MULTIMODAL IMAGING OF GLOTTAL STOP AND CREAKY VOICE: EVALUATING THE ROLE OF EPILARYNGEAL CONSTRICTION

Scott R. Moisik[1], John H. Esling[2], Lise Crevier-Buchman[3,4], Angélique Amelot[3], Philippe Halimi[4]

[1]MPI for Psycholinguistics, Nijmegen; [2]University of Victoria; [3]CNRS-UMR7018, LPP, Univ. Paris 3; [4]HEGP*
Scott.Moisik@mpi.nl; esling@uvic.ca; lbuchman@numericable.fr;
angelique.amelot@univ-paris3.fr; philippe.halimi@egp.aphp.fr

ABSTRACT

This paper investigates the role of epilaryngeal constriction in constricted laryngeal gestures such as creaky voice and glottal stop. Two imaging modalities were used to study canonical phonetic productions anchored by auditory targets. Laryngeal posture was examined in axial and sagittal sections using magnetic resonance imaging, and high-speed laryngoscopic video was used to study posturing and vibratory dynamics. Together these data support the view that epilaryngeal constriction plays a key role in suppressing vocal fold vibration most directly through a vertical compaction mechanism, whereby the vocal folds and ventricular folds come into contact, obliterating the ventricle. The effect of this is suggested to increase mechanical impedance to mucosal wave transmission resulting in the perturbed vibration observed in creaky voice or facilitating vibratory arrest in glottal stop.

Keywords: Glottal stop, creaky voice, epilaryngeal constriction, larynx, glottis.

1. INTRODUCTION

The larynx is a mechanically and geometrically complex articulating structure comprising three pairs of folded tissues (i.e. the vocal, ventricular, and aryepiglottic folds). In numerous laryngoscopic studies (e.g., [7] and [3]), constricted laryngeal gestures used in producing glottal stop and creaky voice are observed to involve more than just activity of the vocal fold level: there is also a decrease in the antero-posterior (AP) separation between the epiglottis at the front and the collection of structures at the back, most prominently featuring the aryepiglottic folds and their cuneiform tubercles. Concomitant with this AP narrowing is medialization of the ventricular folds, and it has been suggested in [9] that, regardless of the extent of medialization (e.g. whether the ventricular folds come into contact with each other at the midline), the functionally important change is that the ventricular folds press into the vocal folds below and mechanically perturb vocal fold function. The upper laryngeal structures are collectively referred to as *the epilarynx*, and they can be considered to form a mostly independent mechanism of laryngeal constriction over the vocal folds.

Some recent laryngoscopic research, e.g. [2] and [4], has casted doubt on whether epilaryngeal action is phonetically relevant at all, and instead reaffirms earlier models of laryngeal gesturing in speech (such as [5]) which place the functional burden almost entirely on the posturing and behaviour of the vocal folds. In [4], glottal stop is said to infrequently occur with *full* ventricular medialization. In [2], it is claimed that the occurrence of creaky phonation is not strictly correlated with ventricular action. These studies would suggest that epilarynx action is irrelevant to the key aspects of phonetic laryngeal behaviour and best regarded as a hyperarticulation of the larynx observed primarily in citation forms and careful phonetic productions.

The goal of the present paper is to evaluate several hypotheses related to the production of constricted laryngeal states, such as creaky voice and glottal stop: that (1) the inferior surface of the ventricular folds presses into the vocal folds; (2) the action of the epilarynx diminishes proportionately with increasing speech rate; and (3) perturbation of vibratory dynamics of the vocal folds is not solely attributable to action at the vocal fold level. To address these hypotheses, two imaging modalities, magnetic resonance imaging (MRI) and high-speed laryngoscopy (HSL), were used to make observations of the larynx during performances of the sounds in question. The objective was to evaluate (1) with MRI to observe the three-dimensional relationship among the laryngeal structures and to evaluate (2) and (3) using the HSL.

2. METHODOLOGY

2.1. Magnetic resonance imaging study

MRI of a trained-phonetician participant was obtained to evaluate the geometrical relationships amongst the laryngeal structures during constricted laryngeal states. The states examined were modal voice (for control), creaky voice, glottal stop, and aryepiglotto-epiglottal stop, all of which were sustained for 8 s (the scanning duration) with the tongue positioned as in [i]. The imaging consisted of independently acquired 2D multi-slice axial and sagittal magnetic resonance imaging. Acquisition was done at the *Hôpital européen Georges-Pompidou (HEGP) in Paris on a Discovery MR750W 3.0T scanner (GE Medical Systems) under the guidance of Prof. Philippe Halimi, Chef de service, Radiologie, HEGP. A GEM head-and-neck coil was used. We used a Fast SPGR 3D sequence (TR = 5.3 mm, TE = 2.1 mm, FOV = 259.9 mm × 259.9 mm, flip angle = 12°). Each sequence produced 56 serial sections with a slice thickness of 2 mm and a 1 mm interslice distance. Resolution of the resulting images was 256 × 256, giving a pixel scaling of 1.0156 mm × 1.0156 mm.

No audio signal could be obtained, but auditory priming prior to each production was given to the participant by another phonetician via the PA system. This helped to ensure the productions were consistent and canonically central to the intended auditory-phonetic category. Beyond this, it is not possible to offer any further guarantee that the productions did not stray from the intended targets, and the results should be interpreted with this in mind.

2.1. High-speed laryngoscopic video study

High-speed laryngoscopic video was obtained to evaluate vocal fold dynamics in relation to laryngeal posturing during constricted laryngeal states. The session was used to examine a wide range of different sounds as produced by two different participants, but, for brevity, only two sets of data will be considered here.

The first of these is the continuous iteration of an [iʔ] syllable, i.e. [iʔiʔiʔi...], produced at different speech rates; these were produced by the same participant who participated in the MRI. Another participant produced a series of strong ejectives in [i] context with oral stricture formed around the barrel of the rigid endoscope, giving, approximately, a series of [ip’i]. These are of interest since at the onset of the closure phase of the [p’] a glottal stop occurs, and thus these data were selected for presentation here as they help clarify the contributions of the vocal folds and the epilarynx.

The laryngoscopy was performed at the HEGP under the supervision of Dr. Lise Crevier-Buchman, ORL-Phoniatre, Unité de la Voix, de la Parole et de la Déglutition, using the HreS Endocam 5562 system (Richard Wolf), which acquired frames at 2000 fps and also simultaneously captured electroglottographic (EGG) and audio signals.

3. RESULTS

3.1. Analysis of MRI

Figs. 1 and Fig. 2 respectively show montages of the sagittal and axial serial sections selected uniformly at 2 mm intervals over regions of interest centred on the larynx. For each figure, the rows show, from top to bottom, the four production targets examined: modal voice, creaky voice, glottal stop [ʔ] (GS), and aryepiglotto-epiglottal stop [ʡ] (ES), all in [i]-context.

The MR images should be regarded as time-averaged representations of the true productions. Although acquisition time was relatively short (8 s), slight changes to posture likely occurred and vocal fold vibration (in the modal and creaky productions) increased the noise in the image around the area of the vocal folds. Note that air spaces appear as dark regions (where water content was likely relatively very low); fatty tissues appear as a light grey.

In both figures, the three productions can be placed on a continuum of laryngeal constriction, increasing from modal voice through creaky voice, glottal stop and peaking with aryepiglotto-epiglottal stop. While the degree of AP narrowing of the epilarynx appears to be high in the constricted productions (arrows 1), it does not reach full closure as in aryepiglotto-epiglottal stop. This being said, in glottal stop, it might be that more AP narrowing had occurred in these productions than what is typical for this speaker (judging from dynamic productions in prior laryngoscopy work). This may have resulted from hyperarticulation and/or breath-holding necessary to sustain the glottal stop for the 8 s scanning duration or may be due to the supine MRI position. Comparison with aryepiglotto-epiglottal stop confirmed that, despite this possible hyperarticulation, the productions are still distinguishable in the MRI and thus can be taken as instances of their intended phonetic categories.

Modal voice (top row) is the control case for evaluating whether or not the hypothesized vocal-ventricular contact occurs as a function of laryngeal constriction. In the sagittal view, the ventricle (v) is visible in modal voice for almost the entire lateral-medial extent sampled, although the ventricle is wider towards the back. In the axial view, the vocal folds (tf) appear above the subglottal space (ss), but in modal voice, at higher sections, the ventricle is manifest and its greater posterior width agrees with the sagittal view. Creaky voice and glottal stop, however, both lack any sign of this lateral ventricular space, although there is more medial patency of the larynx in creaky voice than glottal stop. Furthermore, the anterior surface of the vocal folds is in contact with the ventricular folds (ff) just lateral to the midline, which is especially evident by comparing the modal and creaky voice productions (arrows 2). Thus, it is probable that, in both creaky voice and glottal stop productions, contact between the vocal folds and the ventricular folds had occurred. Also, the larynx is positioned higher by about 2 to 3 mm (consistent with [8]) in these productions relative to modal voice.

Figure 1: Sagittal MRI sections: modal voice, creaky voice, [ʔ] (GS), and [ʡ] (ES); [i]-context. e = epiglottis; ff = ventricular fold; ss = subglottal space; t = tongue; tf = vocal (true) fold; v = ventricle. Left = front. Red box = midsagittal.

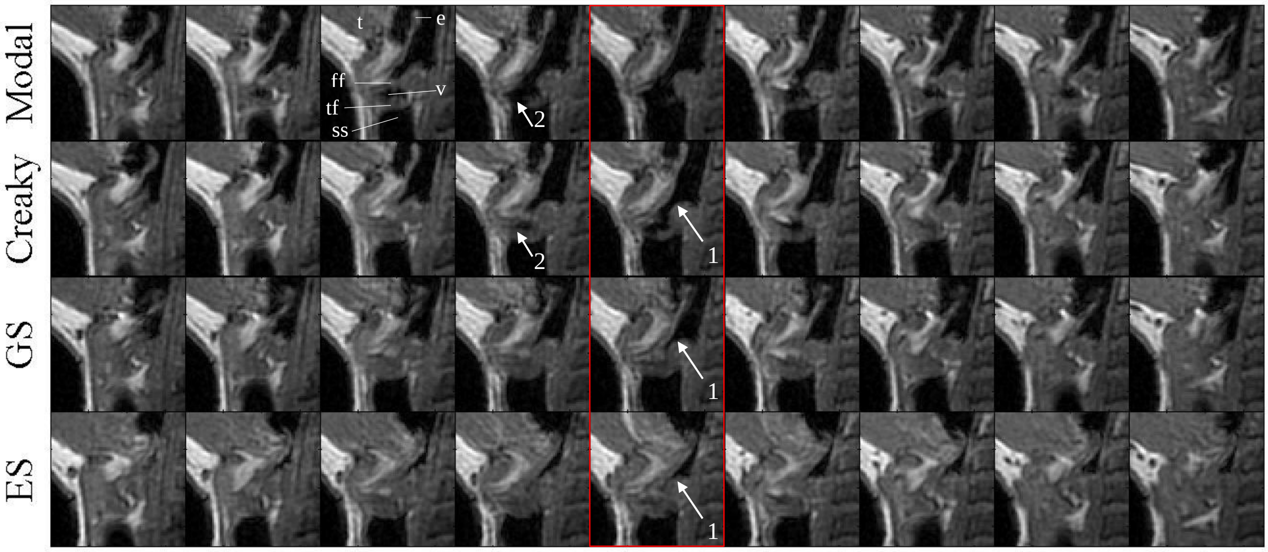
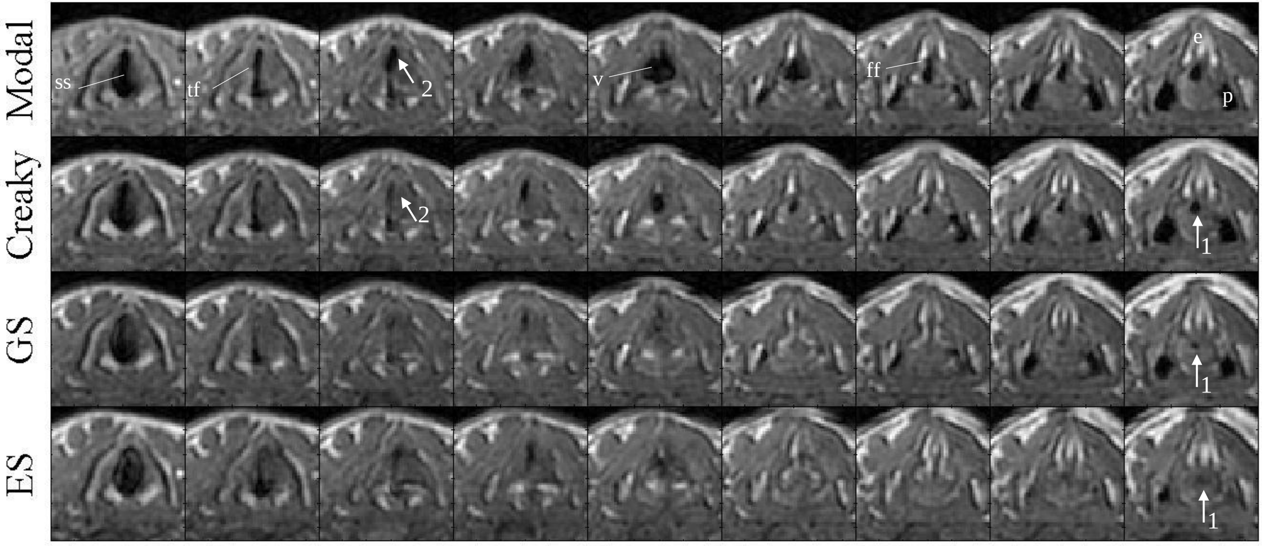


Figure 2: Axial MRI sections: modal voice, creaky voice, [ʔ] (GS), and [ʡ] (ES); all in [i]-context. p = piriform fossa. See Fig. 1 for other abbreviations. Top = front.



Given the short height of the ventricle (for males, 2.77 mm on average in normal phonation, see [1]), an increase in laryngeal height of this magnitude would make contact between the two sets of folds more likely. (Also note that the axial sections have been aligned, but it was observed that the larynx was lower in the modal voice production than in the constricted ones by about 2-3 mm.)

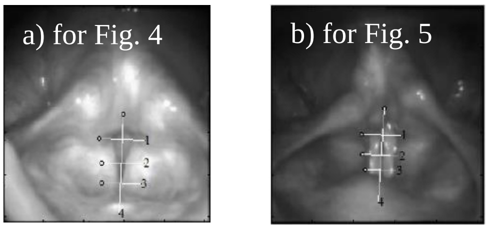
3.2. Analysis of the high-speed laryngoscopy

A kymographic analysis was performed on the HSL data; the locations of the kymographic lines are in Fig. 3. As discussed, one participant performed iterated glottal stops in [i] context at different rates; Fig. 4 shows the fast rate ([ʔ] duration ≈ 40 ms).

A different participant performed ejectives: the phase preceding the sudden larynx raising, which is comparable to the closure in [ʔ], is shown in Fig. 5.

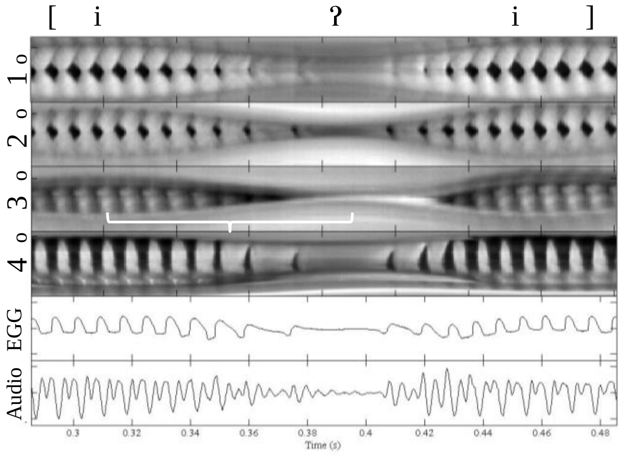
Ventricular adduction was observed to occur even at the fast rate (Fig. 4), although the degree of medialization is clearly a function of rate (the lower the rate, the more medialization). On some iterations of the target [ʔ], and especially at faster rates, vocal fold vibration diminished considerably but did not stop completely, giving a phase of creakiness rather than a true stop (although the perceptual effect was one of a stop). At the fastest rate achievable by the participant ([ʔ] duration ≈ 20 ms, not shown here), the ventricular adduction was no longer appreciable, but it is also the case that vibration was continuous and did not show a substantial diminishment in vibratory intensity.

Figure 3: Kymographic (“kymo”) lines for the analyses shown in (a) Fig. 4, and (b) Fig. 5. Circles indicate orientation of the kymographic lines.



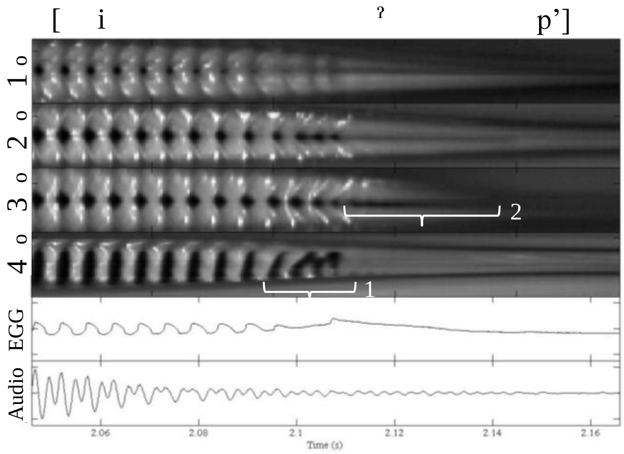
The degree of medialization of the arytenoids (as indicated by the corniculate tubercles) did not change appreciably (the images in Fig. 3 are representative of the degree observed throughout the videos). There was some adductory motion of the posterior vocal folds suggesting increased medial compression in the region of the cartilaginous glottis (kymo-1). The majority of the motion observed was an adductory motion of the ventricular folds accompanied by AP narrowing (kymo-4).

Figure 4: Kymographic data for video of fast glottal stop production. See Fig. 3a for the numbered interpretation of kymographic lines.



The transition in vibratory pattern from the modal flanking-vowel phase to the glottal stop always occurred roughly halfway between peak ventricular fold adduction and abduction (brace in Fig. 4). The transition into vibratory arrest is quite sudden but takes several pulses to decay fully. The behaviour in Fig. 5 (brace 1) is noteworthy since not only was a sudden change in vibration pattern observed, but the vocal folds did not close and instead maintained a slight separation (brace 2).

Figure 5: Kymographic analysis of video of laryngeal movement preceding an ejective. See Fig. 3b for interpretation of kymographic lines.



7. DISCUSSION & CONCLUSION

The evidence leads to the following conclusions regarding the three hypotheses concerning the role of epilaryngeal constriction in glottal stop (and similar phonetic gestures requiring laryngeal constriction) *for these data in particular*. (1) vocal-ventricular fold contact can occur during creaky voice (supporting earlier observations [6]) and glottal stop; however, the productions examined, especially the glottal stop, may have been somewhat hyperarticulated because they were sustained for MRI acquisition. (2) while epilaryngeal action does diminish at increased speech rates, it does not do so by much (and it is the major movement observed laryngoscopically in glottal stop). Finally, (3) the kymographic evidence suggests that simple medial adduction is not enough to explain the perturbation to vibration in the transition into vibratory arrest for glottal stop. With regard to hypothesis (3), we might indeed suspect that the sudden change in vibratory behaviour we observed arises from a sudden contact between the vocal and ventricular folds as the latter descend upon the former. In the case of the ejective, the fact that the vocal fold vibration was disturbed but the vocal folds remained slightly open (insufficient medial compression), supports the idea that an epilaryngeal mechanism is key to vibratory arrest in this case. Specifically, we suspect that, in contacting the vocal folds, the ventricular folds perturb the transmission of the mucosal wave of the vocal folds, resulting in the perturbed vibratory behaviour observed in creaky voice and aiding (possibly even enabling) the arrest of vibration altogether during glottal stop.

7. REFERENCES

[1] Agarwal, M., Scherer, R. C., Hollien, H. 2003. The false vocal folds: Shape and size in frontal view during phonation based on laminagraphic tracings. *J. Voice* 17(2), 97-113.

[2] Brunelle, M., Nguyễn, D. D., & Nguyễn, K. H. 2010. A laryngographic and laryngoscopic study of Northern Vietnamese tones. *Phonetica* 67(3), 147-169.

[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(2), 157-191.

[4] Garellek, M. 2013. Production and perception of glottal stops. PhD dissertation, UCLA.

[5] Halle, M., & Stevens, K. N. 1971. A note on laryngeal features. MIT *QPR* 101, 198-212.

[6] Hollien, H. 1972. Three major vocal registers: A proposal. *Proc. 7th ICPhS* Montréal, 320-331.

[7] Lindqvist-Gauffin, J. A. 1972. A descriptive model of laryngeal articulation in speech. *STL-QPSR* 13(2-3), 1-9.

[8] Moisik, S. R. 2013. The epilarynx in speech. PhD dissertation, University of Victoria.

[9] Moisik, S. R., Esling, J. H. 2014. Modeling the biomechanical influence of epilaryngeal stricture on the vocal folds: A low-dimensional model of vocal–ventricular fold coupling. *JSLHR* 57(2), S687-S704.