

EVALUATING LARYNGEAL ULTRASOUND TO STUDY LARYNX STATE AND HEIGHT

Scott R. Moisk, John H. Esling, Sonya Bird & Hua Lin

University of Victoria, Canada

srmoisk@uvic.ca; esling@uvic.ca; sbird@uvic.ca; luahin@uvic.ca

ABSTRACT

In this work, we evaluate laryngeal ultrasound as a technique to study the state and height of the larynx. We first discuss general structural registration in laryngeal ultrasound (LUS) and illustrate the articulation of glottal stop. We then present a method for quantifying change in larynx height using optical flow analysis on laryngeal ultrasound video data. The analysis is quantitatively validated for accuracy by using independent control data. Then we qualitatively validate the method by performing simultaneous laryngoscopy and laryngeal ultrasound (SLLUS) on canonical productions of Mandarin tones. We conclude that laryngeal ultrasound is best suited for quantification of larynx height but can also provide limited information about larynx state.

Keywords: laryngeal ultrasound, larynx, quantifying larynx height, laryngeal articulation

1. INTRODUCTION

In linguistic research, ultrasound has primarily been used to study the tongue [4, 15]. The larynx has received far less attention using this imaging technique. Some early attempts with M-mode ultrasound were made to study vocal fold oscillation [6, 11]. Although not strictly linguistic in nature, researchers have recently used colour Doppler imaging to study the surface mucosal waves of the vocal folds [12] and ventricular folds [16].

Medical imaging studies [5, 10, 13] attest that laryngeal ultrasound can be used to image a wide array of laryngeal structures, such as the hyoid bone, strap muscles, pre-epiglottic space, and the epiglottis and thyroid, arytenoid, and cricoid cartilages. The vocal folds mainly appear as darkened regions (due to acoustic scattering in muscle tissue), while the ventricular folds are highly visible, and the laryngeal vestibule itself generates bright echoes due to the presence of air pockets. Acoustic ‘blind spots’ generated from the laryngeal air column partially obscure the

arytenoids, posterior commissure, pyriform fossae, aryepiglottic folds, and the hypopharynx. Calcification/ossification of laryngeal cartilages can also negatively impact laryngeal imaging with ultrasound.

Our work demonstrates that laryngeal articulation can be usefully studied with ultrasound. We first show that some articulations can be directly imaged, such as glottal stop. Then we outline how larynx height can be quantified using optical flow analysis; to illustrate this technique, we use canonical productions of Mandarin tones which show correlations between change in F0 and larynx height. We contend that this technique for quantifying larynx height is superior to other techniques such as thyroumbrometry [3] or MRI [7] because it offers the advantages of ease, non-invasiveness, applicability to a wide variety of participants, and high temporal resolution.

2. LARYNGEAL ULTRASOUND METHODOLOGY

This study involves two components. The first component is an examination of laryngeal structure and articulation from the coronal plane. The second component concerns the method for quantifying larynx height and the validation of this method.

For both components of this study we used the same basic setup and approach to laryngeal ultrasound. We used a portable LOGIQ e R5.0.1 system with an 8C-RS probe (both made by GE). The probe pulse frequency was 10 MHz, which allowed for optimal resolution of laryngeal structures. The field of view was consistently set to 120°. The system is pre-calibrated for measurement in the imaging plane; a ruler on the image allows for a pixel-mm scale to be determined. A Sennheiser ME66-K6 shotgun microphone was used to record audio, digitized at 44100 Hz (16 bit), using an M-Audio Mobile Pre-Amp as an external sound card. The video of the ultrasound machine was captured using an XtremeRGB video card at 30 fps (uncompressed,

8-bit greyscale, 1024x768 pixels) and both signals were integrated and manually checked for alignment using Sony Vegas Pro (version 8.0b).

To conduct the laryngeal ultrasound, participants were seated in an examination chair equipped with a head rest to support the head and help provide stabilization. The ultrasound probe was applied manually to the participant's right thyroid lamina near the laryngeal prominence and orientated to obtain a coronal image of the larynx. The probe was held such that the examiner's index finger and thumb were free to anchor on the participant's laryngeal prominence and side of the neck, respectively. This helped to maintain consistent probe placement during the examination. Before elicitation commenced, the participant was instructed to produce an [i] vowel at a normal pitch so that the vocal folds could be located and centered in the ultrasound view. It is during this point of the examination that we obtain basic structural registration.

2.1. Laryngeal ultrasound and larynx state

Using the above methodology, five trained phoneticians and one lay-person (three males and three females; five between 25-30 years of age, one, a female, at 45 years of age) were instructed to produce the sequence [iʔi].

2.2. Simultaneous laryngoscopy and laryngeal ultrasound (SLLUS) of Mandarin tones

Fifteen canonical productions of each of the four Mandarin tones (T1/high-level, T2/mid-rising, T3/low-rising, T4/high-falling) over syllables with unaspirated stop onsets (e.g. [pi]) were produced by a trained phonetician while being examined with SLLUS. To conduct SLLUS, laryngeal ultrasound is performed (as described above) simultaneously with a standard nasoendoscopic laryngoscopy examination. The laryngoscopy equipment used in this study is an Olympus ENF-P3 flexible fiberoptic nasal laryngoscope fitted with a 28 mm wide-angle lens to a Panasonic KS152 camera. The video signal was recorded using a Sony DCR-T4V17 digital camcorder. The camcorder also recorded an additional audio signal to aid in synchronization of the laryngoscopy data with the laryngeal ultrasound. As with the laryngeal ultrasound video, all laryngoscopy video was post-processed using Sony Vegas Pro for alignment of the video and audio signals.

2.3. Optical flow analysis

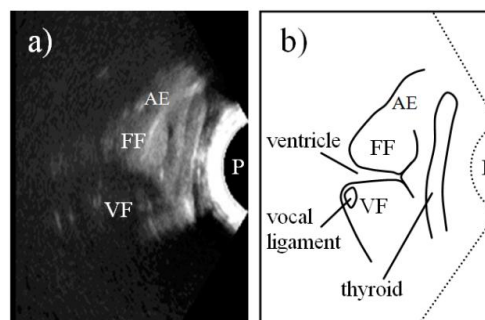
Optical flow analysis is used to quantify larynx height from the LUS data; it involves quantifying motion in video as a discrete velocity vector field [8]. Flow vectors, which by default have units of pixels/s, are converted to mm/s by obtaining a pixel/mm ratio from the ruler superimposed on the image by the ultrasound device. We use a custom optical flow algorithm based on cross-correlation to perform a block-wise calculation of flow vectors. To balance analysis resolution against the influence of ultrasound noise, we use a 15x15 pixel² block size, which is augmented to include all neighbouring analysis blocks (enlargement by a factor of 3), and dilate the pixel data with a 3 × 3 square morphological structure element.

Once the flow field data are created, global movement patterns are extracted by obtaining average flow vectors. Small vectors below a cut-off magnitude and vectors that are statistical outliers are removed to improve the accuracy of this mean. For studying larynx height, the vertical vector component is used. Time series signals of larynx height and F0 are statistically analyzed using smoothing spline ANOVA (SSANOVA) [1].

3. RESULTS

Structural registration in the coronal plane (Figure 1) for all participants consistently allowed for the ventricular folds to be identified, which can be attributed to their hyper-reflectivity [10], but general structural visibility and image quality varied by individual, with males registering better than females, likely due to variation in tissue density and structural orientation [10, 13].

Figure 1: Structural registration in coronal laryngeal ultrasound. (a) LUS image; (b) schematic. AE = aryepiglottic fold; FF = ventricular (false) fold; P = probe location; VF = vocal fold.



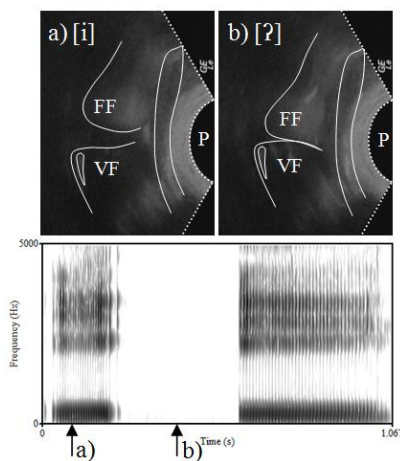
The 45-year-old female exhibited the poorest registration, which may be due to increased ossification of laryngeal cartilages. Vocal fold

oscillation is easily detected as flickering in the video image, particularly associated with the vocal ligament, despite generally poor registration. Using phonation to locate the vocal folds and thereby establishing a frame of reference greatly facilitates interpreting the laryngeal ultrasound image.

3.1. Laryngeal state and glottal stop

Due to poor registration of the vocal folds, it is difficult to ascertain *glottal* state purely from the coronal laryngeal ultrasound image; however, the tendency for the ventricular folds to image well in ultrasound means that laryngeal constriction can be partially observed in the form of *ventricular incursion* [2, 9], whereby the ventricular folds impact into the superior surface of the vocal folds. Every participant was observed to employ ventricular incursion in producing glottal stop. Figure 2 illustrates this for a male participant. The spectrogram is consistent with glottal stop: there are no formant transitions into the consonant and some creakiness precedes and follows the sound.

Figure 2: Laryngeal ultrasound of [iʔi]. (a) first [i]; (b) [ʔ]. FF = ventricular (false) fold; P = probe location; VF = vocal fold. Key structures are traced for easier interpretation of the image. Arrows show temporal location of image (a) and (b) in the spectrogram.

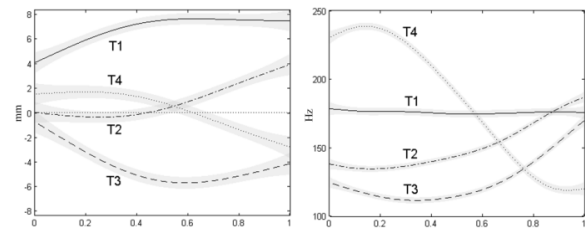


3.2. Quantifying change in larynx height

To ensure that the optical flow algorithm for quantifying larynx height is accurate, control video of a metal bar sliding 11.14 cm along a ruler was analyzed manually and with the algorithm. The normalized RMS error between the two measurements was 12.17% and numerical integration of the velocity data obtained from the algorithm yields 11.35 cm, for an error of 1.8%.

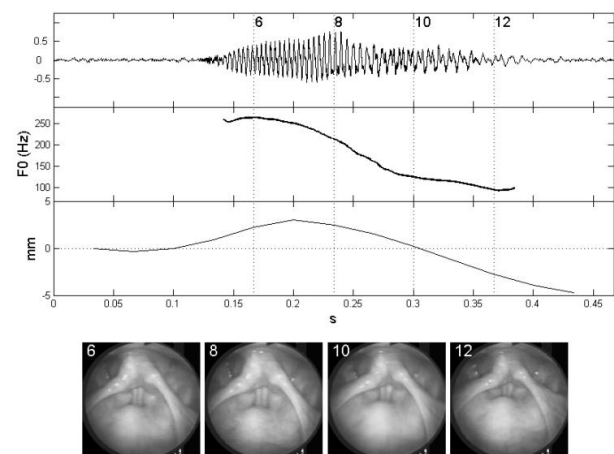
SLLUS is used to evaluate the algorithm for quantifying larynx height using canonical Mandarin tones as linguistic data. Figure 3 shows the results for all tones.

Figure 3: Larynx height quantification (left) results for the four canonical Mandarin tones (T1, T2, T3, T4) along with F0 contours (right) for comparison. Normalized time is on the abscissa. Grey region around each signal is the SSANOVA-based 95% confidence interval [1].



The laryngoscopy component of the SLLUS technique allows us to visually verify that the changes in larynx height indicated by the optical flow based quantification are actually occurring. Figure 4 illustrates this for the high-falling tone (T4). The initial part of the tone is at a high F0 (280 Hz): both the laryngoscopy (frames 6 and 8) and the height change quantification indicate that the larynx rises during this part of the tone, although the larynx height peak lags behind the F0 peak by nearly 50 ms. After this point, the larynx begins to descend in correspondence with F0. The laryngoscopy confirms by visual impression that the larynx does indeed appear to be descending (from frame 8 to 12). The total change in larynx height over the syllable is about 6 mm.

Figure 4: Illustration of SLLUS data for tone 4 (high-falling) on [pi]: waveform (top), F0 (middle), and change in larynx height (bottom). Selected laryngoscopy frames are shown below and vertical dotted lines on plots show their location in time.



4. DISCUSSION

Laryngeal ultrasound has not been conducted extensively in phonetic research. The poor structural registration and complexity of laryngeal anatomy present significant challenges for the researcher attempting to interpret articulatory behaviour using this imaging modality. It is possible to observe laryngeal articulation, especially involving the ventricular folds, which do register well in the image because of high acoustic reflectivity.

More promising is the use of laryngeal ultrasound for quantifying change in larynx height. Consistent results were obtained using the optical flow algorithm to analyze the LUS video data. This technique benefits from non-localized velocity measurement made by averaging vectors from the optical flow field. Thus, despite the general noisiness and frame-by-frame discontinuities of ultrasound video, optical flow is useful for obtaining a global estimate of motion in the LUS video. Generally laryngeal motion in coronal LUS is translational in the vertical dimension, but local divergence from this tendency does not impact the analysis because it is not strongly dependent on local changes in the image.

Other techniques for measuring larynx height are not as flexible as LUS. Thyroumbrometry [3] [14] requires participants with large laryngeal prominences, and measurement made from the thyroid notch will be confounded by rotations of the thyroid. MR imaging (e.g. [7]) is likely the most accurate, but it is costly, difficult to perform, and cannot (currently) capture real-time speech processes.

5. CONCLUSION

Laryngeal ultrasound can be used to image the articulatory contribution of the ventricular folds in glottal stop: ventricular incursion was observed for all participants in this study. Laryngeal ultrasound is also a viable technique, in conjunction with optical flow analysis, for quantifying change in larynx height. The optical flow algorithm was validated using independent video data. Laryngoscopy was used to corroborate larynx height change determined by the algorithm. All data support our claim that the technique provides accurate measurement of change in larynx height.

6. ACKNOWLEDGEMENTS

Research supported by Canadian Foundation for Innovation and SSHRC Canada SRG and PGS Fellowship.

7. REFERENCES

- [1] Davidson, L. 2006. Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *JASA* 120(1), 407-415.
- [2] 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.
- [3] Ewan, W.G., Kronen, R. 1974. Measuring larynx movement using the thyroumbrometer. *Journal of Phonetics* 2, 327-335.
- [4] Gick, B., Bird, S., Wilson, I. 2005. Techniques for field application of lingual ultrasound imaging. *Clinical Linguistics and Phonetics* 19, 503-514.
- [5] Harries, M., Hawkins, S., Hacking, J., Hughes, I. 1998. Changes in male voice at puberty: Vocal fold length and its relationship to the fundamental frequency of the voice. *J. of Laryngology and Otology* 112, 451-454.
- [6] Hertz, C.H., Lindstrom, K., Sonesson B. 1970. Ultrasonic recording of the vibrating vocal folds. *Acta Otolaryngologica* 69, 223-230.
- [7] Honda, K., Hirai, H., Masaki, S., Shimada, Y. 1999. Role of vertical larynx movement and cervical lordosis in F0 control. *Language and Speech* 42, 401-411.
- [8] Horn, B.K.P., Schunk, B.G. 1981. Determining optical flow. *Artificial Intelligence* 17, 185-203.
- [9] Lindblom, B. 2009. Laryngeal mechanisms in speech: The contributions of Jan Gauffin. *Logopedics Phoniatrics Vocology* 34(4), 149-156.
- [10] Loveday, E. 2003. Ultrasound of the larynx. *Imaging* 15, 109-114.
- [11] Mensch, B. 1964. Analyse par exploration ultrasonique du mouvement des cordes vocales isolées. *Compt. Rend. Biol.* 12, 2295.
- [12] Shau, Y., Wang, C., Hsieh, F., Hsiao, T. 2001. Noninvasive assessment of vocal fold mucosal wave velocity using color Doppler imaging. *Ultrasound in Medicine and Biology* 27(11), 1451-1460.
- [13] Sonies, B.C., Chi-Fishman, G., Miller, J.L. 2002. Ultrasound imaging and swallowing. In Bronwyn, J. (ed.), *Normal and Abnormal Swallowing: Imaging in Diagnosis and Therapy*. New York: Springer, 119-138.
- [14] Sprouse, R.L., Solé, M.J., Ohala, J.J. 2010. Tracking laryngeal height and its impact on voicing. *12th Conference on Laboratory Phonology*, University of New Mexico, Albuquerque.
- [15] Stone, M. 2005. A guide to analysing tongue motion from ultrasound images. *Clinical Linguistics and Phonetics* 19(6-7), 455-502.
- [16] Tsai, C., Shau, Y., Hsiao, T. 2004. False vocal fold surface waves during Sygyt singing: a hypothesis. *ICVPB*, Marseille.