriction and is followed by a continued lingual retraction gesture. The latter peaks before the peak of the larynx raising gesture, which occurs towards the offset of the consonant.

The use of optical flow analysis in conjunction with simultaneous laryngoscopy and laryngeal ultrasound offers a promising new means to quantify larynx height, and numerous applications can be imagined beyond the study of pharyngeal consonants, such as the relationship between pitch and larynx height and the role of larynx height in the production of ejectives, implosives, and stops in general.

6. ACKNOWLEDGEMENTS

This research was funded by the Social Sciences and Humanities Research Council of Canada, research grant #410-2007-2375.

7. REFERENCES

[1] Archangeli, D., Pulleyblank, D. 1994. *Grounded Phonology*. Cambridge, MA: MIT Press.

[2] Delattre, P. 1971. Pharyngeal features in the consonants of Arabic, German, Spanish, French and American English. *Phonetica* 23, 129-155.

[3] Edmondson, J.A., Esling, J.H. 2006. The valves of the throat and their functioning in tone, vocal register, and stress: Laryngoscopic case studies. *Phonology* 23, 157-191.

[4] Edmondson, J.A., Padayodi, C.M., Hassan, Z.M., Esling, J.H. 2007. The laryngeal articulator: Source and resonator. *Proc. 16th ICPhS* Saarbrücken, 2065-2068.

[5] Esling, J.H., Harris, J.G. 2005. States of the glottis: an articulatory phonetic model based on laryngoscopic observations. In Hardcastle, W.J., Beck, J.M. (eds.), *A figure of speech: A Festschrift for John Laver*. Mahwah, NJ: Erlbaum, 347-383.

[6] Esling, J.H., Zeroual, C., Crevier-Buchman, L. 2007. A study of muscular synergies at the glottal, ventricular and aryepiglottic levels. *Proc. 16th ICPhS* Saarbrücken, 585-588.

[7] Ewan, W.G., Krones, R. 1974. Measuring larynx movement using the thyroumbrometer. *J. Phon.* 2, 327-335.

[8] Ghazeli, S. 1977. *Back Consonants and Backing Coarticulation in Arabic*. Doctoral dissertation, University of Texas, Austin.

[9] Hess, S. 1998. *Pharyngeal Articulations*. Doctoral dissertation, UCLA.

[10] Honda, K. 2004. Physiological factors causing tonal characteristics of speech: From global to local prosody. *Speech Prosody* Nara, Japan.

[11] Horn, B.K.P., Schunk, B.G. 1981. Determining optical flow. *Artificial Intelligence* 17, 185-203.

[12] Kawahara, H., de Cheveigné, A., Patterson, R.D. 1998. An instantaneous-frequency-based pitch extraction method for high quality speech transformation: Revised TEMPO in the STRAIGHT-suite. *Proc. ICSLP '98* Sydney, Australia.

[13] Lindblom, B. 2009. Laryngeal mechanisms in speech: The contributions of Jan Gauffin. *Logopedics Phoniatrics Vocology* 34(4), 149-156.

[14] McCarthy, J. 1994. The phonetics and phonology of Semitic pharyngeals. In Keating, P. (ed.), *Papers in Laboratory Phonology III*. Cambridge: Cambridge University Press, 191-233.

[15] Moisik, S.R., Lin, H., Esling, J.H. 2010. An investigation of laryngeal behavior during Mandarin tone production using simultaneous laryngoscopy and laryngeal ultrasound. *Proceedings of the 9th Phonetics Conference of China*.

[16] Shahin, K. 2002. *Postvelar Harmony*. Amsterdam & Philadelphia: Benjamins.

[17] Sprouse, R.L., Solé, M.J., Ohala, J.J. 2010. Tracking laryngeal height and its impact on voicing. Presented at the *12th Conference on Laboratory Phonology* Albuquerque.