Part 3: MELODY – INTONATION AND PITCH

Część 3: MELODIA – INTONACJA I WYSOKOŚĆ TONU

Laryngeal aperture in relation to larynx height change: an analysis using simultaneous laryngoscopy and laryngeal ultrasound

Analiza regulacji otwarcia krtani względem jej wysokości z wykorzystaniem jednoczesnej laryngoskopii i ultradźwiękowego obrazowania krtani

John H. Esling and Scott R. Moisik

University of Victoria
Victoria, British Columbia, Canada
esling@uvic.ca srmoisik@uvic.ca

ABSTRACT

We present several case studies of laryngeal and pharyngeal articulations using simultaneous laryngoscopy and laryngeal ultrasound (SLLUS). We provide qualitative data of laryngeal state from the laryngeal ultrasound and laryngoscopic video; we also provide data for larynx height, which is quantified by means of optical flow analysis.

STRESZCZENIE

Przedstawiamy kilka analiz przypadku artykulacji krtaniowej i gardłowej z wykorzystaniem jednoczesnej laryngoskopii i ultradźwiękowego obrazowania krtani (ang. SLLUS). W oparciu o ultradźwiękowe obrazowanie krtani i obraz wideo uzyskany na podstawie laryngoskopii dostarczamy danych jakościowych dotyczących stanu krtani; przedstawiamy również dane związane z wysokością krtani, określone ilościowo za pomocą przepływu optycznego.

1. Overview

Wiktor Jassem's work has contributed in numerous areas of phonetics, but most particularly in the measurement of the articulatory attributes of stress and intonation. As with James (Tony) Anthony's work on breathing as a concomitant of speaking gestures, we take inspiration from Wiktor's work on pitch in our descriptions of the voice quality component of speech prosody. We present evidence, using a variety of instrumental phonetic techniques, of parallel operations within the larynx which go beyond the function of the vocal folds, implicating the coupling of higher valves of the larynx, integrating the role of the epilaryngeal tube, and demonstrating how the pitch of the voice is related to voice quality, incorporating such processes as phonation type, states of the larynx, laryngeal constriction, and aerodynamic flow. The acknowledgement of what the relationships could be between the concepts of pitch and voice 'quality' appears in Jassem [1].

2. Laryngeal Articulatory Axes: Height, Constriction, and Pitch

In addition to glottal aperture control, there are two principal articulatory axes of the larynx: the axis of height and the axis of constriction. The former is causally associated with changing the volume of the pharyngeal cavity; the latter denotes action of the mechanism that induces stricture of the epilaryngeal tube [2]. Both mechanisms have an impact on the pitch control system. Larynx height positively correlates with pitch: larynx lowering results in cricoid rotation favouring reduction of vocal fold tension [3]; larynx raising increases vocal fold tension through anterior traction on the hyoid-thyroid complex and through thyrohyoid contraction in conjunction with contributions from other factors such as tracheal pull [4, 5]. We assume that larynx height is relatively slow acting compared to the longitudinal pitch-control mechanism. Laryngeal constriction is a function of three primary physiological components that act synergistically [6]: tongue retraction (which causes posterior displacement of the epiglottis), larynx raising, and contraction of the thyroarytenoid muscle complex and possibly the aryepiglottic muscle [7]. Pitch is thought to be impacted by concomitant ventricular incursion of epilaryngeal stricture [8]: the ventricular folds impinge upon the upper surfaces of the vocal folds and mechanically couple with them; the increase in oscillating mass results in lowered frequency response of vocal fold vibration.

The system of laryngeal articulation is complicated by the interdependencies amongst the axes of height, constriction and the primary pitch-control mechanism, i.e. the cricothyroid muscles. Figure 1 provides a depiction of these relationships; the arrows indicate the direction of causality. Larynx height plays a dual role in laryngeal control: it acts agonistically in the mechanisms for pitch control and for laryngeal constriction. As discussed in [9], the primary pitch control mechanism stands in antagonistic relationship to the laryngeal constriction mechanism; this is because engagement of the laryngeal constrictor acts along the postero-anterior dimension to narrow the epilaryngeal tube, this is opposite to the widening of the laryngeal aperture caused by engagement of the cricothyroid muscles.

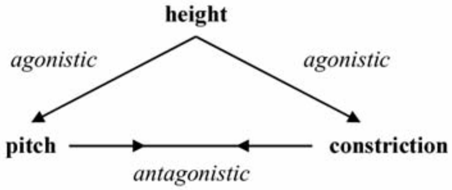

Figure 1: The relationship amongst larynx height, constriction, and pitch.

This study examines the temporal relationship between the height and constriction axes of laryngeal control in the context of a variety of sounds that depend upon laryngeal constriction for their realization. The purpose is to gain deeper insight into the physiological constraints governing the articulatory capacity of the larynx and its attendant effects on vocal fold dynamics.

3. Methods for Observing Laryngeal and Pharyngeal Gestures

We used simultaneous laryngoscopy and laryngeal ultrasound (SLLUS) to obtain empirical data on the temporal sequencing of larynx height and the other components of laryngeal constriction. Since this study focuses on the physiological nature of phonetic production as opposed to language specific production strategies, our data consist of careful phonetic production of cardinal phonetic categories by two participants (JE and SM). Specifically, we have selected glottal stop, creaky voice, and two different manners of pharyngeal articulation (stop and voiceless trill). Two types of data are considered: qualitative observations of laryngeal state based on the laryngeal ultrasound (LUS) and laryngoscopy (LS) using selected frames and quantification of larynx height obtained using an optical flow algorithm to calculate the vertical component of larynx velocity.

3.1. Simultaneous Laryngoscopy and Laryngeal Ultrasound (SLLUS)

To conduct the SLLUS technique, a standard laryngoscopic examination was performed, and the attending physician was seated in front of the participant, while the laryngeal ultrasound examination was performed simultaneously by approaching the participant from the side. The laryngeal ultrasound provides a peri-coronal view of the larynx by placing the ultrasound probe over the participant's thyroid lamina, approximately 1 cm 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 (in units of *cm*) was used to determine the pixel-to-mm ratio, which is necessary for measuring the velocity of the larynx (in the optical flow analysis). For more details concerning the use of laryngeal ultrasound for speech research see [10].

The laryngoscopy equipment is an Olympus ENF-P3 flexible fibreoptic nasal laryngoscope or a Kay 9106 rigid endoscope 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 a 12L-RS probe (General Electric Corporation). The probe pulse frequency was 10 MHz, which allowed for optimal resolution of laryngeal structures. Two audio signals were also recorded (at 44100 Hz, 16 bit) for the purposes of signal synchronization and analysis. 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, carefully aligned through manual inspection, and segmented into isolated tokens using Sony Vegas. F0 traces were obtained using the STRAIGHT algorithm [12]; it should be noted that STRAIGHT failed to register the creaky voice sections of our recordings.

3.2. Optical flow analysis

Change in larynx height was observed using laryngeal ultrasound and quantified by means of an optical flow algorithm based on a block-wise, absolute differences method. Gradient methods for optical flow were avoided since ultrasound data do not meet assumptions of smoothness in the brightness pattern of an image sequence [11]. The optical flow field is calculated as follows. Each frame from the video sequence is broken down into an analysis grid. For each node in the grid, an analysis block centered about the grid node is obtained from the current frame, and from the next frame an expanded region of pixels three times the size of the analysis block is obtained. To obtain the flow vector for this frame pair, the current pixel block is ‘swept’ across the next pixel block; the algorithm is in some respect like a two-dimensional convolution of images, but takes the sum of the absolute differences between the current and next pixel blocks rather than their products. The formula for calculating this is

$$D := \sum_{j=1}^{m-k} \sum_{i=1}^{m-k} |N_{i:i+k,j:j+k} - C|$$

where C is a $k \times k$ matrix for the pixel block of the current frame, N is an $m \times m$ matrix for the expanded pixel block of the next frame, and D is an $m\text{–}k \times m\text{–}k$ of the resulting sum of matrix differences[1]. The indices of the global minimum in D (with the center entry defined as the origin of the index-coordinate space) are taken to reflect the vertical and horizontal components of object displacement vector imaged by the pair of frames. The negative of D is thresholded at 95% of its brightness range producing $\mathrm{D}_{\mathrm{thresh}}$; the ratio of ones to zeros in $\mathrm{D}_{\mathrm{thresh}}$ defines a weighting σ that represents the accuracy of the result (smaller ratios indicate greater accuracy). Displacements in units of pixels are then converted to velocity by dividing by the frame period and applying the pixel-to-mm scaling. The optical flow field for a pair of frames constitutes the set of all velocity vectors obtained over the field defined by the analysis grid. A video sequence of f frames will yield f–1 velocity fields.

Vertical larynx displacement is calculated as follows. For each velocity field, the weighted average of all its vertical components is obtained using the σ values as weights. The results constitute a discrete representation of the vertical velocity; this velocity function must then be numerically integrated (using the trapezoidal method) to obtain a function for change in larynx height over time.

Validation of the optical flow algorithm was conducted on control data of a metal bar sliding 11.14 cm along a ruler. The velocity function was calculated with manual measurements and with the algorithm. The normalized RMS error between these two velocity functions is 12.17% and numerical integration of the velocity data obtained from the algorithm yields 11.35 cm, for an error of 1.8%, which we take as an acceptable level of error relative to a manual analysis.

3.3. Qualitative analysis of laryngoscopy

In this study, laryngoscopy data are qualitatively evaluated with reference to four visual indicators of laryngeal constriction (cf. [7] and [9]). Figure 2 presents three of these indicators: (i) postero-anterior dimension of the epilaryngeal tube, (ii) the positioning of the ventricular folds, and (iii) the angle of the aryepiglottic folds in the axial plane relative to the glottal midline. The fourth indicator is the elevation of the larynx. In laryngoscopy, changes to larynx height appear as a change in scale of the laryngeal structures; if the larynx elevates, it generally appears larger and vice versa for larynx lowering [13].

---

[1] Note that for notational convenience, the ‘:’ is used to define submatrices of N and the vertical bar notation indicates entry-wise absolute value of the difference matrix, not its determinant.

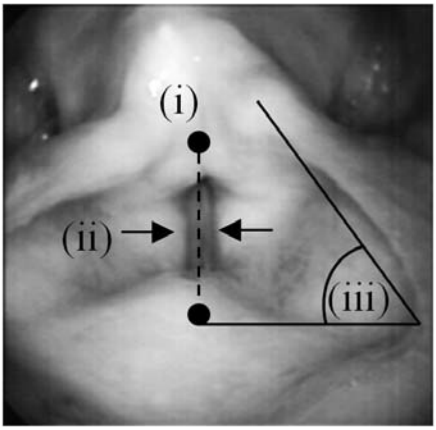

Figure 2: Indicators of laryngeal constriction.

4. Results

Figure 3 (also see [10]) shows two LUS frames selected from an [iʔi] sequence. As images (a) and (b) reveal, the ventricular fold impinges upon upper surface of the vocal fold during the glottal stop; this is part of the gesture referred to as *ventricular incursion* [2], which constitutes an articulatory component of laryngeal constriction. This vertical dimension of ventricular incursion is not easily observed; there are only scattered speculations and a few studies that report similar findings of vertical vocal-ventricular contact [8, 14, 15]. Since registration of the laryngeal structures in ultrasound is generally poor, the task of interpreting the images involves some estimation of the exact structural contours. Nevertheless, the video data are strongly suggestive that vocal ventricular contact does indeed occur.

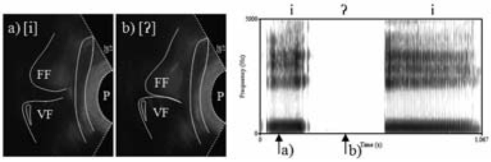

Figure 3: Coronal LUS view of ventricular incursion during glottal stop.
FF = false (ventricular) folds; VF = vocal folds; P = probe.

An illustration of the SLLUS data[2] for glottal stop produced by SM is in Figure 4. Frames 24, 28, and 32 show the progressive adduction of the ventricular folds during the glottal stop; some anticipatory adduction is evident during the vowel (frame 24) and full adduction is attained early during the onset phase. The adduction evidently continues to increase in degree of medial compression (frame 28) until offset occurs

---

[2] The audio data show a high noise floor; this is due to the laryngoscopy and ultrasound equipment.

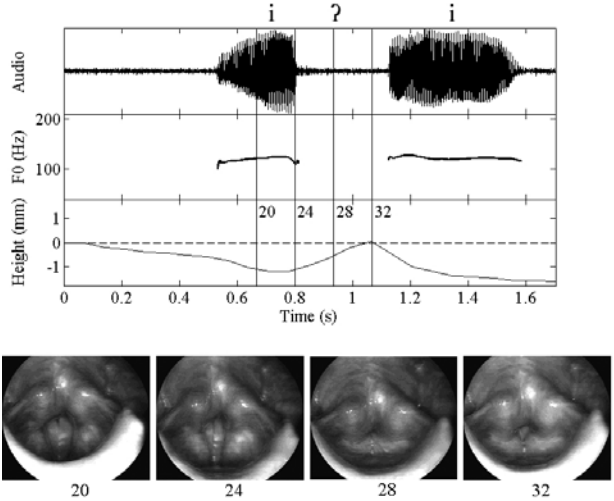

Figure 4: SLLUS data for glottal stop.

where release of adduction must take place in anticipation of the upcoming vowel. During the adduction, there is a reduction in the postero-anterior dimension of the epilaryngeal tube which correlates with a slight rise in larynx height. The change is on the order of 1.24 mm and occurs in about 0.275 s.

Figure 5 illustrates the SLLUS data for creaky phonation produced by SM, which was gradually engaged during the midpoint of an extended [i] sequence. The larynx raises by 3.38 mm over the course of 1.100 s. The spectrogram is included in this image to show how F2 exhibits arc-like raising (dashed-underline) in correspondence with the larynx raising[3]. Note that the raising can be discerned in the laryngoscopy frames as a slight enlargement of the appearance of the structures. Ventricular incursion once again is present, although complete adduction of the ventricular folds does not occur: throughout the creaky section of the performance there is a posterior gap at the ventricular level. Postero-anterior narrowing of the epilaryngeal tube is comparable with that which occurs for the glottal stop.

The production in Figure 5 can be compared with that in Figure 6, which represents an attempt to produce creaky phonation while suppressing the larynx raising and laryngeal constriction (again, produced by SM). While it was found to be possible to produce creak in this lowered larynx state without any obvious ventricular adduction, there is a clear state change that occurs with the onset of creakiness at frame 50. There is a very slight elevation of the larynx at this point which is accompanied by a reduction of the postero-anterior dimension of the epilaryngeal tube; this continues to increase in degree throughout the creaky region (frame 65). At this point, the vocal folds abruptly switch vibratory mode from one which allows for loose vibration along their entire longitudinal extent, to one allowing only for limited lateral displacements confined to a short section along their longitudinal extent (compare frames 20 and 35 with 50 and 65).

---

[3] The target vowel for these sequences was [i], but the quality tends to be centralized because of the presence of the laryngoscope in the oral cavity.

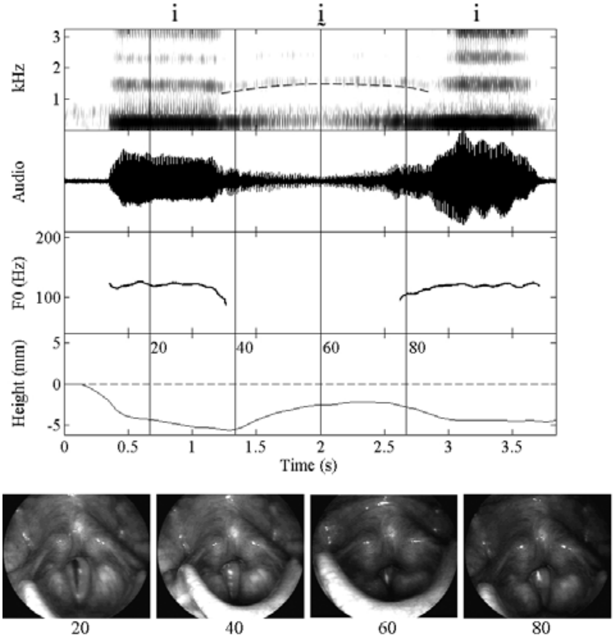

Figure 5: SLLUS data for creaky voice without height suppression.

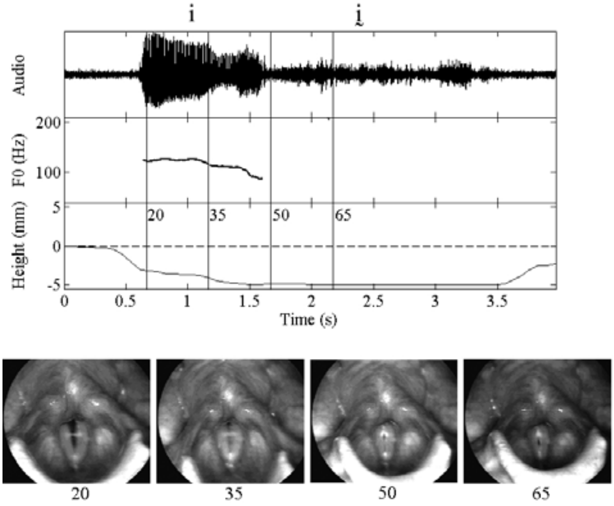

Figure 6: SLLUS of creaky voice with height suppression.

The next two case studies (produced by JE) demonstrate the relationship between larynx height and laryngeal constriction in the context of sounds that are traditionally labeled pharyngeal consonants. We present a pharyngeal stop— or more accurately an *aryepiglotto-epiglottal stop*—in Figure 7 and an pharyngeal or aryepiglotto-epiglottal

(AE) fricative-trill in Figure 8. The aryepiglotto-epiglottal narrowing illustrated in these studies is a hallmark of the postero-anterior dimension of more extreme degrees of laryngeal constriction and it can occur regardless of whether the vocal folds are in an adducted or abducted state.

In the AE stop, the upper laryngeal airway progressively collapses: the ventricular folds adduct (frame 19), as observed during glottal stop, but in this case the concomitant postero-anterior narrowing continues to the point of contact between the aryepiglottic

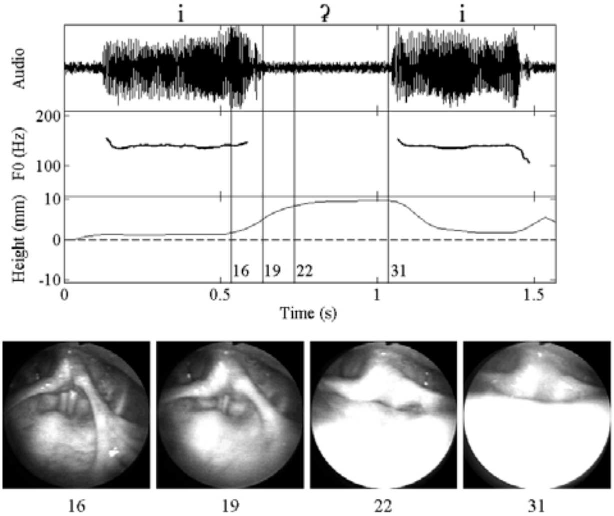

Figure 7: SLLUS data for aryepiglotto-epiglottal stop.

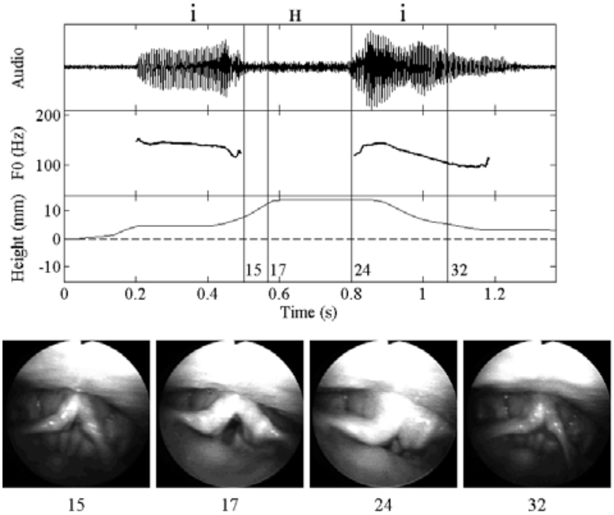

Figure 8: SLLUS data for a voiceless aryepiglotto-epiglottal trill.

folds and the epiglottis (frame 22). Larynx height rises by 8.62 mm from frame 16 to 31, where it reaches its peak and then begins its descent during the following vowel. Comparing the LS with the LUS, it is evident that larynx raising is slower acting relative to epiglottis retraction and advancement of the aryepiglottic folds. Epiglottis (and presumably lingual) retraction appears to increase in degree throughout the production and peaks towards the offset of the sound (frame 31).

For the voiceless AE trill in Figure 8, larynx height once again peaks shortly after postero-anterior narrowing of the epilaryngeal tube is complete.

Trilling begins at a point between frames 15 and 17 while the larynx is rapidly moving upwards. Total larynx displacement is 7.05 mm measured from the mid-point of the first vowel to the point show in frame 24. Note the visual correlation between descending larynx height and falling F0 during the vowel following the trill.

5. Discussion

To understand the structural and temporal relationships observed in the data, it is useful to consider the laryngeal architecture from a mechanical perspective. The fact that larynx height consistently lags behind intrinsic laryngeal articulations is attributable to inertial factors, namely the relative sizes of masses involved: the ventricular folds, aryepiglottic folds, and epiglottis are all relatively quick acting, whereas the larynx height mechanism is relatively slow; vertical velocity of larynx for the various productions was on the order of 10 to 100 mm/s (in either direction).

Ventricular incursion plays a key role in laryngeal constriction, particularly in glottal stop and creaky voice: it mechanically alters the vocal fold dynamics associated with these states. For glottal stop, ventricular incursion likely aids in rapidly arresting the motion of the vocal folds, which will tend to continue oscillating assuming no change in subglottal pressure. If vocal-ventricular coupling occurs, then the damping (due to energy transmission into the body of the ventricular folds) and increased mass of the oscillating system will greatly increase its mechanical impedance, inhibiting phonation.

Although it could not be clearly visualized with the laryngoscopy, we also suspect that ventricular incursion interferes with the vocal fold mucosal waves, which play an important role in promoting self-sustained oscillation of the vocal folds. The result may be to destabilize the system, producing a change in phonatory regime characterizable as creaky phonation. Suppression of larynx height during creaky phonation (as in Figure 6) evidently mitigates the latero-medial degree of ventricular incursion, but we posit that ventricular incursion still occurs in this case based on the engagement of the postero-anterior epilaryngeal tube narrowing visible in the laryngoscopy at the moment creakiness begins, a slight medial movement of the ventricular folds, and the accompanying change to the appearance of the vocal folds and their movement.

The pharyngeal consonants rely more upon larynx raising for their execution, but the compaction of the epilaryngeal tube in the axial plane is complete prior to full engagement of larynx height. The fact that epiglottis retraction also appears to persist late into the production of the consonant suggests to us that this may be due to relatively late peak in lingual retraction, which would have the effect of further displacing the epiglottis towards the aryepiglottic folds. Early epiglottis retraction might be driven by the more quick acting intrinsic laryngeal muscles connected with the epiglottis, such as the external thyroarytenoids, thyroepiglotticus, and possibly the aryepiglottic muscles [7]. It might be the case then that the internal mechanism of laryngeal constriction is the primary agonist of epilaryngeal stricture, while tongue retraction and larynx raising play supportive roles.

6. Conclusion

Timing relationships of the lower laryngeal vocal tract are difficult to evaluate empirically; we have demonstrated the simultaneous laryngoscopy and laryngeal ultrasound technique (SLLUS), which provides a means to obtain information on these relationships, with the help of optical flow analysis of the laryngeal ultrasound data as a means to quantify larynx height changes.

With this technique we have observed the cascade of gestures used for sounds that predominantly employ the larynx in their production. We have shown that larynx height plays a critical role in laryngeal constriction; there is evidence of its activity for sounds not typically considered to employ larynx raising, such as glottal stop and creaky phonation. Somewhat paradoxically, the role that the larynx serves for pitch production is in opposition to its role in laryngeal constriction: raising the larynx can raise pitch, but it also is essential in producing laryngeal constriction; furthermore, with laryngeal constriction engaged, pitch will tend to be low, as in creaky voice, if vocal fold dynamics are altered by vocal-ventricular coupling.

The SLLUS technique promises to be an exciting new domain of exploring laryngeal articulation. Its utility in the analysis of careful phonetic productions encourages future research of natural phonetic performance of native speakers.

REFERENCES

[1] Jassem, W. 1971. Pitch and compass of the speaking voice. *Journal of the International Phonetic Association* 1. 59–68.

[2] Edmondson, J. A. and 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.

[3] Honda, K., Hirai, H., Masaki, and Shimada, Y. 1999. Role of vertical larynx movement and cervical lordosis in F0 control. *Language and Speech* 42. 401–411.

[4] Vilkman, E., Sonninen, A., Hurme, P. and Körkkö, P.. 1996. External laryngeal frame function in voice production revisited: A review. *Journal of Voice* 10. 78–92.

[5] Honda, K. 1995. Laryngeal and extra-laryngeal mechanisms of F0 control. [In:] F. Bell-Berti & L. Raphael (Eds.) *Producing Speech: Contemporary Issues — for Katherine Safford Harris*. New York: AIP Press. 215–245.

[6] Esling, J. H., Zeroual, C. and Crevier-Buchman, L. 2007. A study of muscular synergies at the glottal, ventricular and aryepiglottic levels. [In:] J. Trouvain and W. J. Barry (Eds.) *Proceedings of the 16th International Congress Phonetic Sciences*. Saarbrücken: Universität des Saarlandes, Germany. 585–588.

[7] Painter, C. 1986. The laryngeal vestibule and voice quality. *Archives of Oto-Rhino-Laryngology* 243. 329–337.

[8] Laver, J. 1980. *The Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press.

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

[10] Moisik, S. R., Esling, J. H., Bird, S. and Lin, H. 2011. Evaluating laryngeal ultrasound to study larynx state and height. [In:] W.-S. Lee and E. Zee (eds.) *Proceedings of the 17th International Congress of Phonetic Sciences*. 136–139.

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

[12] Kawahara, H., de Cheveigné, A. and Patterson, R. D. 1998. An instantaneous-frequency-based pitch extraction method for high quality speech transformation: revised TEMPO in the STRAIGHT-suite. Paper presented at the *International Conference on Spoken Language Processing* 1998, Sydney, Australia. December 1998.

[13] Kagaya, R. 1974. A fiberscopic and acoustic study of the Korean stops, affricates, and fricatives. *Journal of Phonetics* 2. 161–180.

[14] Hollien, H. & Allen, E. L. 1972. Laminagraphic investigation of pulse register (vocal fry) phonation (A). *Journal of the Acoustical Society of America* 52 (1A). 124.

[15] Lindqvist-Gauffin, J. 1972. A descriptive model of laryngeal articulation in speech. *QPSR, Speech Transmission Laboratory, Royal Institute of Technology* 13(2–3). 1–9.