

PUTTING THE LARYNX IN THE VOWEL SPACE: STUDYING LARYNX STATE ACROSS VOWEL QUALITY USING MRI

Scott R. Moisik¹, John H. Esling², Lise Crevier-Buchman³, & Philippe Halimi⁴

¹Nanyang Technological University, Singapore

²University of Victoria, Victoria, BC, Canada

³CNRS-UMR7018, Paris III, Hôpital Foch, ⁴HEGP, Paris, France

scott.moisik@ntu.edu.sg; esling@uvic.ca; lise.buchman1@gmail.com; philippe.halimi@egp.aphp.fr

ABSTRACT

This study addresses how vowel quality interacts with larynx state. Magnetic Resonance Imaging (MRI) data of the larynx during the articulation of phonetically produced vowels under different voice and phonatory quality conditions – including modal and creaky phonatory qualities, glottal and epiglottal stop, and raised larynx voice (RLV) pharyngealized quality – were acquired for two trained phoneticians. The analysis consists of qualitative observations of laryngeal tissues and quantitative analysis of anatomical landmarks. The results indicate that vocal-ventricular fold contact occurs in the context of creaky voice and glottal and epiglottal stop, but the contact pattern is complex with a residual medial air space. The larynx is low for [a, o, u]; front vowels have typically higher larynx height. Lingual articulation in RLV resembles a ‘double bunched’ [ɹ], while the larynx position is both advanced and raised.

Keywords: vowel quality, laryngeal articulation, epilarynx, laryngeal ventricle.

1. INTRODUCTION

This study addresses how vowel quality interacts with larynx state. It is known that larynx height varies in relation to vowel quality, but inconsistently so [6, 11]. Furthermore, some have suggested that the intrinsic f_0 of vowels arises from lingual-laryngeal interaction in producing vowels [17, 21], and others point to wider-reaching phonological patterns [20, 3, 15]. Despite evidence supporting the existence of such effects, uncertainty remains about the nature of the articulatory changes involved, in particular those in the lower vocal tract that concern the internal configuration of laryngeal tissues in response to vowel articulation or the relation between cricothyroid rotation and larynx height [10]. Furthermore, very little work has examined how the interactions play out under the influence of different voice (and phonatory) qualities, such as raised larynx voice (RLV) which is laryngeally constricted [12].

A general tendency is that vowel height tends to be inversely related to larynx height [18, 13, 6, 5]. As

summarized in Table 1, [u] seems to be universally lower in larynx height, while [i] and open vowels like [a] and [ɑ] vary in larynx height by language; [i] often being unexpectedly high.

Table 1: Vowel ranking, high to low larynx height.

Language	Ranking	Source
N. A. English	[æ, ɑ, i, ɪ, ε, ɔ, u]	[18], p. 40
S. Brit. English	[æ, ɑ, i, ɔ, u]	[22], pp. 26-27
Egyptian Arabic	[æ, ɑ, i, ɔ, u]	[22], pp. 26-27
English	[i, a, u]	[5], p. 27
French	[i, a/u]	[5], p. 28
Japanese, (YK)	[i/e, a, u/o]	[5], p. 29
Japanese, (SH)	[a, i, u]	[5], p. 30
Taiwanese	[i, a, u]	[5], p. 31
Mandarin	[a/e, i/y, u, o]	[5], p. 32
Vietnamese	[a, i, u]	[5], p. 33
Thai	[i/a, u]	[5], p. 34
English	[i/æ/e, ɑ/ʌ, o, u]	[5], p. 41
German	[i, a, u]	[5], p. 41

Ewan [5] also points out that the patterns are more complex, being both pitch and speaker dependent. Larynx height variation may result from competing action of hyoid (and larynx) lowering to aid jaw opening for open vowels, and, in [i], of hyoid advancement in tongue root articulation.

The goal of our study is to provide detailed imaging data of the vocal tract when producing different vowels under the influence of different laryngeal articulatory postures, of increasing epilaryngeal constriction, to ascertain how larynx height is balanced between serving vowel articulation and serving laryngeal articulation. As this task involves comparison of articulations not found to co-occur within a single language, we focus on careful phonetic productions of the target states.

2. METHODS

The data for this study come from two partially independent MRI sessions with two phonetician subjects, each with somewhat different objectives and spaced widely apart in time. Common to both is that different states of constriction of the laryngeal articulator [3, 15] (with or without a phonation type)

were sustained in a given vocalic context. What follows is a description of each MRI session.

The first session features data from JHE, a trained phonetician producing [i, e, æ, a, o, u] across different laryngeal states. The states examined were modal voice (for control), glottal stop, and (aryepiglottal)epiglottal stop, all sustained coarticulatorily for 8 s. Imaging consisted of 2D multi-slice sagittal sequences. 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 spoiled gradient recalled echo (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. No audio was collected during this session. The participant did not have auditory-priming assistance.

The second session features data from SRM, also a trained phonetician. This session was designed to obtain a wider range of vowel qualities than the first session (but including those vowels as a subset) in the context of constricted laryngeal states with phonation. The states examined were modal voice (for control), creaky voice, and RLV (pharyngealized quality). The vowel qualities examined were [i, ɪ, e, ɛ, æ, a, ɔ, o, ʊ, u]. Acquisition was at the Donders Institute for Brain, Cognition and Behaviour (in the Donders Centre for Cognitive Neuroimaging or DCCN) in Nijmegen (NL). We used an Avanto 1.5T MRI system (Siemens Healthcare) with six-channel head and neck RF receiver coil array (also by Siemens) and a T2-weighted turbo spin-echo sequence (acquisition time = 10 s, TE = 9.9 ms, TR = 380 ms, flip angle = 150°, slice thickness = 3 mm, field of view = 240 mm × 240 mm, voxel size = 3.0 mm × 0.94 mm × 0.94 mm), yielding only midsagittal slices. Audio was recorded, but we only analyse f_0 here. The participant had no auditory-priming support in the productions.

Analysis of the resulting images was both qualitative and quantitative. Qualitative analysis was facilitated by various means of visualizing the scans. The vocal tract was manually traced and two comparisons were made, following [15], one between the modal state and a moderately constricted state (either glottal stop or creaky voice) and another between the modal state and the heavily constricted state (either epiglottal stop or RLV).

For the first session only (JHE), volumetric visualizations were constructed using the marching cubes algorithm [8] on a region bounded inferiorly by the base of the cricoid lamina, anteriorly by the laryngeal prominence of the thyroid cartilage and

superiorly, obliquely, by a plane tangent to the apex of the epiglottis and the upper surface of the posterior cartilage complex formed by the cuneiforms, corniculates, and arytenoids.

Quantitative analysis was performed on a subset of linear and angular measures obtained from a landmarking analysis of the raw data (using a custom interface in MATLAB®). In this paper, we only consider cricoid height (a measure of larynx height obtained by comparing the average position of the base of C7 across vowel states within a single participant's data and the superior midline point of the cricoid lamina) and the posterior-anterior distance of the epilaryngeal tube (the smallest distance between its anterior and posterior margins), selected on the theoretical grounds that it is a key dimension of laryngeal articulator action [3, 4].

3. RESULTS

Qualitative results are given in Fig. 1-5. First in Fig. 1-3, are 3D visualizations of the laryngeal and hypopharyngeal airway of JHE (from a few centimetres below the vocal folds up to the upper epilaryngeal border). Fig. 1 provides an annotated view of the airways as they appear for the vowel [i] across the three voice qualities (modal, GS, and ES).

Figure 1. Lower airway renderings (JHE) for [i]; modal (top), glottal stop (GS; middle), epiglottal stop (ES; bottom). Views: side view (from left); 3/4 front-view; front view. Abbrev.: ea = apex of the epiglottis; ep = petiole of the epiglottis; et = epiglottic tubercle; ff = ventricular (false) fold; tf = vocal (true) fold; pf = piriform fossa; tr = trachea; vent = ventricle (space); vest = vestibule.

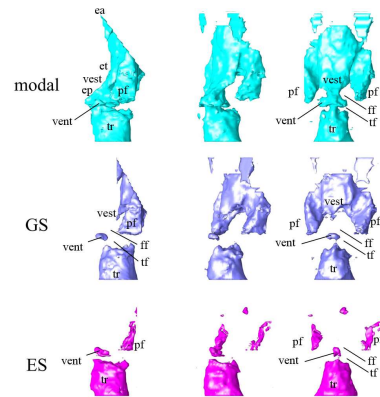


Fig. 2 and 3 present side and front views, respectively, of these renderings for all of the vowels produced by JHE. Fig. 2 illustrates the angle of the epiglottis as it rotates and flattens in the production of vowels with greater degrees of retraction (especially [æ] and [a]). One can also get an immediate impression of the variation in larynx height (in

absolute terms). In modal voice, [e] is highest while [a] and [u] are lowest. In GS and ES, [a] is lowest. Both Fig. 2 and 3 give a sense of how constriction changes the ventricle. It does not disappear entirely: a small anterior pocket remains across all vowels in both GS and ES. Despite this, the remainder of the mid-laryngeal airway has collapsed, which suggests that there is contact between the vocal folds and ventricular folds over their entire lateral-medial extent as one moves posteriorly. The vestibule collapses to an extreme degree in both GS and ES. A thin channel (vest) extends below the level of the tubercle of the epiglottis (et; side view in Fig. 1).

Figure 2: Lower airway (JHE), side views.

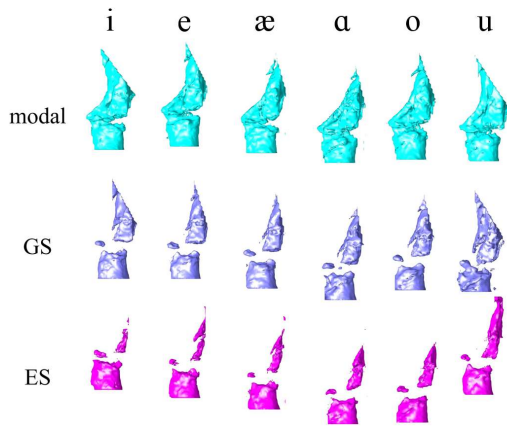
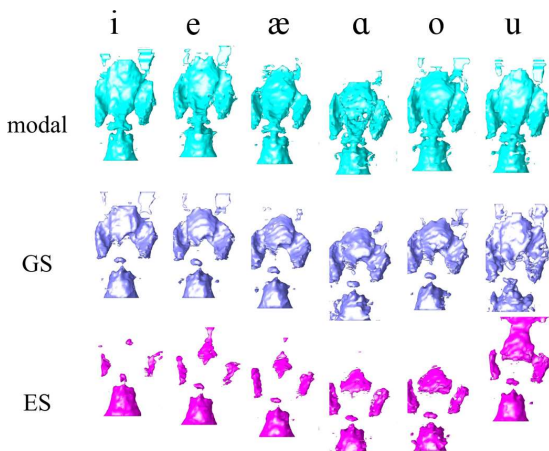


Figure 3: Lower airway (JHE), front views.



Near-mid-sagittal sections appear in Fig. 4, for JHE, and in Fig. 5, for SRM. In Fig. 4, the tongue appears to make contact with the posterior pharyngeal wall (arrow 1) – and indeed it does – but, we can ascertain from parallel scans that this does not mean the airway is closed off (as for a linguo-pharyngeal stop). Rather, this is an artefact of a slightly non-orthogonal alignment of the participant’s head, making it difficult to capture both the airway at the

level of the tongue and the narrow ventricular patency (arrow 2, Fig. 4) at the level of the larynx. In fact, this linguo-pharyngeal contact occurs even in modal voice for the vowels [æ], [a], and [o] but not [i], [e], and [u], which is consistent with laryngoscopic observations [7] that suggest pharyngeal bracing for the former set of vowels.

Figure 4. MRI montage of vowels (JHE): modal (top) glottal stop (middle); epiglottal stop (bottom). Traces: modal voice (black outline) vs. GS (gray area; top) and vs. ES (gray area; bottom).

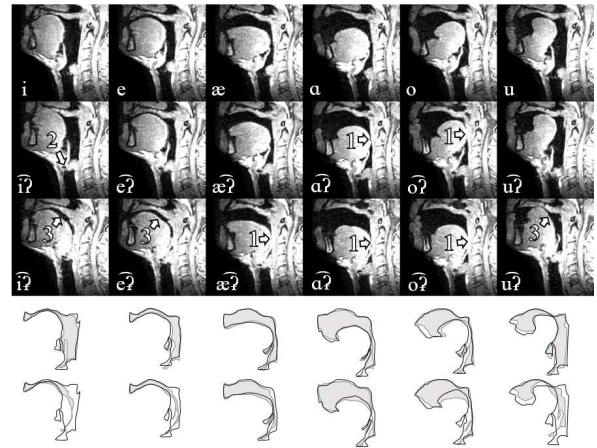
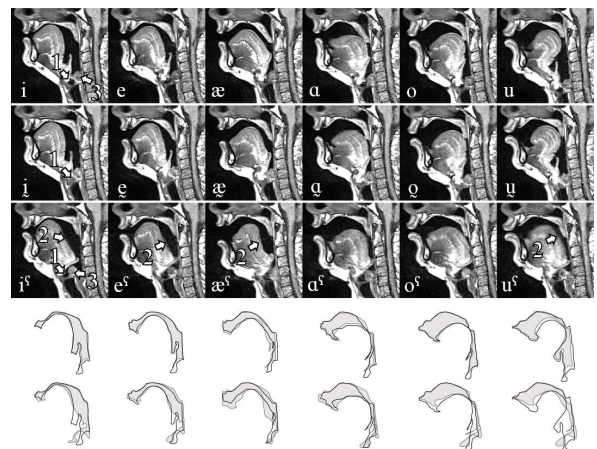


Figure 5. MRI montage of vowels (SRM): modal (top); creaky voice (middle); RLV (bottom). Traces: modal (black outline) vs. creaky voice (gray area; top) and vs. RLV (gray area; bottom).



Turning to Fig. 5, the positioning of the tongue between the modal and creaky voice series reveals that there is some slight lingual retraction in creaky voice, which makes the overall pharyngeal airway appear slightly reduced in volume. RLV shows a reconfiguration of the pharyngeal airway associated with ‘cinching’ of the tongue (arrow 2, Fig. 5; cf. arrow 3, Fig. 4); it also shows striking advancement of the larynx as a whole but most visible in the

anterior movement of the posterior structures (cf. arrow 3, Fig. 5).

We suspect that muscle fibres of the medial portion of the genioglossus muscle are pulling the dorsum anteriorly [19]. At the same time, the posterior tongue infringes the laryngopharyngeal space, displacing the epiglottis. The overall effect is that a significant portion of the pharyngeal airspace volume is shifted upwards, particularly notable for [i^s], [e^s], and [u^s]. This distortion of tongue shape, which bears a resemblance to the type-4 variety of American English /r/ [2] and the ‘double bunching’ identified by Catford [1], parallels the lingual shapes observed for ES.

The ventricle seems to be obliterated in creaky voice (cf. arrow 1, Fig. 5), but it must have been at least slightly open (otherwise voicing would not be possible). The apparent closure probably reflects loss of the lateral portions of the ventricle caused by contact between the vocal and ventricular folds [16].

Figure 6. Laryngeal vowel space (cricoid height vs. postero-anterior epilaryngeal tube distance) for JHE (top) and SRM (bottom). Cricoid height: lower values = lower larynx height; epilarynx distance: lower values = narrower epilarynx.

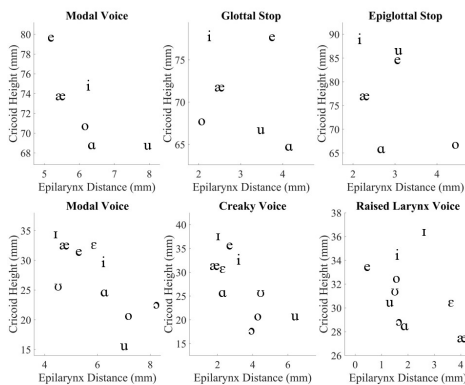


Fig. 6 shows possible dimensions of what could be called the ‘laryngeal vowel space’. The postero-anterior epilarynx distance is plotted against cricoid height. For SRM, mean f_0 was 87 Hz, 91 Hz, and 101 Hz for modal voice, creaky voice, and RLV, respectively. Spearman’s rank-order correlation tests were run to check the strength of association between f_0 and cricoid height, but none were significant (which probably more reflects being underpowered). Thus, these values are suggestive that the observed effects are due to larynx state and not pitch, but pitch cannot be definitively ruled out given the limited data.

4. DISCUSSION AND CONCLUSION

The data here reflect careful, canonical performances by two trained phoneticians, and interpretations

should be made with this in mind. As a few researchers have observed [6, 17, 22], larynx height is not independent of vowel quality but instead covaries with it. We add considerable detail in qualitative visualization and quantification of larynx height and postero-anterior epilarynx distance. This study also introduces the interaction effects that different types of laryngeal articulation produce – from nil (modal voice), to moderate (glottal stop and creaky voice), and extreme forms of constriction (epiglottal stop and raised larynx voice).

An important result of this work is the finding that during glottal stop, epiglottal stop, and creaky voice, there is manifest contact between the vocal folds and the ventricular folds. Such contact is also implied in [9], described by [14] and [12], and further confirmed in [16]. In our results here, although the posterior portion of the ventricular space is obliterated (for both participants), a medial-anterior patency of the ventricle was also evident. Creakiness and especially RLV quality preserve a supraglottal space within the larynx to allow for vibration, but with mainly perturbed vibratory quality in creakiness and mainly perturbed resonance in RLV. Our findings for larynx height are mixed for glottal stop and creaky voice. In addition to variation associated with vowel quality, it is possible that low pitch and increased constriction compete to determine laryngeal height. This differs from epiglottal stop, where epilaryngeal constriction seems to dominate in controlling larynx height over the other factors because the larynx is always raised in this state.

Although the vowels show a range of positions in the laryngeal vowel space across conditions and speakers, some consistencies hold. While larynx height appears to be inversely related to epilarynx distance (lower larynx, wider epilarynx), epilarynx distance is typically narrowest for [æ] – except, surprisingly, in RLV. The general pattern of larynx height across the vowels is more or less consistent with that observed in previous studies, with front vowels [i] and [a] (or [æ], in this study) being higher in larynx height than [u]. We can state further that front vowels are typically higher than vowels [a, o, u]. One reason for the lower larynx position in [a] may be that lowering the larynx, via chain-linkage with the hyoid bone, helps the tongue retract into the pharynx. If larynx height is not suppressed, then we suspect phonatory vibration would tend towards constricted types (e.g., creaky voice) or cease altogether. Thus, because of the internal laryngeal compaction brought about by tongue retraction, open vowels are at risk of perturbed phonatory quality if compensatory adjustment is not made by lowering the larynx. Generally we have shown that lower-vocal-tract states significantly influence vowel quality.

5. REFERENCES

- [1] Catford, J.C. 1983. Pharyngeal and laryngeal sounds in Caucasian languages. In: Bless, D.M., Abbs, J.H. (eds), *Vocal Fold Physiology: Contemporary Research and Clinical Issues*. San Diego: College Hill Press, 344–350.
- [2] Delattre, P., Freeman, D.C. 1968. A dialect study of American R's by x-ray motion picture. *Linguistics* 6, 44, 29–68.
- [3] Esling, J.H. 2005. There are no back vowels: The laryngeal articulator model. *Canadian Journal of Linguistics* 50, 13–44.
- [4] Esling, J.H., Moisić, S.R., Benner, A. and Crevier-Buchman, L. 2019. *Voice quality: The laryngeal articulator model*. Cambridge University Press.
- [5] Ewan, W.G. 1979. Laryngeal behaviour in speech. *Report of the Phonology Laboratory 3*. Berkeley.
- [6] Ewan, W.G., Krones, R. 1974. Measuring larynx movement using the thyroumbrometer. *Journal of Phonetics* 2, 327–335.
- [7] Gauffin, J., Sundberg, J. 1978. Pharyngeal constrictions. *Phonetica* 35, 157–168.
- [8] Hammer, P. 2011. *Marching Cubes*. MathWorks File Exchange:
<https://www.mathworks.com/matlabcentral/fileexchange/32506-marching-cubes>.
- [9] Hollien, H. 1972. Three major vocal registers: A proposal. *Proc. 7th ICPhS Montreal*, 320–331.
- [10] Honda, K., Hirai, H., Masaki, S., Shimada, Y. 1999. Role of vertical larynx movement and cervical lordosis in f0 control. *Language and Speech* 42, 4, 401–411.
- [11] Ladefoged, P., DeClerk, J., Lindau, M., Papçun, G. 1972. An auditory-motor theory of speech production. *UCLA Working Papers in Phonetics* 22, 48–75.
- [12] Laver, J. 1980. *The Phonetic Description of Voice Quality*. Cambridge University Press.
- [13] Lindau, M. 1975. Features for vowels. *UCLA Working Papers in Phonetics* 30, 1–55.
- [14] Lindqvist-Gauffin, J. 1972. A descriptive model of laryngeal articulation in speech. *Quarterly Progress and Status Report, Speech Transmission Laboratory, Royal Institute of Technology* 13, 2–3, 1–9
DOI:https://doi.org/http://www.speech.kth.se/prod/publications/files/qpsr/1972/1972_13_2-3_001-009.pdf.
- [15] Moisić, S.R. 2013. *The Epilarynx in Speech*. Doctoral dissertation, University of Victoria.
- [16] Moisić, S.R., Esling, J.H., Crevier-Buchman, L., Amelot, A., Halimi, P. 2015. Multimodal imaging of glottal stop and creaky voice: Evaluating the role of epilaryngeal constriction. *Proc. 18th ICPhS Glasgow*, paper 247.
- [17] Ohala, J.J. 1987. Explaining the intrinsic pitch of vowels. In Channon, R., Shockey, L. (eds), *In honor of Ilse Lehiste*. Dordrecht: Foris, 207–215.
- [18] Perkell, J.S. 1969. *Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study*. MIT Research Monograph No. 53.
- [19] Takano, S., Honda, K. 2007. An MRI analysis of the extrinsic tongue muscles during vowel production. *Speech Communication* 49, 1, 49–58
DOI:<https://doi.org/10.1016/j.specom.2006.09.004>.
- [20] Trigo, L. 1991. On pharynx-larynx interactions. *Phonology* 8, 1, 113–136
DOI:<https://doi.org/10.1017/S0952675700001299>.
- [21] Whalen, D.H., Gick, B., Kumada, M., Honda, K. 1998. Cricothyroid activity in high and low vowels: Exploring the automaticity of intrinsic F0. *Journal of Phonetics* 27, 125–142.
- [22] Wood, S. 1979. A radiographic analysis of constriction location for vowels. *Journal of Phonetics* 7, 25–43.