

# A high-speed laryngoscopic investigation of aryepiglottic trilling<sup>a)</sup>

Scott R. Moisik and John H. Esling

*Department of Linguistics, University of Victoria, Victoria, British Columbia V8W 3P4, Canada*

Lise Crevier-Buchman

*Laboratoire de Phonétique et Phonologie, UMR 7018, CNRS/Sorbonne-Nouvelle, 19 rue des Bernardins, 75005 Paris, France*

(Received 19 December 2008; revised 17 November 2009; accepted 30 November 2009)

Six aryepiglottic trills with varied laryngeal parameters were recorded using high-speed laryngoscopy to investigate the nature of the oscillatory behavior of the upper margin of the epilaryngeal tube. Image analysis techniques were applied to extract data about the patterns of aryepiglottic fold oscillation, with a focus on the oscillatory frequencies of the folds. The acoustic impact of aryepiglottic trilling is also considered, along with possible interactions between the aryepiglottic vibration and vocal fold vibration during the voiced trill. Overall, aryepiglottic trilling is deemed to be correctly labeled as a trill in phonetic terms, while also acting as a means to alter the quality of voicing to be auditorily harsh. In terms of its characterization, aryepiglottic vibration is considerably irregular, but it shows indications of contributing quasi-harmonic excitation of the vocal tract, particularly noticeable under conditions of glottal voicelessness. Aryepiglottic vibrations appear to be largely independent of glottal vibration in terms of oscillatory frequency but can be increased in frequency by increasing overall laryngeal constriction. There is evidence that aryepiglottic vibration induces an alternating vocal fold vibration pattern. It is concluded that aryepiglottic trilling, like ventricular phonation, should be regarded as a complex, if highly irregular, sound source. © 2010 Acoustical Society of America. [DOI: 10.1121/1.3299203]

PACS number(s): 43.70.Gr, 43.70.Fq, 43.70.Jt, 43.70.Bk [AL]

Pages: 1548–1558

## I. INTRODUCTION

There are numerous structures in the vocal tract apart from the vocal folds that can generate phonatory oscillations when subject to appropriate aerodynamic conditions: the esophageic sphincter, ventricular folds, and aryepiglottic folds are some examples (Catford, 1968; Lindestad *et al.*, 2001; Sakakibara *et al.*, 2004). The use of these structures in generating a voice source has been observed for pathological cases, where vocal fold phonation is not possible or done inefficiently (Fröhlich *et al.*, 2000; Schwarz, 2007). The aryepiglottic folds are unique among these structures because they roughly parallel the vocal folds in their ability to produce both articulations and phonatory voicing sources. The aryepiglottic folds contribute a very distinct category of phonation, typically referred to as growl or harsh voice. Singers such as Louis Armstrong, Louis Prima, Koko Taylor, and Tom Waits readily come to mind as famously exploiting aryepiglottic vibration to produce their unique singing styles, illustrating the application of the sound as a voice quality. Esling (1996) and Edmondson and Esling (2006) argued that the aryepiglottic folds play a central role in forming the primary constriction in the pharynx associated with pharyngeal articulations in general. More specifically, these researchers re-attribute the former categories of epiglottal and pharyngeal to the (active) aryepiglottic and the (passive) epiglottal

place of articulation, which forms the locus of pharyngeal constriction, achieved supplementarily by retraction of the tongue (and epiglottis) and by raising of the larynx. Their laryngoscopic evidence shows that numerous languages employ the aryepiglottic folds in a variety of ways to produce linguistic contrasts (Carlson *et al.*, 2004; Edmondson *et al.*, 2005; Esling and Edmondson, 2002; Esling *et al.*, 2005). Articulatory possibilities include stop, fricative, approximant, trill, and possibly even tap.<sup>1</sup> The aryepiglottic sphincter is shown to play a role in the production of creaky<sup>2</sup> and harsh voice, which also can exhibit aryepiglottic trilling/vibration. Additionally, languages containing pharyngeal fricatives and approximants can use aryepiglottic vibration as a means of enhancing the distinctiveness of these sounds (Esling, 2005).

The aryepiglottic folds are anatomically situated at the top of the epilaryngeal tube (also referred to as the laryngeal vestibule; see Fig. 1). The tube can be reduced in volume and ultimately form a protective hermetic seal by means of laryngeal constriction (in the sense of Gauffin, 1977; also see Fink, 1975, pp. 85–102). Titze (2008) argued that this tube plays an important role in coupling the glottal source to the rest of the vocal tract, specifically when the epilaryngeal tube is reduced in cross-sectional area through laryngeal constriction (less than or equal to 0.5 cm<sup>2</sup>). Two interactions between the vocal folds and the epilaryngeal tube are identified: First, flow pulse skewing can energize glottal harmonics below the first resonance of the vocal tract and, second, vocal

<sup>a)</sup>This paper is derived from research originally presented at the 6th ICVPB, Tampere, Finland, 6–9 August 2008.

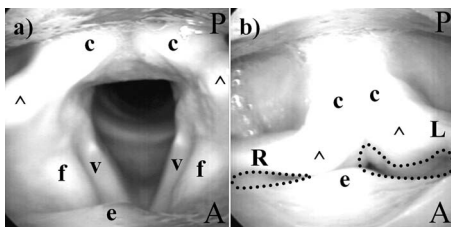


FIG. 1. A superior view of the epilaryngeal tube during the video of a voiced aryepiglottic trill at 100 Hz. Image (a) shows the tube in its unstricted state (frame 19), and image (b) shows the tube in its constricted state (frame 396). Dotted-line=aryepiglottic aperture, c=corniculate tubercle, ^=cuneiform tubercle, e=epiglottic tubercle, f=ventricular (false) fold, v=vocal fold, R=right aryepiglottic fold, L=left aryepiglottic fold, A=anterior (front), and P=posterior (back).

fold oscillatory efficiency (the ratio of radiated acoustic power to aerodynamic power) can be enhanced. Using computational models, both of these effects are evinced to be augmented under conditions of reduced epilaryngeal tube area at the expense of destabilizing the oscillation of the vocal folds. Additionally, the epilaryngeal tube also contributes to the resonances of the vocal tract. With respect to the acoustics of the singing voice, Sundberg (1977) claimed that the epilaryngeal tube (or, in his terms, the outlet of the larynx) becomes independent of the vocal tract and produces its own resonance when the cross-sectional area of the epilaryngeal tube is one-sixth the area of the pharynx. Titze and Story (1997) further argued that the epilaryngeal tube is most suitably characterized as a quarter-wave resonator with its first resonance between 2500 and 3000 Hz (also known as the singer's formant). Under voiceless conditions, a turbulent noise source can be produced from an air jet impinging on the posterior surface of the epiglottis and the medial surfaces of the ventricular folds. Stevens (1998, p. 430) estimated that as much as two-thirds of the acoustic energy in glottal fricatives [h] is produced at the surface of the epiglottis. Stevens' (1998) analysis assumes that the pharyngeal airway is not constricted, which presumably means that the epilaryngeal tube is also considered to be unstricted (i.e., unreduced in cross-sectional area). It is suspected, however, that increased constriction of this tube would also yield an effect on the character of laryngeal-pharyngeal noise sources, which characterize the voiceless pharyngeal fricative /h/. For Arabic /ħ/, Butcher and Ahmad (1987) reported high intensity friction noise values with a lower limit between 500 and 1500 Hz and distinct concentrations of noise energy in formant regions associated with the vowel context.

Overall, laryngeal constriction causes the epilaryngeal tube to change the acoustic and aerodynamic properties of the vocal tract. In addition to these effects, the oscillatory capacity of the epilaryngeal tube represents a further degree of complexity of the laryngeal system as a whole. Specifically, it is posited that aryepiglottic trilling can be regarded as a vocal tract source, one that can interact with the glottal source but that is also independent from it. The purpose of this paper is to adduce evidence that supports this characterization and provides an illustration of the oscillatory nature of aryepiglottic trilling in general. The assertion that the aryepiglottic folds are source generators will be supported by

comparing aryepiglottic aperture data and kymographic traces with the corresponding acoustic waveforms and electroglottograph (EGG) signals for several aryepiglottic trills with various glottal settings (i.e., pitch and the presence or absence of voicing). The appropriateness of the phonetic label *trill* will also be a central point of discussion, and the decision to use the label is motivated by the data brought forth in this study.

## II. HIGH-SPEED LARYNGOSCOPIC DATA

High-speed laryngoscopic videos of six different aryepiglottic trills were studied. All of the trilling sequences were performed by Esling, a 60-year-old male phonetician, who was trained in the Edinburgh tradition of phonetics. Two sets of aryepiglottic trills were performed with an [ə] vocalic target: one voiced and one voiceless. In the voiced set, glottal pitch was targeted at 100 and 200 Hz. In the voiceless set, the subject produced the trills by emulating the voiced configurations required for 100 and 200 Hz while keeping the vocal folds abducted (so that the sound produced would be voiceless). An additional parameter for both voiced and voiceless trills was explored: degree of laryngeal constriction. These are called "tense" trills (cf. Heselwood, 2007), although no claims as to the exact degree of muscular tension are made; rather, the term reflects the subject's own interpretation of his performance, expressing the extra effort or tension in producing these sounds. The performances were facilitated by the use of pre-recorded target sounds, which the subject listened to using headphones during the video recording session.

The system used to capture the trilling sequences was an SL Kamera 500 connected from a rigid oral endoscope to a SpeedCam Lite interface (Weinberger, Erlangen, Germany), which was set to a frame rate of 500 Hz (one frame for every 2 ms) with a resolution of 262 × 256 pixels. Recording was done at the Sorbonne-Nouvelle Paris III/CNRS-LPP-UMR 7018 research site at l'Hôpital Européen Georges Pompidou, Paris. The average sequence length was 1.945 s (or around 972 frames). Throughout the recording session, the audio signal of the trills was obtained using a head mounted AKG C410 microphone, positioned at 4 cm from the lips, with an angle of 45° from sagittal plane, sampled at 44 100 Hz, 16-bit resolution. The EGG signal was obtained using a Glottal Enterprises EG2-PCX electroglottograph; the surface electrodes were placed externally on each thyroid lamina, which was constant for all trill performances. The audio and video signals were digitized using ANAVOX (custom software; Vannier-Phototec, Antony, France). A time-to-live signal was used to ensure synchronization of the audio with the video signal. The videos were digitized frame by frame in uncompressed format and later stored as uncompressed TIFF files for analysis in MATLAB.

## III. AUTOMATED IMAGE ANALYSIS

An automated image analysis technique was employed to analyze the high-speed laryngoscopic videos. This involved assessing the percent of aperture formation for each of the aryepiglottic folds during oscillation. These data are

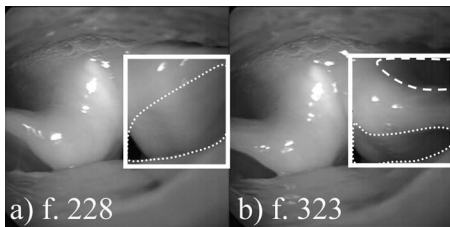


FIG. 2. Non-tense voiceless aryepiglottic trilling. The left fold undergoes extreme displacement, making ROI placement (solid rectangle) over the aperture (dotted-line) difficult. In (a) the left aperture is wide and requires a large ROI in frame 228 of the sequence. The same aperture in (b), however, at frame 323 encloses area over the piriform fossa (dashed-line) in addition to the aryepiglottic aperture.

accompanied by kymographic images and EGG signal for comparison. The following sections discuss this analysis in detail.

### A. Aperture area analysis

The aperture analysis involved the identification of regions of interest (ROIs) across the sequence of video frames for both of the aryepiglottic apertures (defined by the subject's right and left aryepiglottic folds; see Fig. 1) and then processing these regions to yield an image that could be measured using MATLAB's image processing functions, e.g., `bwarea()`. Two problematic circumstances arose that required special ROI placement for the left aryepiglottic aperture. For the normal and tense voiceless aryepiglottic trills at 100 Hz, the left aryepiglottic aperture fluctuated so much that it was difficult to select a ROI that maximally excluded non-aperture related areas (primarily the piriform fossa). This is illustrated in Fig. 2(a), where a ROI (solid line; not used in the actual analysis) is placed over an area large enough to enclose an extreme displacement of the left aryepiglottic fold (dotted-line). Later in the trilling sequence [Fig. 2(b)], the same ROI encloses a large portion of the space associated with the piriform fossa (dashed-line), in addition the aryepiglottic aperture itself. Another problem was observed for the voiced aryepiglottic trill at 200 Hz, where the epiglottis occluded the view of the left aryepiglottic aperture entirely, as seen in Fig. 3. In both of these extenuating cases, it was decided that oscillation information for the left aryepiglottic aperture could be still obtained by placing the ROIs over the left piriform fossa instead. Thus, for these cases, the ROIs used in the analysis monitor the position of the posterior surface of the left aryepiglottic fold as it occludes the view of the left piriform fossa. The resultant signal was inverted to match the phase of the actual aperture, which was assumed to be opposite to the signal (i.e., the observed piriform fossa area enlarges when the left aryepiglottic aperture diminishes). The drawback of this approach is that any subtle, minor vibrations (i.e., surface mucosal waves) only visible on the inner portion of the fold are lost; this was deemed an acceptable sacrifice so that at least major pulse information could be collected for the offending fold.

ROI processing is illustrated in Fig. 4; the leftmost image shows frame 236 of the voiced aryepiglottic trill at 100 Hz. In this frame, both the right and left ROIs are outlined. The processing steps used for all ROIs are demonstrated for

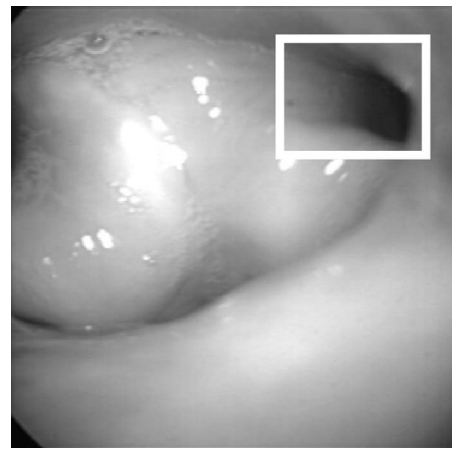


FIG. 3. Frame 166 of the "200 Hz" voiceless aryepiglottic trill. The left aperture is barely visible, but oscillation information was obtained by tracking a ROI (solid rectangle) placed over the left piriform sinus (upper right hand corner) and then inverting the signal. The right aperture is visible as a dark sliver (bottom left hand corner).

the right aperture in the images on the right [Figs. 4(a)–4(c)]. The ROIs were first extracted [Fig. 4(a)] from their corresponding frame and then inverted [Fig. 4(b)]. Before area measurement can be made, the image needs to be in binary representation [Fig. 4(c)]. This step requires a black-white luminance threshold value to be selected. The threshold values were determined automatically (prior to the actual measurement routine) by calculating the average grayscale value for each ROI from its grayscale histogram using the `imhist()` function in MATLAB. An average of all of the mean grayscale values for all ROIs in a given sequence was obtained and converted to a luminance threshold value by dividing by 255 (the number of values in an 8-bit grayscale image). Aperture data are expressed as a percentage of the maximum aperture area value recorded for a particular sequence and fold. Exact values for all parameters of the image analysis are presented in Table I. The overall process was automated using simple MATLAB scripts.

As a complement to the aperture data, kymographic images (e.g., Švec and Schutte, 1996) were created for both the left and right aryepiglottic apertures. A line of pixels (the kymographic line) was selected for each aperture, which was

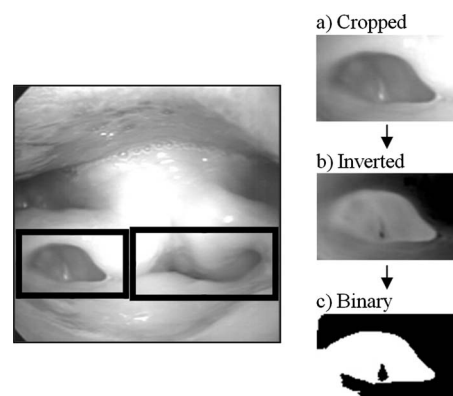


FIG. 4. Illustration of ROI definition over the left (larger rectangle) and right (smaller rectangle) aryepiglottic apertures and image processing sequence to arrive at a binary image that can be measured for aperture area.



TABLE I. A listing of values for the aperture analysis and related details of the high-speed videos of voiced (vd.) and voiceless (vl.) aryepiglottic trilling. Height and width dimensions are in pixels. CC analysis=cross-correlation analysis between aperture and kymographic data. “Yes” in the “inverted data” row indicates that the aperture data were inverted (see text for explanation).

Sequence	100 Hz				200 Hz				100 Hz tense			
	vd.		vl.		vd.		vl.		vd.		vl.	
Start frame	467		191		120		443		42		1041	
End frame	1441		2046		335		1555		1047		1861	
No. of frames	974		1855		215		1112		1005		676	
Duration (s)	1.948		3.710		0.430		2.224		2.010		1.352	
Aperture	Right	Left	Right	Left	Right	Left	Right	Left	Right	Left	Right	Left
Threshold	0.10	0.09	0.54	0.73	0.46	0.51	0.24	0.21	0.42	0.51	0.38	0.67
ROI width	101	131	100	102	53	84	43	151	45	99	115	93
ROI height	57	73	68	68	74	129	49	86	44	35	85	99
CC analysis	0.976	0.978	0.885	0.864	0.921	0.907	0.940	0.937	0.946	0.980	0.997	0.939
Inverted data	No	No	No	Yes	No	Yes	No	No	No	No	No	Yes

different in every video. The lines were selected such that they were perpendicular to the visible section of the fold (at rest) and bisected it. In some cases, due to camera position (as in Fig. 2), only a small portion of the fold was visible. In this case, the line was placed such that it passed through the point on the visible fold that exhibited the greatest displacement. Line length was decided by viewing the entire video and determining the minimal length required to capture as much of the aperture oscillation as possible while excluding non-aperture related image information. This process was complicated by the fact that in several videos the position of the larynx changes slightly through the performance of the aryepiglottic trill, which means that the image of the larynx migrates somewhat. Ultimately, the lines were only used if all of the researchers agreed that they represented the optimal choice given the video data.

The data obtained from the kymographic images were then converted into an oscillatory signal simply by summing the grayscale values for each line in the kymographic image. In the case that a kymographic image included non-aperture data (e.g., an image of the piriform fossa due to migration of the image of the larynx), these areas were masked by setting the pixel values to zero. The kymographic “signal” was then compared to the aperture data using cross-correlation analysis to quantify how similar the two signals were. Successful matching of the signals was defined to be indicated by a normalized zero-lag component of the cross-correlation analysis above 0.7071.<sup>3</sup> None of the signals obtained in this method resulted in a cross-correlation less than 0.7071; values of the analysis are presented in Table I.

The final technique used to analyze the aperture data was peak detection analysis, using a MATLAB script developed by O’Haver (2006), which allowed the average frequency and jitter of the EGG and aperture signals to be calculated. A peak detection analysis window of 50 ms was used for all signals; in the case of the aperture data, this window typically included two to five aryepiglottic pulses. There was a 25 ms overlap interval between analysis windows. The frequency of the signal for each analysis window was determined to be the inverse of the average peak-to-peak

(the signal was always positive) interval time; the frequency for the entire signal was the average of all analysis window results. Jitter was calculated by taking the standard deviation of all peak-to-peak durations in the signal.

Sections III A 1–III A 4 present a selection of data from the analysis described above that serve to characterize the nature of aryepiglottic trilling for our subject. Sections III A 1 and III A 2 examine the basic aryepiglottic trill in voiced and voiceless contexts; in both cases a contiguous set of 25 video frames is presented alongside an array of synchronized plots of the aryepiglottic aperture data, the kymographic images, and the audio signal. Section III A 3 documents the tense voiceless aryepiglottic trill. Section III A 4 presents data for all trills including acoustic spectra and Table II, which contains root mean square (rms) intensity values and peak detection analysis values.

### 1. Voiced aryepiglottic trill

Twenty-five frames of the voiced aryepiglottic trill produced at a glottal frequency of 100 Hz are presented in Fig. 5. The lines used to create the corresponding kymographic images have been superimposed on the first frame. This set corresponds exactly to a time period between 0.800 and 0.850 s in the data plots displayed in Fig. 6. In this time period, two right aryepiglottic fold pulses can be seen; the left fold also exhibits two major pulses, each with an additional minor pulse evident both in the kymographic image and the aperture signal. Throughout the entire trill, the right aryepiglottic fold oscillates semi-periodically; it typically makes complete contact with the surface of the epiglottis. Additional pulses tend to occur when complete contact is not made (e.g., at 0.860, 1.030, and 1.170 s). The gross movements of the left aryepiglottic fold appear to be largely in phase with the right fold during the section shown in the plots and have roughly the same frequency (see Table II); however, very often in this sequence the left fold oscillates with major and minor pulses, particularly between 0.800 and 1.100 s; after this point the pulsing diminishes in magnitude and the alternating major-minor pattern is not discernible. By

TABLE II. Intensity and frequency values for the audio signals of the voiced (vd.) and voiceless (vl.) aryepiglottic trills. rms intensity was obtained for the 200 ms region between the listed analysis frames. Frequency data are averaged across the entire trilling sequence.

Sequence	100 Hz				200 Hz				100 Hz tense			
	vd.		vl.		vd.		vl.		vd.		vl.	
Analysis frames	967–1067		341–441		221–321		1193–1393		947–1047		1191–1291	
rms intensity (dB)	−9.0309		−17.4958		−7.6700		−17.8252		−9.7551		−10.8040	
Actual $F_0$ (Hz)	104.0		...		200.3		...		110.6		...	
Glottal jitter (ms)	0.32		...		0.53		...		0.07		...	
Aperture	Right	Left	Right	Left	Right	Left	Right	Left	Right	Left	Right	Left
Average frequency (Hz)	65.9	82.4	67.4	42.4	64.7	69.8	55.8	100.4	92.0	100.8	73.8	71.6
Jitter (ms)	5.8	6.7	3.6	8.9	3.4	7.4	5.9	2.7	2.6	1.6	5.1	3.0

observing the videos themselves, the minor pulse of the left aryepiglottic fold appears to be a mucosal wave traveling on the aryepiglottic fold in the cranial (cephalad) direction. Comparison of the EGG signal with the audio and aperture data signals indicates that the aryepiglottic pulse actually corresponds to perturbations of the vibratory pattern of the vocal folds. A double peaked (or plateau-peak type) glottal pulse tends to occur every time the right aryepiglottic aperture is open (e.g., the pulse immediately after 1.000 s). The increase in noise in the acoustic signal at this point indicates that turbulent air is escaping from the glottis, which is open for a prolonged duration during the open phase of the right aryepiglottic pulse. When the right aryepiglottic aperture is closed, the glottis tends to open and close very rapidly. The left aryepiglottic aperture does not appear to have any discernible impact on the vocal folds, although its continually changing size is likely to impact supraglottal pressure.

## 2. Voiceless aryepiglottic trill

The first twenty-five frames and corresponding aperture and kymographic data of the voiceless aryepiglottic trill produced at an emulated 100 Hz can be seen in Figs. 7 and 8 (see Table II for frequency and intensity data). It is crucial to note that there was no actual glottal pulse, which is corroborated by the (essentially) flat EGG signal. Although the audio signal is considerably noisy (indicative of increased airflow due to glottal voicelessness), a distinctive pulse pattern can be seen, which roughly corresponds to the signals in the aperture data. Unlike the voiced aryepiglottic trill at 100 Hz, this trill exhibits massive left aryepiglottic fold displacement, as seen in the video frames (Fig. 7). Particularly interesting is the behavior of the left cuneiform cartilage, which can be seen to oscillate like a rubbery/highly elastic rod.

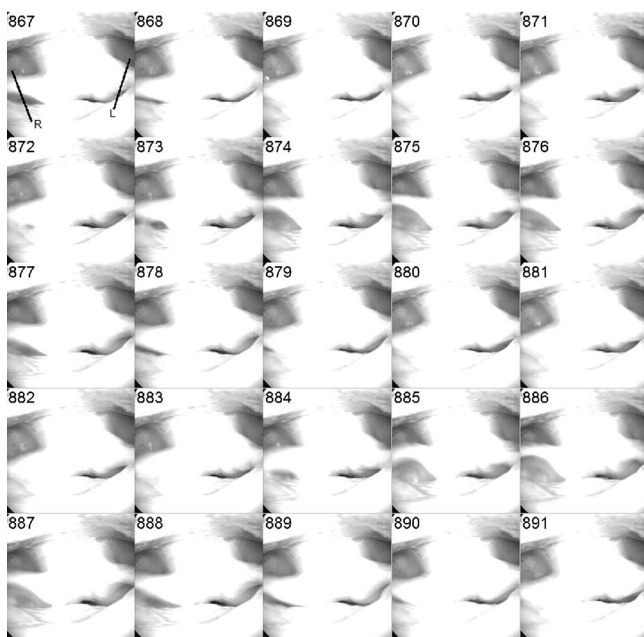


FIG. 5. Twenty-five frames of voiced aryepiglottic trilling (100 Hz glottal pitch, 25 frames, 50 ms). First frame (867) shows right and left lines used to extract kymographic images. Corresponds to Fig. 6 between 0.800 and 0.850 s.

## 3. The tense voiceless aryepiglottic trill

An alternative voiceless aryepiglottic trill was performed by our subject; this sequence is considered to be a tense variant of the voiceless trill, although the degree of tension is not strictly quantifiable. Despite the vagary of its characterization, the data in Fig. 9 indicate that this trill merits consideration. It should be noted again that the sound is voiceless, confirmed by the relatively flat EGG signal. The right aryepiglottic pulse pattern shows alternating major pulse and minor pulses: The acoustic effect associated with the major pulse can be clearly seen in the audio signal as a distinctive pressure spike, particularly in the section between 0.450 and 0.700 s. The negative part of this spike appears to coincide with the opening phase of the right aryepiglottic pulse, while the positive part occurs when the aperture is more fully open. Between 0.300 and 0.500 s, the left aryepiglottic fold tends to exhibit the major-minor pattern. The minor pulses must occur for the entire fold (not just the interior surface) because the left aryepiglottic aperture data, in this case, are based on measurements above the piriform fossa. Between 0.500 and 0.700 s, the left aryepiglottic fold begins to exhibit a more steady oscillatory pattern with approximately equal intensity pulses.

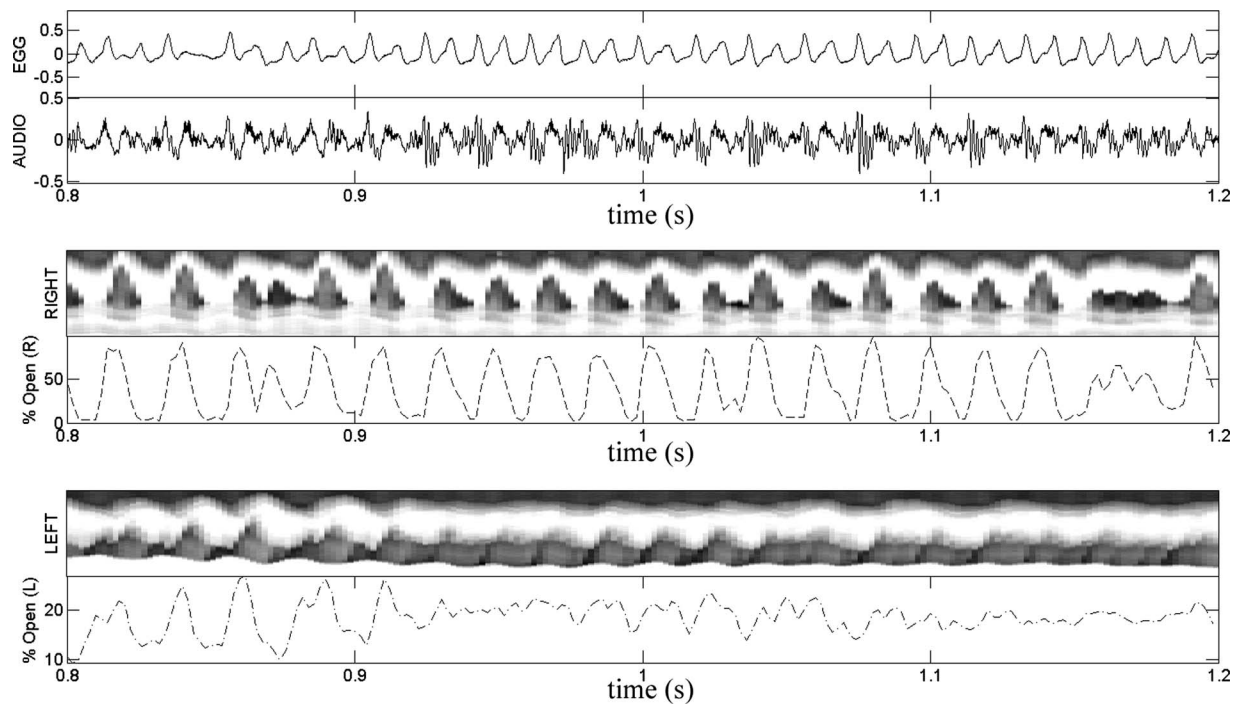


FIG. 6. Voiced aryepiglottic trilling (100 Hz glottal pitch): EGG signal and oscillogram (top), right aryepiglottic fold kymograph and aperture data (middle), and left aryepiglottic kymograph and aperture data (bottom).

#### 4. Aryepiglottic trill frequency data

Figure 10 contains spectra for all six of the trills. All of the spectra were obtained using a fast Fourier transform calculated over a 100 ms section of relative signal stability, using a 20 Hz Gaussian window. As the signals are highly noisy, the lower harmonic content was best visualized by limiting the spectrum display from 0 to 500 Hz. The voiced aryepiglottic trills all have fairly well defined and intense

glottal fundamentals, although the higher harmonics are not always clearly defined. The aryepiglottic pulses arguably explain the presence of nearly harmonic energy around the 50 Hz mark (in the 100 and 200 Hz voiced trills), although the noise in the signal and the jitter of the aryepiglottic pulse likely smear the pulse energy over a broader frequency range than the glottal pulse. The voiceless trills show harmonic spikes at values that correspond well with the values measured in the peak detection analysis (see Table II). The tense voiceless aryepiglottic trill is the exception to this where the spectral energy is exceedingly noisy; the peak at 70 Hz might correspond with the measured frequency of the folds, but the stronger peak at  $\sim 30$  Hz is possibly attributable to the major pulse of the right aryepiglottic fold that characterizes this trill. As mentioned in Sec. III A 3, the fundamental frequency of 73.8 Hz for the right fold during this trill was estimated based on both the major and minor pulses. If the minor pulses are ignored, then the mean frequency for the right fold in the tense trill is 31.2 Hz, which agrees with the spectrum of this sound.

Table II provides a listing of all measurements pertaining to the intensity of the overall signal and frequency of the vocal and aryepiglottic fold pulses. In all voiced cases, the average frequency of the glottal pulse (calculated from the EGG signal) corresponds well with the desired target pitches (100 and 200 Hz). In the case of our subject, the right aryepiglottic fold exhibits a frequency range between 55.8 and 92.0 Hz. The subject's left aryepiglottic fold was more difficult to measure (as discussed above in Sec. III A); it ranged between 42.4 and 100.8 Hz. Although not entirely consistent, the tense trills can be said to increase the pulse frequency for both of our subject's aryepiglottic folds by roughly 10–40 Hz.

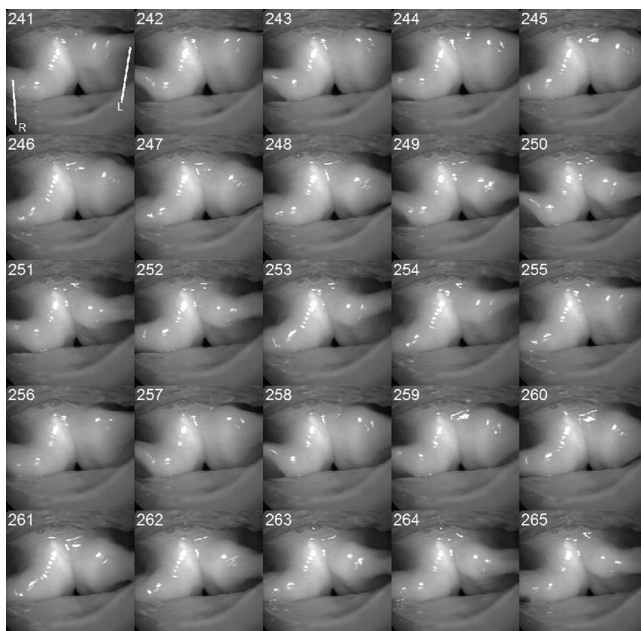


FIG. 7. Twenty-five frames of voiceless aryepiglottic trilling (emulated 100 Hz glottal pitch configuration, 25 frames, 50 ms). First frame (241) shows right and left lines used to extract kymographic images. Corresponds to Fig. 8 between 0.100 and 0.150 s.



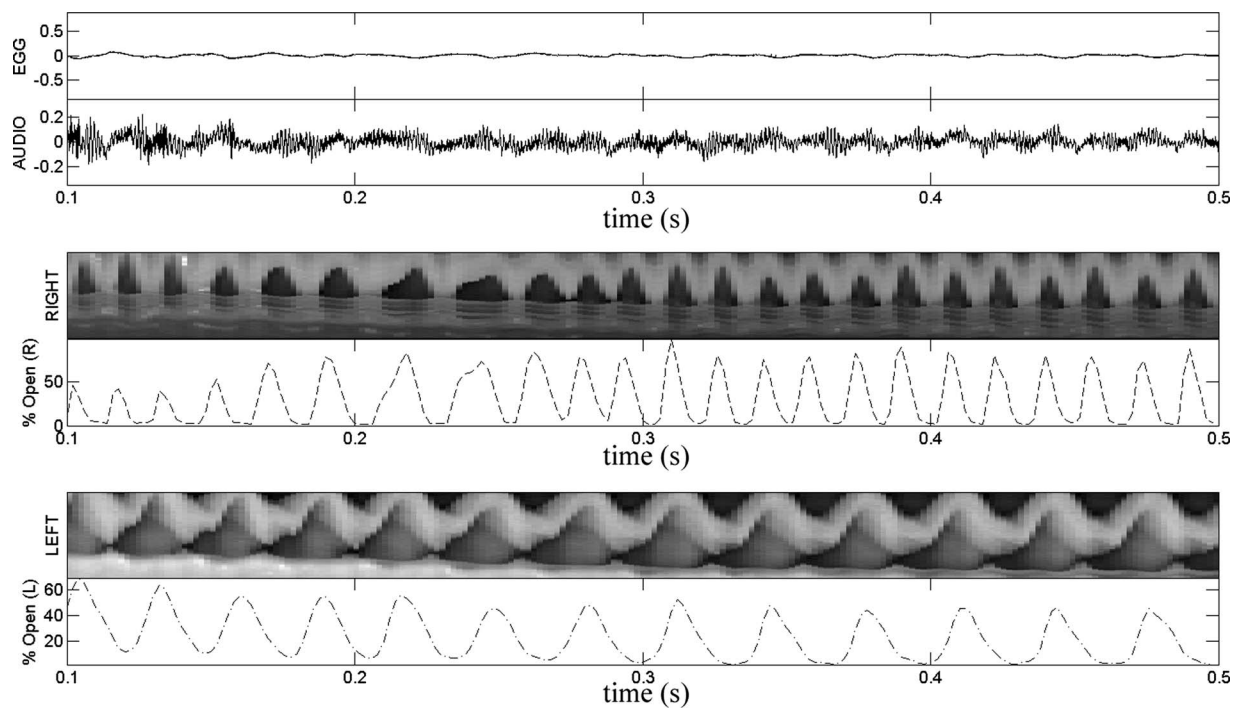


FIG. 8. Voiceless aryepiglottic trilling (emulated 100 Hz glottal pitch configuration): EGG signal and oscillogram (top), right aryepiglottic fold kymograph and aperture data (middle), and left aryepiglottic kymograph and aperture data (bottom).

The rms intensity values for the signals reveal a distinction between the non-tense voiced and voiceless trills: The voiced trills were approximately 8–10 dB more intense than the voiceless ones; however, the tense trills (voiced and voiceless) were of similar intensity. The acoustic spectra (Fig. 10) also show that the voiced trills are generally more intense than the voiceless ones. This difference is likely due to increased turbulent losses during the voiceless trills. The

aryepiglottic folds are evidently not as efficient at contouring the pulmonic airstream into a pulsatile flow as the vocal folds are.

#### IV. DISCUSSION

Even within the performance of a single individual, the variability of aryepiglottic fold trilling behavior is mani-

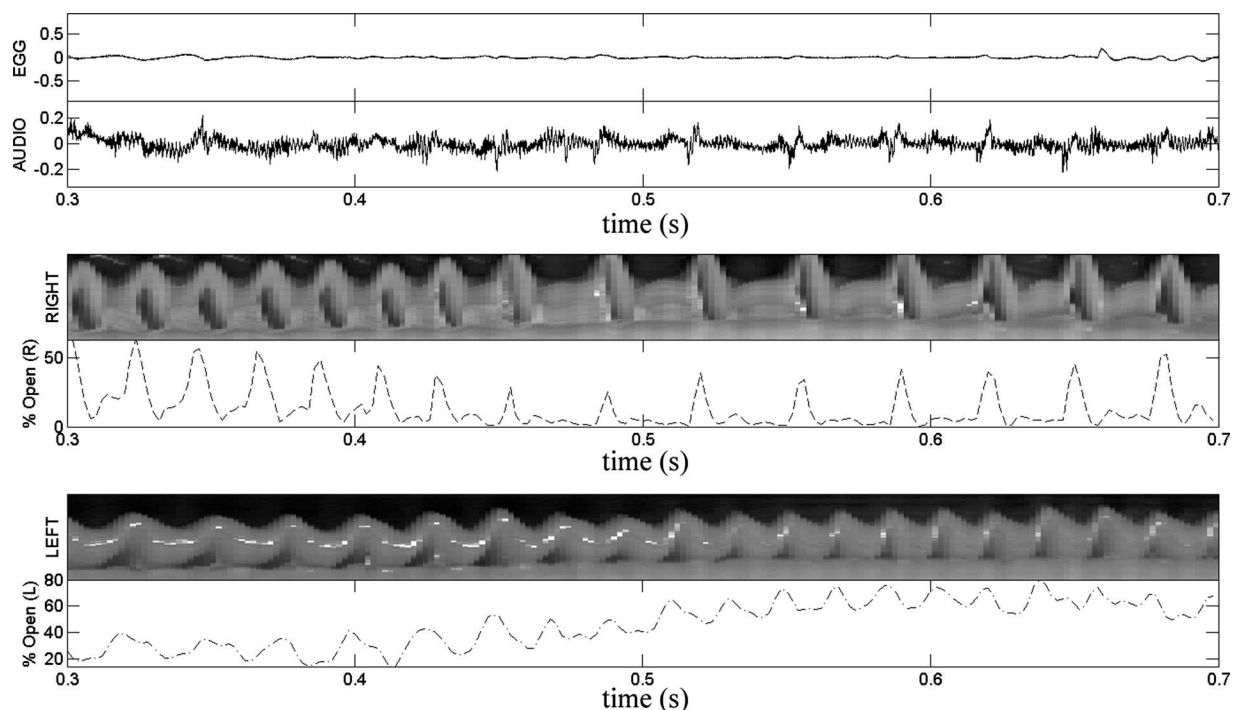


FIG. 9. Tense voiceless aryepiglottic trilling (emulated 100 Hz glottal pitch configuration): EGG signal and oscillogram (top), right aryepiglottic fold kymograph and aperture data (middle), and left aryepiglottic kymograph and aperture data (bottom).

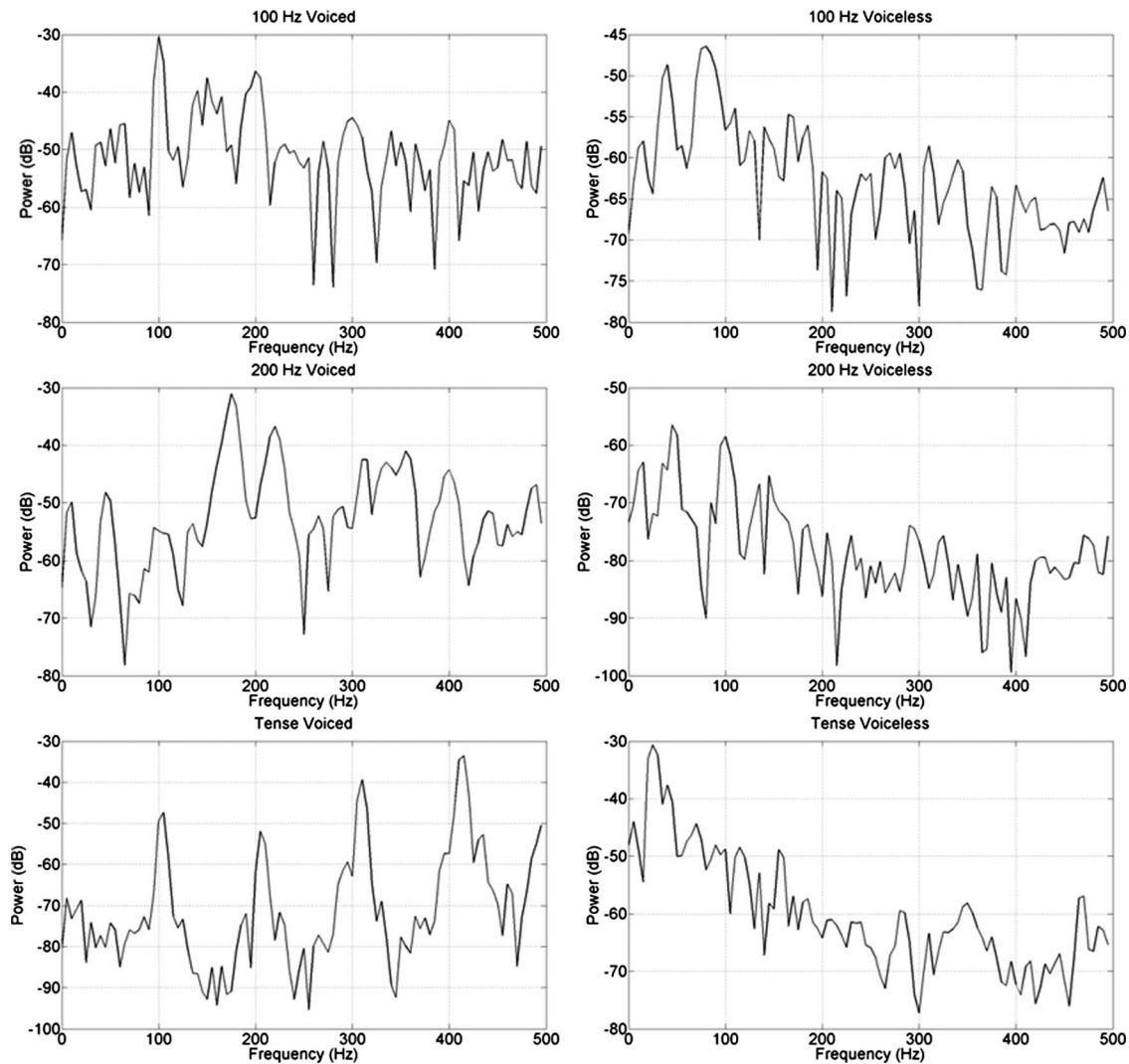


FIG. 10. Spectra of all trills (steady 100 ms interval used). Left columns are voiced trills (100 Hz, ~200 Hz, 100 Hz tense), and right columns are voiceless trills (emulated 100 Hz, 200 Hz, tense).

fested. Three major parameters were employed to study aryepiglottic trilling: the presence or absence of vocal fold vibration, glottal fundamental frequency, and the overall extent of laryngeal constriction. Although the latter parameter is only described qualitatively, it must be possible to vary constriction due to the fact that the aryepiglottic sphincter varies in configuration from being fully open to being fully closed, which is well documented in the languages of the world (e.g., [Esling, 1996](#); [Edmondson and Esling, 2006](#)). The 100 Hz set of aryepiglottic trills exhibits similar oscillatory behavior despite one trill being voiced and the other being voiceless. The reader is reminded that the left aryepiglottic fold frequency value measured for the voiceless trill (see [Table II](#)) is only indicative of the major pulse, as the actual aperture could not be measured. Thus, it is possible that the value reported is half of its actual value, which depends on the presence of the minor pulses, which can only be observed on the interior of the fold. The 200 Hz set revealed two notable changes in the oscillatory frequency. The voiced trill of this set behaves similarly to its 100 Hz counterpart, with the exception that the left aryepiglottic fold exhibited slower oscillations. The voiceless trill of this set, however, was re-

markably different: The right aryepiglottic fold decreases in frequency, while the left aryepiglottic fold considerably increases in frequency. The aperture data from this trill (not presented in this article) show that the left aryepiglottic fold oscillates in a particularly irregular way and with variable intensity; it does, however, consistently make full contact with the base of the epiglottis. It is noteworthy that, despite a doubling of glottal pitch, the aryepiglottic folds themselves do not double in frequency. Finally, the tense trills generally exhibited increased oscillatory frequency for both of the aryepiglottic folds (except for the left fold in the voiceless trill, which dropped in frequency slightly compared to the “non-tense” 100 Hz trills). For the voiced trill of this set, it is speculated, based on the frequency values recorded in [Table II](#), that there is greater coupling between the glottal and aryepiglottic systems. With increased tension, the aryepiglottic folds could be more sensitive to the driving pressure generated by the glottal pulse, which is reflected in the fact that aryepiglottic fold frequencies almost match that of the glottal pulse for this trill.

There are two major theoretical questions of interest for this study. The first is whether the aryepiglottic aperture can



be considered a sound source. Both oscillographic and spectral evidence adduced in this study suggest that the aryepiglottic folds contribute to the harmonic spectrum of the signal. This is particularly evident for the voiceless trills where the EGG evidence indicates that the (near) periodic signal cannot be emanating from the glottis. The high degree of jitter in aryepiglottic fold oscillation, however, makes it difficult to clearly assess the exact harmonic content they are generating, particularly since both folds can oscillate at different frequencies, with major and minor pulsations, and in and out of phase with each other. Furthermore, both folds show open oscillations (where the aryepiglottic fold does not make complete contact with the epiglottis) and close oscillations, where collision with the posterior surface of the epiglottis might also generate acoustic energy. From the aperture analysis data and spectra, the frequency of the aryepiglottic folds was determined to tend toward a range between 50 and 100 Hz in the case of our subject.

While the voiced aryepiglottic trill poses a considerable challenge in determining what the exact contribution of the aryepiglottic fold oscillation to the overall laryngeal sound source is, the voiceless aryepiglottic trill provides some evidence that the folds do indeed contribute a pulsatile excitation to the vocal tract, suggesting that the folds can act as an alternative laryngeal source. For example, when the vocal folds are entirely absent (i.e., due to surgical resection of the vocal folds; Fröhlich *et al.*, 1998; Fröhlich *et al.*, 2000; Schwarz, 2007), the aryepiglottic folds function as the voicing source. The aryepiglottic source, however, appears far less efficient at converting the pulmonic airstream into acoustic energy than the vocal folds: The voiceless aryepiglottic trills were consistently less intense than the voiced trill, and the oscillograms and spectra both indicate that the acoustic signals these trills generate are highly noisy. This could be attributed to the fact that much of the air is escaping non-uniformly through one or both of the aryepiglottic apertures, in addition to the medial opening between the arytenoid cartilages, which is never fully sealed during any of the trills. Undoubtedly this amounts to considerable turbulent loss of aerodynamic power. The exact acoustic and aerodynamic nature of the turbulence generated by the aryepiglottic constriction is crucial to understanding pharyngeal fricatives and requires further study.

Throughout this study, the term trill has been used as a phonetic characterization of the oscillatory behavior of the aryepiglottic folds. This investigation reveals that this label of manner of articulation is appropriate for a number of reasons. While aryepiglottic trilling is unique due to the potential for two structures (the left and right aryepiglottic folds) to vibrate simultaneously and yet with considerably different characters (in terms of frequency, phase, and displacements), they arguably exhibit the same physiological conditions and restrictions of the other trilling structures, such as the tongue tip, uvula, and lips. Anatomically, trilling is claimed to be optimized when the tissue undergoing the trill is well supported by adjacent structures and forms a “funnel-like constriction, with the articulator opening along the direction of the airstream” (Barry, 1997, p. 36). Fundamentally, this configuration takes advantage of the Bernoulli effect when a

sufficient airstream is generated (Catford, 2001, p. 173). When the aryepiglottic folds are in their constricted configuration, their opening is indeed in the outward direction of the glottal airstream. The fact that all of the trills in this study were produced with considerable antero-posterior approximation of the epilaryngeal tube and with pharyngeal narrowing reflects the requirement for the aryepiglottic folds to be supported by adjacent structures. No aryepiglottic trilling has been attested, as far as the authors are aware, without these conditions holding (for example, in an entirely unconstricted state as occurs for modal voice). Furthermore, regardless of whether one or both of the folds are oscillating, the end acoustic product is uniformly describable as a trilling sound produced at the aryepiglottic place of articulation.

Aryepiglottic oscillations are also trill-like in the sense that there are no real intrinsic or specialized mechanisms for aryepiglottic pitch control. Vocal fold vibration can be controlled intrinsically with the contraction of the thyroarytenoid muscles and extrinsically relies on the highly specialized cricothyroid mechanism of pitch control. As this study demonstrates, the aryepiglottic folds are considerably limited in terms of their frequency range. The highest frequencies were observed for the tense trills, which indicates that general laryngeal constriction can alter the biomechanical nature of the folds to some degree, perhaps increasing the tension on the folds and resulting in a slightly increased oscillatory frequency (of about 20–40 Hz). Other trills are similar in this regard; they do not permit excessive manipulation of their frequency.

Despite these shared characteristics, the aryepiglottic trill must still be regarded as unique among trills because of its low position in the vocal tract. While it is possible to excite the resonances of the vocal tract using buccal or uvular trills, producing nearly unintelligible speech, aryepiglottic trilling is phonatory in the sense that it can be used as a quality of voicing (described as a category of *harsh voice*), and this function is incorporated into the linguistic systems of several languages, such as !Xóǀ (Tuu; Traill, 1986), N|uu (Tuu; Miller *et al.*, 2009), and Jianchuan Bai (Tibeto-Burman; Edmondson and Esling, 2006).

The aryepiglottic trilling documented here might also be compared to the phonation described for various ethnic singing styles of Asia, such as Mongolian throat singing. One of the important conclusions of Lindestad *et al.* (2001) is that a ventricular oscillation half-periodic to the glottal pulse acts to dampen the glottal source on every second glottal pulse. They regarded this mode of phonation as a complex sound source, where the ventricular damping is considered to contribute to a very rich harmonic spectrum that is characteristic of the singing style (which employs overtone enhancement). A similar alternating glottal-damping effect might occur for voiced aryepiglottic trill at 100 Hz (e.g., Fig. 6). In this particular trill, the aryepiglottic occlusion is not complete: Only the right aryepiglottic fold fully closes; the left aryepiglottic aperture, on the other hand, only reduces in area. Thus, the glottal damping in this case is not complete. Rather, once the right aryepiglottic aperture opens, the vocal folds destabilize (exhibiting a double peaked oscillation) and the acoustic signal becomes more noisy. Furthermore, the oscillatory behav-

ior of the aryepiglottic folds is generally more irregular than that of the ventricular folds, which appear to oscillate in sympathy with the vocal folds in Lindestad *et al.*'s (2001) data.

A note must be made about the question of how the acoustic impedances are changed in the vocal tract under aryepiglottic oscillation. As already described, under conditions of static laryngeal constriction, Titze's (2008) *non-linear source-filter coupling in phonation theory* predicts that the epilaryngeal tube contributes a positive reactance (inertance) to the vocal tract, which can both enhance the efficiency of glottal phonation as well as destabilize it. It is a point for future research to determine how this interaction might change under conditions of aryepiglottic trilling, where the epilaryngeal tube can potentially be separated into two outlet channels (the left and right apertures) that deform over time.

A significant challenge will be the simulation of aryepiglottic trilling in a computational model. One of the purposes of the present research is to contribute to the general understanding of aryepiglottic trilling with the goal of developing a computational simulation of the phenomenon. Much research remains, however, particularly with regard to understanding the acoustic nature of the trill and the aerodynamic conditions under which it occurs. Furthermore, there is little documentation available on the biomechanical properties of the aryepiglottic folds themselves. Computational modeling of the structures would greatly benefit from specific knowledge of their biomechanical characteristics and even their dimensions and mass.

## V. CONCLUSION

The purpose of this study has been to understand the oscillatory nature of aryepiglottic trilling. Evidence derived from image analysis of high-speed laryngoscopic videos of six trills with various glottal parameters indicates that the aryepiglottic folds can behave as an acoustic source to the vocal tract under conditions of glottal voicelessness, and perturb the vibration of the vocal folds by triggering alternation in the character of the glottal pulse. On account of this, the voiced aryepiglottic trill should be considered a complex sound source that modulates the quality of voicing and introduces noise as a result of prolonged glottal opening. The voiceless aryepiglottic trills are characterized by being much less intense than the voiced ones, although semi-harmonic excitation of the vocal tract occurs due to the aryepiglottic fold oscillations. Most of the acoustic power for the voiceless trills is converted into noise, which is characteristic of this sound. The aryepiglottic trill is deemed to be appropriately labeled as a trill primarily due its independence of the vocal folds and limited means of pitch control. It was found that increasing the "tension" of the overall laryngeal constriction used in producing the trill could increase the oscillating frequency of the aryepiglottic folds.

Future research must be carried out on other subjects and using linguistic data to determine the strength of these claims. Our next objective will be to analyze aryepiglottic trilling as it occurs in attested words of Iraqi Arabic with the

image analysis techniques used for this study. One of the ultimate goals is to create a computational simulation of aryepiglottic trilling. The data collected here will be used as a basis for such a model.

<sup>1</sup>Based on Esling (1996), the International Phonetic Alphabet symbols used to represent these sounds are as follows: stop [ʔ], fricative [ħ], approximant [ʕ], and trills [ʕʕ]. The pharyngeal tap [ʕ̤] is also attested.

<sup>2</sup>Lindqvist (1972) also argued that the aryepiglottic sphincter played an essential role in the production of creaky voice.

<sup>3</sup>A value of 1 indicates total positive correlation; in MATLAB this type of normalization is obtained by setting the `xcorr()` option flag to "coeff." The limiting value of 0.7071 was selected because it represents the zero-lag component of a cross-correlation analysis of two sine-waves that are exactly 45° out of phase.

- Barry, W. (1997). "Another [r]-tickle," *J. Int. Phonetic Assoc.* **27**, 35–45.
- Butcher, A., and Ahmad, K. (1987). "Some acoustic and aerodynamic characteristics of pharyngeal consonants in Iraqi Arabic," *Phonetica* **44**, 156–172.
- Carlson, B. F., Esling, J. H., and Harris, J. G. (2004). "A laryngoscopy phonetic study of Nlaka'pamux (Thompson) Salish glottal stop, glottalized resonants, and pharyngeals," in *Studies in Salish Linguistics in Honor of M. Dale Kinkade*, Occasional Papers in Linguistics Vol. **17**, edited by D. B. Gerdts and L. Matthewson (University of Montana Press, Missoula, MT), pp. 58–71.
- Catford, J. C. (1968). "The articulatory possibilities of man," in *Manual of Phonetics*, edited by B. Malmberg (North-Holland, Amsterdam), pp. 309–333.
- Catford, J. C. (2001). "On Rs, rhotacism and paleophony," *J. Int. Phonetic Assoc.* **31**, 171–185.
- Edmondson, J. A., and Esling, J. H. (2006). "The valves of the throat and their functioning in tone, vocal register and stress: Laryngoscopy case studies," *Phonology* **23**, 157–191.
- Edmondson, J. A., Esling, J. H., Harris, J. G., and Huang, T. (2005). "A laryngoscopic study of glottal and epiglottal/pharyngeal stop and continuant articulations in Amis—An Austronesian language of Taiwan," *Language and Linguistics* **6**, 381–396.
- Esling, J. H. (1996). "Pharyngeal consonants and the aryepiglottic sphincter," *J. Int. Phonetic Assoc.* **26**, 65–88.
- Esling, J. H. (2005). "There are no back vowels: The laryngeal articulator model," *Can. J. Ling.* **50**, 13–44.
- Esling, J. H., and Edmondson, J. A. (2002). "The laryngeal sphincter as an articulator: Tenseness, tongue root and phonation in Yi and Bai," in *Phonetics and Its Applications: Festschrift for Jens-Peter Köster on the Occasion of His 60th Birthday*, edited by A. Braun and H. R. Masthoff (Franz Steiner, Stuttgart), pp. 38–51.
- Esling, J. H., Fraser, K. E., and Harris, J. G. (2005). "Glottal stop, glottalized resonants, and pharyngeals: A reinterpretation with evidence from a laryngoscopic study of Nuuchahnulth (Nootka)," *J. Phonetics* **33**, 383–410.
- Fink, B. R. (1975). *The Human Larynx: A Functional Study* (Raven, New York).
- Fröhlich, M., Michaelis, D., and Strube, H. W. (1998). "Acoustic 'breathiness measures' in the description of pathologic voices," Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP 1998, Seattle, WA) **2**, 937–940.
- Fröhlich, M., Michaelis, D., Strube, H. W., and Kruse, E. (2000). "Acoustic voice analysis by means of the hoarseness diagram," *J. Speech Lang. Hear. Res.* **43**, 706–720.
- Gauffin, J. (1977). "Mechanisms of larynx tube constriction," *Phonetica* **34**, 307–309.
- Heselwood, B. (2007). "The 'tight approximant' variant of the Arabic 'ayn,'" *J. Int. Phonetic Assoc.* **37**, 1–32.
- Lindestad, P. A., Sodersten, M., Merker, B., and Granqvist, S. (2001). "Voice source characteristics in Mongolian 'throat singing' studied with high-speed imaging technique, acoustic spectra, and inverse filtering," *J. Voice* **15**, 78–85.
- Lindqvist, J. (1972). "A descriptive model of laryngeal articulation in speech," *Speech Transm. Lab. Q. Prog. Status Rep.*, Stockholm, KTH **13**, 1–9.
- Miller, A., Brugman, J., Sands, B., Namaseb, L., Exter, M., and Collins, C.

- (2009). "Differences in airstream and posterior place of articulation among N|uu clicks," *J. Int. Phonetic Assoc.* **39**, 129–161.
- O'Haver, T. (2006). "Peak finding and measurement. Custom scripts for the Matlab platform," <http://www.wam.umd.edu/~toh/spectrum/PeakFindingandMeasurement.htm> (Last viewed 6/10/2008).
- Sakakibara, K. I., Fuks, L., Imagawa, H., and Tayama, N. (2004). "Growl voice in pop and ethnic styles," Proceedings of the International Symposium on Musical Acoustics 2004.
- Schwarz, R. (2007). "Model-based quantification of pathological voice production," Ph.D. thesis, Universität Erlangen-Nürnberg, Erlangen, Germany.
- Stevens, K. (1998). *Acoustic Phonetics* (MIT, Cambridge, MA).
- Sundberg, J. (1977). "The acoustics of the singing voice," *Sci. Am.* **236**(3), 82–91.
- Švec, J. G., and Schutte, H. (1996). "Videokymography: High-speed line scanning of vocal fold vibration," *J. Voice* **10**, 201–205.
- Titze, I. R. (2008). "Nonlinear source-filter coupling in phonation: Theory," *J. Acoust. Soc. Am.* **123**, 2733–3749.
- Titze, I. R., and Story, B. H. (1997). "Acoustic interactions of the voice source with the lower vocal tract," *J. Acoust. Soc. Am.* **101**, 2234–2243.
- Traill, A. (1986). "The laryngeal sphincter as a phonatory mechanism in !Xóó (Bushman)," in *Variation, Culture and Evolution in African Populations: Papers in Honour of Dr. Hertha de Villiers*, edited by R. Singer and J. K. Lundy (Witwatersrand University Press, Johannesburg), pp. 123–131.