

Research Article

The Relationship Between Physiological Mechanisms and the Self-Perception of Vocal Effort

Victoria S. McKenna,^a Manuel E. Diaz-Cadiz,^a Adrianna C. Shembel,^b
Nicole M. Enos,^{c,d} and Cara E. Stepp^{a,c,e}

Purpose: This study aimed to examine the relationship between a large set of hypothesized physiological measures of vocal effort and self-ratings of vocal effort.

Method: Twenty-six healthy adults modulated speech rate and vocal effort during repetitions of the utterance /ifi/, followed by self-perceptual ratings of vocal effort on a visual analog scale. Physiological measures included (a) intrinsic laryngeal tension via kinematic stiffness ratios determined from high-speed laryngoscopy, (b) extrinsic suprahyoid and infrahyoid laryngeal tension via normalized percent activations and durations derived from surface electromyography, (c) supraglottal compression via expert visual-perceptual ratings, and (d) subglottal pressure via

magnitude of neck surface vibrations from an accelerometer signal.

Results: Individual statistical models revealed that all of the physiological predictors, except for kinematic stiffness ratios, were significantly predictive of self-ratings of vocal effort. However, a combined regression model analysis yielded only 3 significant predictors: subglottal pressure, mediolateral supraglottal compression, and the normalized percent activation of the suprahyoid muscles (adjusted $R^2 = .60$).

Conclusions: Vocal effort manifests as increases in specific laryngeal physiological measures. Further work is needed to examine these measures in combination with other contributing factors, as well as in speakers with dysphonia.

Vocal effort, defined as “perceived exertion” of the voice (Baldner, Doll, & van Mersbergen, 2015), has been reported in approximately 10% of older healthy adults (Merrill, Roy, & Lowe, 2013) and upward of 50% of speakers with voice disorders (Merrill et al., 2013; E. Smith et al., 1998). Excessive vocal effort is reported in speakers who fall into different etiologic categories, including vocal hyperfunction (Altman, Atkinson, & Lazarus, 2005; Roy, Merrill, Gray, & Smith, 2005), vocal fold paresis and paralysis (Bach, Belafsky, Wasylik, Postma, &

Koufman, 2005; Hartl, Hans, Vaissiere, Riquet, & Brasnu, 2001), spasmodic dysphonia (Cannito, Doiuchi, Murry, & Woodson, 2012; Isetti, Xuereb, & Eadie, 2014), and vocal fatigue from occupational voice demands (de Alvear, Baron, & Martinez-Arquero, 2011; E. Smith, Gray, Dove, Kirchner, & Heras, 1997). Yet, the relationship between the sensation of vocal effort and possible underlying physiological contributions has not been fully elucidated, leading to clinical ambiguity when trying to assess and remediate vocal effort in these speakers.

Because of its prevalence, researchers have focused on identifying possible structural, physiological, acoustical, and cognitive-emotional factors contributing to vocal effort. Thus far, respiratory and laryngeal aerodynamics (e.g., phonation threshold pressure) seem to be promising physiological predictors of vocal effort (Chang & Karnell, 2004; Sandage, Connor, & Pascoe, 2013; Solomon, Glaze, Arnold, & van Mersbergen, 2003), with further evidence that viscoelastic properties of the vocal folds (affected by both surface and systemic hydration) may contribute to self-perceived vocal effort as well (Solomon & DiMattia, 2000; Tanner et al., 2016; Verdolini et al., 2002; Verdolini, Titze, & Fennell, 1994). Vocal loading tasks that induce vocal fatigue have consistently reported increases in perceived

^aDepartment of Speech, Language, and Hearing Sciences, Boston University, MA

^bDepartment of Otolaryngology–Head and Neck Surgery, New York University Langone Medical Center and Voice Center, New York

^cDepartment of Biomedical Engineering, Boston University, MA

^dDepartment of Electrical and Computer Engineering, Boston University, MA

^eDepartment of Otolaryngology–Head and Neck Surgery, Boston University School of Medicine, MA

Correspondence to Victoria S. McKenna: vmckenna@bu.edu

Editor-in-Chief: Julie Liss

Editor: Michelle Ciucci

Received May 21, 2018

Revision received July 16, 2018

Accepted November 30, 2018

https://doi.org/10.1044/2018_JSLHR-S-18-0205

Disclosure: The authors have declared that no competing interests existed at the time of publication.

vocal effort, though simultaneous analysis of speech acoustics are inconclusive (for a review, see Fujiki & Sivasankar, 2017). Combined, these studies provide evidence for a working hypothesis that the sensation of vocal effort is derived from a compensatory physiological response to maintain the same vocal output (both in quality and intensity) in the presence of changes in vocal fold tissue properties and/or reduced muscular endurance (McCabe & Titze, 2002; Titze, 1999).

Despite the existing literature, it remains unclear exactly which physiological mechanisms are the primary factors driving self-reported vocal effort. Most studies have evaluated individual physiological measures in isolation, resulting in a lack of information on the comprehensive physiological profile of vocal effort. It follows that this study sought to examine multiple physiological contributions to self-perceived vocal effort with a specific focus on measures obtained from the laryngeal and paralaryngeal structures. The goal was to determine which measures of laryngeal and paralaryngeal function were most salient to the self-perception of vocal effort to streamline future research on vocal effort.

Proposed Physiological Mechanisms of Vocal Effort

At present, a series of physiological mechanisms specific to the laryngeal and paralaryngeal areas are reported to be associated with increased vocal effort, including increased intrinsic laryngeal tension, extrinsic laryngeal tension, supraglottal compression, and subglottal pressure. All of these mechanisms have been reported to increase when vocally healthy speakers purposefully increase effort and strain (Lien, Michener, Eadie, & Stepp, 2015; McKenna, Murray, Lien, & Stepp, 2016; Rosenthal, Lowell, & Colton, 2014; N. R. Smith et al., 2016) and are reported in speakers with voice disorders characterized by symptoms of excessive vocal effort (e.g., vocal hyperfunction).

Activation and tension in the intrinsic laryngeal muscles are necessary for voice production in healthy speakers (e.g., thyroarytenoid, cricothyroid; Y. Koike, 1967; Shipp, 1975). Changes in the timing and amplitude of intrinsic laryngeal muscle activation has been reported across modulations of voice onset types (Hirano, 1971; Hirose & Gay, 1973; Y. Koike, 1967) and in speakers with voice disorders (McCall, Colton, & Rabuzzi, 1973). These findings have led to the hypothesis that excessive tension in the intrinsic laryngeal muscles contributes to voice disorders and voice symptoms.

Because of the challenges of measuring intrinsic laryngeal tension (e.g., invasive nature of intramuscular electromyography [EMG], small size of the intrinsic laryngeal muscles), researchers have turned toward characterizing intrinsic laryngeal tension via indirect methodology. As such, *kinematic stiffness ratios*—derived from less invasive laryngoscopic images—have been investigated as a clinical correlate of intrinsic laryngeal tension (Cooke, Ludlow, Hallett, & Selbie, 1997; Dailey et al., 2005; McKenna et al., 2016; Munhall & Ostry, 1983; Stepp, Hillman, &

Heaton, 2010). Analysis via a one-joint virtual trajectory model revealed that increasing stiffness parameters in specific intrinsic laryngeal muscles (i.e., thyroarytenoid, posterior cricoarytenoid, and lateral cricoarytenoid) was strongly associated with increases in kinematic stiffness ratios (Stepp, Hillman, et al., 2010). Furthermore, kinematic stiffness ratios have tended to be smaller during typical voice productions and larger during modulations of hard glottal attack (Cooke et al., 1997) and vocal strain (McKenna et al., 2016). Despite these findings, current research is lacking an evaluation of the relationship between intrinsic laryngeal tension and the perception of vocal effort, leading to ambiguity in clinical targets for vocal therapy.

Similarly, excessive extrinsic laryngeal tension is hypothesized to contribute to dysphonia (Aronson, 1990), symptoms of laryngeal pain (Roy, Bless, Heisey, & Ford, 1997), and excessive vocal effort (McCabe & Titze, 2002). It follows that excessive extrinsic laryngeal muscle tension has been identified as a clinical marker of voice disorders (Angsuwarangsee & Morrison, 2002) and has been targeted diagnostically and therapeutically (e.g., laryngeal palpation, circumlaryngeal massage; Roy, Ford, & Bless, 1996; Roy & Leeper, 1993). Yet, studies seeking to quantitatively evaluate extrinsic laryngeal muscle activation patterns as a means for discriminating between voice types and health status (i.e., healthy speakers vs. those with voice disorders) have yielded conflicting results (Redenbaugh & Reich, 1989; N. R. Smith et al., 2016; Stepp, Heaton, Jette, Burns, & Hillman, 2010; Stepp et al., 2011; Van Houtte, Claeys, D'Haeseleer, Wuyts, & Van Lierde, 2013).

To date, few studies have directly examined the relationship between measures of extrinsic laryngeal tension and self-reported vocal effort. Dietrich and Abbott (2012) evaluated extrinsic muscle activation levels in healthy speakers under vocally stressful situations. Findings indicated that the percent activation of submental (i.e., suprahyoid muscles) and infrahyoid muscle groups significantly increased during stressful speaking situations and that self-perceived vocal effort also increased during stressful events. Therefore, increased tension in extrinsic laryngeal muscles may result in simultaneous increases in self-perceived vocal effort in some speakers.

Supraglottal compression, often referred to as supraglottal constriction, is due to increased constriction of the muscles superior to the glottis (e.g., ventricularis, aryepiglottic, thyroepiglottic; Kotby, Kirchner, Kahane, Basiouny, & el-Samaa, 1991; Moon & Alipour, 2013; Reidenbach, 1996, 1998; Sakakibara, Kimura, Imagawa, Niimi, & Tayama, 2004; Yanagisawa, Estill, Kmucha, & Leder, 1989). Although excessive supralaryngeal muscle activation and tension are postulated to be a clinical indicator of vocal hyperfunction (Lawrence, 1987; M. D. Morrison, Rammage, Belisle, Pullan, & Nichol, 1983; Sama, Carding, Price, Kelly, & Wilson, 2001), supraglottal compression is frequently reported in healthy speakers during typical speech (e.g., glottal stops; Pemberton et al., 1993; Stager, Bielamowicz, Regnell, Gupta, & Barkmeier, 2000) and singing (Guzman et al., 2016;

Pershall & Boone, 1987). More likely, the degree of supraglottal compression may be a relevant objective indicator of aberrant vocal behavior. For example, Stager et al. (2000) reported greater incidences of compression in both the mediolateral (M-L) and anterior–posterior (A-P) directions in speakers with vocal hyperfunction when compared to healthy speakers. Similarly, a greater degree of compression was noted in three out of four vocally healthy speakers in a small study that induced vocal fatigue (Solomon et al., 2003). The speakers also reported elevated amounts of vocal effort as well, though a direct analysis of the relationship between the two measures was not completed. The present study sought to add to the existing literature by evaluating the relationship between the degree of compression in both the A-P and M-L directions and the perception of vocal effort.

Finally, subglottal pressure, the pressure from the lungs that assists in initiating and maintaining vocal fold oscillation for phonation, has been implicated as one of the physiological mechanisms contributing to the self-perception of vocal effort (Colton, 1973). Subglottal pressure estimates, often captured indirectly via intraoral pressure estimates, are consistently elevated in speakers with nonphonotraumatic vocal hyperfunction (i.e., muscle tension dysphonia; Dastolfo, Gartner-Schmidt, Yu, Carnes, & Gillespie, 2016; Espinoza, Zanartu, Van Stan, Mehta, & Hillman, 2017; Hillman, Holmberg, Perkell, Walsh, & Vaughan, 1989; Zheng et al., 2012), phonotraumatic vocal hyperfunction (e.g., nodules; Dastolfo et al., 2016; Espinoza et al., 2017; Holmberg, Doyle, Perkell, Hammarberg, & Hillman, 2003; Kuo, Holmberg, & Hillman, 1999), and vocal fold lesions from glottic cancer (Friedman, Hillman, Landau-Zemer, Burns, & Zeitels, 2013; Zietels, Burns, Lopez-Guerra, Anderson, & Hillman, 2008). Increased subglottal pressure may be a strategy to initiate or maintain voicing in the presence of structural or functional changes to the larynx. There is further evidence that modulations of vocal effort in vocally healthy speakers affect subglottal pressure measures (Lien et al., 2015; McKenna, Llico, Mehta, Perkell, & Stepp, 2017; Rosenthal et al., 2014). Rosenthal et al. (2014) examined a series of aerodynamic measures across three voice conditions: comfortable voice, minimal vocal effort, and maximal vocal effort. Results indicated that subglottal pressure estimates were significantly different between all three voicing conditions, with the largest subglottal pressure measures found during maximal vocal effort productions. As such, self-perceived effort may be related to increases in subglottal pressure in both healthy speakers and speakers with voice disorders.

Research Questions and Hypotheses

The present investigation aimed to evaluate a large set of hypothesized physiological measures of vocal effort specific to the laryngeal and paralaryngeal regions and determine their relationship with self-perceived vocal effort. To our knowledge, no study has undertaken such a large

analysis, which has resulted in an inability to pinpoint which measures may be most relevant to the perception of vocal effort. We addressed the following research questions:

Research Question 1: What are the individual relationships between each physiological measure and the self-perceptual ratings of vocal effort?

Hypothesis 1: We hypothesized that each of the physiological measures gathered in this study would significantly predict self-perceptual ratings of vocal effort when analyzed in individual statistical models. We further suspected that each measure would have a positive relationship with self-perceptual ratings, meaning that as self-ratings of vocal effort increased, the physiological measure would also increase (e.g., measures of extrinsic tension would also increase).

Research Question 2: How do the relationships between the physiological measures and self-perceptual ratings of vocal effort change when analyzed in combination with one another?

Hypothesis 2: We hypothesized that there would be a reduction in the number of significant physiological predictors in the combined statistical model, in comparison to the individual models. We suspected that only one measure from each of the four hypothesized mechanisms would significantly predict self-perceptual ratings of vocal effort in the combined model.

Method

Participants

Twenty-six healthy adult participants, aged 18–29 years (16 women; $M = 20.9$ years, $SD = 2.8$ years), were recruited to Boston University for completion of the study. A greater number of women were enrolled in the study in order to be consistent with the estimates of the sex distribution of voice disorders for men and women (Roy et al., 2005). All participants were speakers of Standard American English with no history of speech, language, hearing, neurological, pulmonary, or voice disorders. They did not have any trained singing experience beyond grade school and were nonsmokers. All participants were screened for healthy vocal function via auditory–perceptual assessment and flexible nasendoscopic laryngeal imaging by a certified speech-language pathologist (SLP). A written consent, approved by the Boston University Institutional Review Board, was acquired from all participants before the start of the protocol.

Procedure

Participant Training

Participants were trained to produce vowel–consonant–vowel utterances of /ifi/. An /ifi/ string was two sets of four /ifi/ productions, resulting in eight total /ifi/ productions per string (e.g., /ifi ifi ifi ifi/, pause, /ifi ifi ifi ifi/). The combination of phonemes in the utterance /ifi/ provided the abductory and adductory vocal fold gestures needed to calculate kinematic estimates of laryngeal

stiffness and created an open pharyngeal configuration to better view the larynx during flexible nasendoscopy (McKenna et al., 2016). The participants were instructed to produce the /ifi/ strings at different vocal rates (slow, regular, and fast) and different levels of vocal effort (mild, moderate, and maximal) for a total of six voice conditions. A metronome was used to train vocal rate at three levels: Slow rate was at 50 beats per minute (bpm), regular rate was at 65 bpm, and fast rate was at 80 bpm. These targets were chosen because previous research has indicated that increased speech rate increases stiffness of oral articulators (Hertrich & Ackermann, 2000; Ostry & Munhall, 1985) and intrinsic laryngeal muscles (Stepp, Hillman, et al., 2010). Next, participants received instructions to vary their vocal effort using the following script: “Now we would like you to increase your effort during your speech as if you are trying to create tension in your voice as if you are trying to push your air out. Try to maintain the same volume while increasing your effort.” They were instructed to maintain their comfortable speaking rate and vocal volume. Mild effort was described as “mildly more effort than your regular speaking voice.” Moderate effort was described as “more effort than your mild effort,” and maximal effort was “as much effort as you can, while still having a voice.” Participants practiced these productions for approximately 10 min with a certified SLP to verify appropriate productions of rate and effort.

Participants were trained to make self-perceptual ratings of vocal effort on a horizontal 100-mm visual analog scale (VAS) after each /ifi/ string. The VAS is sensitive to small changes in self-perceived effort when analyzed within speaker and has been used with anchors at the extremes of the scale when rating vocal effort immediately following voice productions (Fujiki, Chapleau, Sundarajan, McKenna, & Sivasankar, 2017; Sandage et al., 2013; Solomon & DiMattia, 2000; Solomon et al., 2003; Sundarajan, Fujiki, Loerch, Venkatraman, & Sivasankar, 2017; Tanner et al., 2016). The left side of the scale was anchored with the description of “no effort at all” and labeled as *No Effort*, and the right side of the scale was anchored with the description of “the most effort you can imagine” and labeled as *Most Effort*. The labels *No Effort* and *Most Effort* were provided superior to the line at each end, whereas the numbers “0” and “100” were inferior to the line (placed on the left and right ends, respectively). A set of 13 lines were presented on a single sheet of paper and given to each participant on a clipboard. Participants were instructed to mark the scale with a single vertical line to indicate the amount of vocal effort employed in each /ifi/ string. No specific experiential anchors were provided, and no retraining was provided at the time of data acquisition.

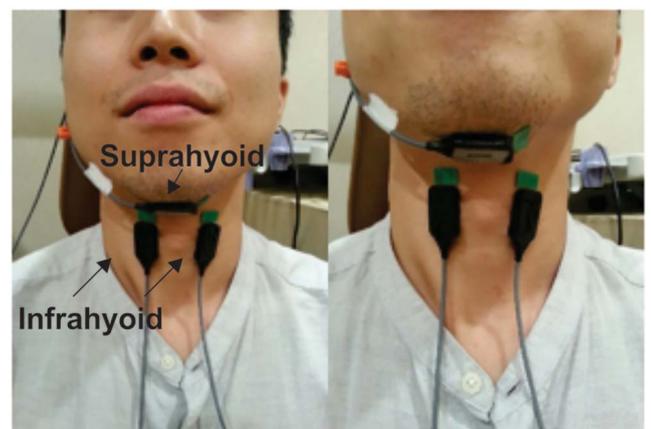
Experimental Setup and Calibration

Participants were seated throughout the study. First, three Delsys Bagnoli surface EMG (sEMG) sensors were affixed with adhesive tape to the anterior surface of the neck. Prior to sensor application, the skin on the anterior neck

was abraded and exfoliated (Stepp, 2012). A single-differential sensor was configured in the submandibular region, just posterior to the mandible. This sensor captured suprahyoid muscle activation of the mylohyoid muscles (and less so, the geniohyoid and the anterior belly of the digastrics due to variation in these muscle fiber orientations). Then, two double-differential sensors were placed approximately 1 cm to the right and left side of the thyroid prominence to target extrinsic infrahyoid muscles, including the thyrohyoids, omohyoids, and sternohyoids. A double-differential electrode was chosen to reduce conduction volume by minimizing cross-talk from the surface musculature common to all electrode contacts (Rutkove, 2007). Of note, the platysma muscle (a thin, superficial muscle of the neck) overlies all of the suprahyoid and infrahyoid muscles targeted in this study, meaning that the signals gathered from the surface electrodes may have also included platysma activity. A certified SLP identified and verified sensor locations via palpation during various tasks (e.g., hum, swallow). Figure 1 provides an example of sensor placement on the anterior neck. A ground electrode was placed on the acromion of the right shoulder to account for environmental and physiological noise (e.g., heartbeat). Participants completed a series of maximal voluntary contraction (MVC) tasks including neck flexion, saliva swallows, throat clears, and isometric contraction against resistance (see Appendix A). The MVCs were later used to normalize sEMG voltages.

Next, a BU series 21771 accelerometer (Knowles Electronic) was placed on the anterior neck with double-sided adhesive tape, superior to the thyroid notch, and inferior to the cricoid cartilage. A directional headset microphone (Shure SM35 XLR) was placed 45° from midline of the vermilion and 7 cm from the corner of the lips. In order to calculate sound pressure level (dB SPL) for all voice recordings, electrolaryngeal pulses were played at the lips while a sound pressure level meter measured dB SPL at the microphone. The known sound pressure levels were later used to calibrate the microphone recordings.

Figure 1. Example of surface EMG sensor placement on the submandibular region (suprahyoid) and anterior neck (infrahyoid).



Experimental Recordings

Once the training, equipment setup, and equipment calibration procedures were completed, a flexible pediatric endoscope (Pentax, Model FNL-7RP3, 2.4 mm) was passed transnasally over the soft palate into the hypopharynx to visualize the larynx. A numbing agent was not administered, so as not to affect laryngeal sensory feedback (Dworkin, Meleca, Simpson, & Garfield, 2000); however, a nasal decongestant was provided to decrease discomfort while the endoscope was passed through the nasal cavity. Participants completed a minimum of two recordings per condition (slow rate, regular rate, fast rate, mild effort, moderate effort, maximal effort), for a total of 12 recordings. If the endoscopist was unsure whether an adequate view of the vocal folds was appropriately captured, the condition was repeated. The need for repetition ultimately produced a total of seven extra recordings analyzed in the study (four slow rate, two regular rate, and one maximal effort). Immediately following each recording, participants completed a self-rating of their vocal effort on the 100-mm VAS. A paper version on a clipboard was used so that the participants could easily move the clipboard to a place they could see during the flexible endoscopy. The total time of the laryngoscopy was approximately 5–10 min, whereas the time leading up to the experimental recordings (including consent, training, equipment setup, and calibration) was approximately 1 hr.

The pediatric endoscope was attached to a FASTCAM Mini AX100 camera (Model 540K-C-16GB) operating at a resolution of 256×256 pixels with a 40-mm optical lens adapter. A steady xenon light was used for imaging (300 W KayPentax Model 7162B), and video images were acquired with Photron Fastcam Viewer software (Version 3.6.6). Because of the frame rate and memory capacities of the camera system, each recording was limited to 8 s in duration. The microphone and accelerometer signals were preamplified (Xenyx Behringer 802 Preamplifier) and digitized at 30 kHz with a data acquisition board (DAQ; National Instruments 6312 USB). Neck sEMG signals were acquired using a 16-channel Delsys Bagnoli EMG System (DS-160) and an analog bandpass filtered with roll-off frequencies of 20 and 450 Hz and a gain of 1,000. The sEMG signals were recorded at 30 kHz through the same DAQ as the microphone and accelerometer. Recordings were triggered via a custom MATLAB algorithm that time-aligned the endoscopic video images with the signals from the accelerometer, microphone, and sEMG sensors at the time of acquisition.

Postrecording Tasks

Following the experimental recordings, the endoscope and sEMG sensors were removed, leaving the accelerometer and headset microphone in place. Participants then completed a corresponding subglottal pressure task with the Phonatory Aerodynamic System (PAS; Model 6600, PENTAX Medical) to determine the relationship between intraoral estimates of subglottal pressure and measurements made from the accelerometer. Full details on the task and processing can be found in Appendix A. In brief, participants

produced a series of /pi/ repetitions while varying vocal effort. Correlations between intraoral estimations of subglottal pressure and simultaneous measurements made with the accelerometer revealed that all participants met the prespecified cutoff criteria ($r > .50$), verifying the accelerometer measure as an indirect indicator of subglottal pressure.

Measures

Microphone and Neck Surface Accelerometer

A semiautomated algorithm was developed to extract the root-mean-square (RMS) of each vowel for each individual /ifi/ production in the accelerometer signal, referred from here forward as NSV_{Mag} (i.e., the magnitude of neck surface vibration). The accelerometer signal was first full-wave rectified and filtered using a first-order low-pass Butterworth filter at 12 Hz. A threshold to distinguish voicing onset and offset was empirically determined during pilot testing as four times the mean amplitude of a 500-ms period of rest in the filtered signal. The RMS was calculated for the vowel segment between voicing onset and offset in the raw accelerometer signal, resulting in an NSV_{Mag} (V_{RMS}) value for each /ifi/ production. NSV_{Mag} values were then averaged across each recording.

The vowel segments were extracted from the corresponding microphone signal and calibrated to the sound pressure level gathered at the microphone with the sound pressure level meter. Sound pressure level was then averaged across each recording and is referred to as *mean SPL*. Because previous studies have reported strong correlations between sound pressure level and subglottal pressure estimates (Fryd, Van Stan, Hillman, & Mehta, 2016; Holmberg, Hillman, & Perkell, 1988; Lamarche & Ternstrom, 2008; Sundberg, Titze, & Scherer, 1993; Tanaka & Gould, 1983), we examined the variance inflation factors (VIF) between NSV_{Mag} and mean SPL when predicting ratings of vocal effort. Results revealed VIF values of less than 10, indicating no violation of multicollinearity (Hair, Anderson, Tatham, & Black, 1995) and no need to correct for mean SPL.

Neck sEMG

The sEMG signals were digitally bandpassed with a second-order Butterworth filter between 20 and 500 Hz to ensure minimization of ambient noise. Individual /ifi/ productions were segmented from each /ifi/ string. The initiation and termination of voicing were determined from the accelerometer signal (see Appendix A for more information on data processing), with the addition of a 250-ms prephonatory segment to account for muscle activity prior to the manifestation of voicing in the acoustic signal (Shipp, 1975; Stepp et al., 2011). Two target measures were then extracted for each recording: (a) percent activation, the activation amplitude compared to each sensor MVC, and (b) percent duration, the percentage of time the signal was “active” above a designated quiet rest level for each sensor.

Percent activation. The RMS of each individual /ifi/ segment was divided by the MVC value determined

for each sensor during the calibration procedure. The entire /ifi/ segment was chosen for analysis because muscle activation has been shown to increase at both the initiation and termination of voicing due to quick dynamic laryngeal movements for speech (Hirose & Gay, 1973; Sawashima, Kakita, & Hiki, 1973). The resulting activation value was a percentage of the possible maximum at each sensor for each /ifi/ segment. These were then averaged across each recording and are represented as a single percent activation per recording for each sensor placement.

Percent duration. First, sEMG signals were rectified and low-pass filtered at 12 Hz with a first-order Butterworth filter. A segment of quiet rest was extracted over 500 ms of recording during a slow rate production in which the participant was not voicing or swallowing (confirmed via the accelerometer signal). From the filtered rest signal, the mean and standard deviation of quiet rest were determined. We then empirically assessed a range of threshold values. The mean plus 3 *SDs* away from the mean was a sufficient threshold to minimize activation during quiet rest (i.e., in which no activity was occurring), yet still provide reasonable activation durations during /ifi/ segments that met a normal distribution without a ceiling effect. Finally, /ifi/ segments (also processed with the same specifications of the rest threshold) were directly compared to the rest thresholds for each speaker. This comparison resulted in a percent duration of each /ifi/ segment during which the sEMG signal was greater than the rest threshold. These percent durations were averaged over each recording for each sensor.

High-Speed Video Imaging

Kinematic stiffness ratios. High-speed videoendoscopic data were processed by trained technicians via a user-assisted algorithm. Appendix A provides a description of the training protocol and in-depth information on the user-assisted algorithm. In brief, the user-assisted algorithm determined vocal fold glottic angles extending from the anterior commissure along the medial vocal fold edge to the vocal process (see Figure 2a). The glottic angle was extracted over a series of images during the gross abductory and adductory gestures surrounding the /f/ in each /ifi/ production. The raw angles were plotted over time (McKenna et al., 2016; Stepp, Hillman, et al., 2010) and smoothed with a zero-phase 15th-order finite impulse response low-pass filter at 25 Hz. The maximum angular velocity during the adductory gesture was determined from the smoothed data within a range of 20%–80% of the maximum abductory angle (Dailey et al., 2005; Stepp, Hillman, et al., 2010) in order to minimize the effect of vibratory artifacts in the signal. The maximum angular velocity was then divided by the maximum abductory angle during the /f/ and reported as the kinematic stiffness ratio for each /ifi/ instance. These ratios were then averaged for each recording. Figure 2b provides a schematic of the raw angle waveform, the smoothed data, and determination of the maximum abductory angle and angular velocity.

Supraglottal compression ratings. A certified SLP was trained to complete supraglottal compression ratings (see Appendix A for training information) with the Voice-Vibratory Assessment With Laryngeal Imaging (VALI; Poburka, Patel, & Bless, 2017) in both the M-L direction (in which the false vocal folds compress medially to cover the true vocal folds) and the A-P direction (in which the distance between the arytenoids and petiole of the epiglottis is shortened). The VALI uses a 0–5 rating scale for M-L and A-P ratings, with 0 representing no constriction and a rating of 5 representing complete obstruction of the true vocal folds.

During data extraction, the SLP was blinded to participant and voice condition. First, the SLP watched a muted video of the entire recording to provide context into the relationship between the structures of the larynx. Then, separate ratings were made for A-P and M-L compression on images extracted ahead of time. Images were extracted from the midpoint time of each /i/ vowel in which the membranous portion of the vocal folds were adducted (± 10 ms from vowel center). The midpoint was chosen as it represents a more static compression that is due to overall glottal positioning, instead of dynamic, quick supraglottic articulatory actions that could occur during phonemic changes in running speech (Stager et al., 2000). Only one certified SLP was trained to complete compression ratings due to the large number of ratings (2,280 /ifi/ productions \times 2 /i/ vowels each \times 2 compression ratings = 9,120 total ratings). Ratings took approximately 2–3 hr per participant.

Interrater and intrarater reliability on the experimental data were completed on two randomly selected participants (i.e., 367 images, or 734 ratings), with the second rater being the first author of this article. The reliability was completed on averaged data for each recording as those were the data used in the experimental statistical analysis. Both raters were blinded to the participant, voice condition, and previous ratings. A two-way intraclass correlation coefficient (ICC) analysis for consistency revealed interrater reliability of ICC(2, 1) = .61 for M-L compression and ICC(2, 1) = .78 for A-P compression. Intrarater reliability analysis revealed an ICC(2, 1) = .56 and .71 for M-L and A-P compression, respectively.

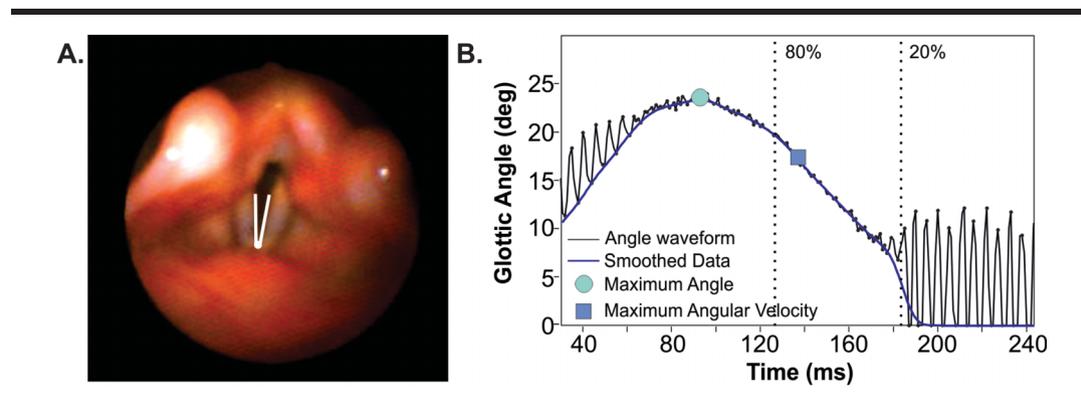
Self-Perceptual Ratings

Each self-perceptual rating on the 100-mm VAS was manually measured with a ruler and reported to the nearest millimeter. The measurements were made by a single technician and checked by the same technician at least 1 month after the first measurement. No discrepancies arose. The self-ratings were used as the outcome variable in this study.

Statistical Analysis

To be included in the final analysis, each physiological measure had to be calculated from at least three /ifi/ repetitions per recording (chosen as the slow rate productions

Figure 2. (a) View of the vocal folds under flexible laryngoscopy. The glottic angle has been marked from the anterior commissure to the vocal processes. (b) Raw vocal fold angles with smoothed data overlay. Maximum angle (circle), maximum angular velocity (square), and the range of 20%–80% of the maximum angle are identified.



often only had four /ifi/ repetitions). To evaluate our first hypothesis, individual mixed-effects regression models were completed. The independent variables were each physiological predictor and participant (random factor), whereas the dependent variable was the self-perceptual ratings of vocal effort (rated on the 100-mm VAS). Alpha level was set a priori to $p < .05$, and the adjusted coefficient of determination (adjusted R^2) was determined for each individual model. To evaluate our second hypothesis, a mixed-effect backward stepwise regression model was completed with participant (random factor) and all physiological predictors. The outcome variable of this combined model was the self-perception of vocal effort. Predictor variable significance was set to $p < .05$, and adjusted R^2 was determined for each model iteration. Variable effect sizes (η_p^2) were calculated for each significant predictor in the final model.

Results

Three hundred nineteen recordings were made with an average of 7.15 /ifi/ productions each and 2,280 /ifi/ repetitions in total. Across all recordings, there were eight instances in which there were fewer than three usable kinematic stiffness ratios available for averaging, accounting for approximately 2% of missing data points for that measure. These missing data points occurred across five different participants in the conditions of slow rate, regular rate, mild effort, moderate effort, and maximal effort. There was also one instance in which M-L compression and A-P compression could not be rated for any vowel in the high-speed video recording due to an unclear image (i.e., mucous on the endoscope). Therefore, the only missing data across all 10 physiological measures were these 10 instances, resulting in 3,181 total data points for analysis (319 recordings \times 10 measures – 10 missing data points).

All variables met the assumptions of the statistical models, except that the percent activation and percent duration measures of the left and right infrahyoid were too highly related to one another in the combined model (VIF > 10; Hair et al., 1995). In order to reduce multicollinearity,

the two sides were collapsed into a single measure by averaging the left and right values together. The measures are here forward referred to as *averaged percent activation of infrahyoids* and *averaged percent duration of infrahyoids*. Averaging these measures reduced the number of physiological predictors from 10 total measures to eight measures. Table 1 provides summary information of the mean and 95% confidence interval for each of the eight physiological measures and the self-ratings of vocal effort for each speaking condition.

Individual Mixed-Effects Regression Models

Individual statistical models revealed that all of the physiological predictors, except for kinematic stiffness ratios, significantly predicted vocal effort ratings. The strength of the predictions ranged from weak to moderate with adjusted R^2 values of .09–.53. Inspection of the beta coefficients revealed that all of the physiological measures in the study were positively associated with self-perceptual ratings of vocal effort. Results of the individual models can be found in Table 2.

Combined Mixed-Effects Backward Stepwise Regression Model

A mixed-effects backward stepwise regression model was calculated to analyze the relationship between the eight physiological measures and self-ratings of vocal effort. During the analysis, the variable with the largest p value was excluded in each iteration until all remaining variables met the criterion of $p < .05$. Results indicated that NSV_{Mag} , M-L compression, and percent activation of the suprahyoids were significant predictors of the self-perception of vocal effort. NSV_{Mag} had a large effect size, M-L compression had a medium effect size, and percent activation of the suprahyoids had a small effect size (Witte & Witte, 2010). The model accounted for 60% of the variance in self-ratings of vocal effort (adjusted $R^2 = .60$). Inspection of the beta coefficients revealed that all three predictors

Table 1. Mean and 95% confidence interval for each physiological measure and the self-rating of vocal effort for each condition.

Measure	Condition					
	Slow rate	Regular rate	Fast rate	Mild effort	Moderate effort	Maximal effort
Kinematic stiffness ratios (1/s)	14.35 [13.46, 15.24]	14.50 [13.74, 15.26]	15.29 [14.41, 16.17]	14.63 [13.78, 15.48]	14.70 [13.80, 15.59]	14.97 [14.09, 15.85]
Percent activation of suprahyoids	10.05 [9.95, 11.15]	10.01 [8.76, 11.26]	10.44 [9.20, 11.69]	10.38 [9.03, 11.73]	11.62 [10.03, 13.21]	14.07 [12.01, 16.13]
Percent duration of suprahyoids	68.52 [60.45, 76.60]	68.23 [61.01, 75.45]	72.40 [65.94, 78.85]	68.39 [60.06, 76.72]	74.17 [66.51, 81.84]	81.75 [75.38, 88.11]
Average percent activation of infrahyoids	9.59 [7.76, 11.42]	10.13 [8.13, 12.13]	10.15 [8.33, 11.96]	10.22 [8.35, 12.09]	10.92 [8.92, 12.92]	11.79 [9.55, 14.02]
Average percent duration of infrahyoids	68.40 [62.14, 74.67]	68.72 [61.88, 75.57]	72.34 [65.45, 79.23]	72.21 [65.74, 78.67]	73.94 [66.98, 80.91]	77.47 [71.26, 83.68]
A-P compression (0–5 scale)	2.20 [1.97, 2.43]	2.19 [1.96, 2.42]	2.25 [2.01, 2.50]	2.28 [2.06, 2.50]	2.43 [2.20, 2.66]	2.50 [2.27, 2.73]
M-L compression (0–5 scale)	1.57 [1.40, 1.75]	1.57 [1.42, 1.73]	1.71 [1.52, 1.91]	1.84 [1.63, 2.05]	1.95 [1.75, 2.15]	2.15 [1.95, 2.35]
NSV _{Mag} (V _{RMS})	0.14 [0.12, 0.15]	0.14 [0.12, 0.16]	0.15 [0.13, 0.17]	0.16 [0.14, 0.18]	0.21 [0.18, 0.23]	0.27 [0.23, 0.31]
Vocal effort rating (mm)	13.14 [10.64, 15.65]	14.69 [11.37, 18.00]	17.22 [12.95, 21.49]	26.09 [23.06, 29.11]	43.77 [40.02, 47.52]	68.17 [62.48, 73.85]

Note. A-P = anterior–posterior; M-L = mediolateral; NSV_{Mag} = magnitude of neck surface vibration.

increased as the self-perception of effort increased. A summary of significant findings can be found in Table 3, and the order of variable elimination and adjusted R^2 of each model can be found in Table 4.

Discussion

This study sought to evaluate physiological mechanisms of vocal effort, specific to the larynx, in order to determine which measures are the most salient to self-perceptual ratings of vocal effort. As expected, the majority of physiological measures significantly predicted self-ratings of vocal effort when analyzed in separate statistical models, with positive relationships with vocal effort ratings. Unexpectedly, the measure of intrinsic laryngeal tension was not a significant predictor in the individual

model, although the results approached significance ($p = .057$). Therefore, our first hypothesis was not supported because not all of the physiological measures proposed in this study were significantly predictive of self-ratings of vocal effort.

To address our second research question, we completed a combined statistical analysis to determine which measures were the most salient to vocal effort ratings. Importantly, when all eight physiological measures were analyzed in the initial combined regression model, the adjusted R^2 value was .61. After removal of nonsignificant measures (when the p value was greater than .05, see Table 4 for a review), the adjusted R^2 of the final model was a value of .60. These results indicate that removal of five of the physiological variables only reduced the model fit by 1%, further supporting the design of the study and the need for combined

Table 2. Results of mixed-effects regression models between individual physiological predictors and self-ratings of vocal effort.

Physiological measure	Beta Coef.	p	Adjusted R^2
Kinematic stiffness ratios (1/s)	1.1	.057	.09
Percent activation of suprahyoids	240.0	< .001*	.18
Percent duration of suprahyoids	46.0	< .001*	.15
Average percent activation of infrahyoids	159.8	.003*	.11
Average percent duration of infrahyoids	51.0	< .001*	.13
A-P compression (0–5 scale)	15.9	< .001*	.15
M-L compression (0–5 scale)	22.2	< .001*	.26
NSV _{Mag} (V _{RMS})	229.1	< .001*	.53

Note. Asterisks (*) are placed for significant predictors. Coef. = coefficient; A-P = anterior–posterior; M-L = mediolateral; NSV_{Mag} = magnitude of neck surface vibration.

Table 3. Summary of significant variables in the final mixed-effects backward stepwise regression statistical model.

Physiological measure	Beta Coef.	SE Coef.	t	p	Effect size (η_p^2)	Effect size interpretation
NSV _{Mag}	194.6	13.7	14.24	< .001	.41	Large
M-L compression	12.6	2.1	6.12	< .001	.11	Medium
Percent activation of suprahyoids	91.6	29.1	3.15	.002	.03	Small

Note. Coef = coefficient; SE = standard error NSV_{Mag} = magnitude of neck surface vibration; M-L = mediolateral.

statistical analysis in vocal effort research. The final model only yielded three significant predictors of vocal effort that fell across three of the four hypothesized mechanisms: subglottal pressure, supraglottal compression, and extrinsic laryngeal tension. Therefore, our second hypothesis was not supported as we had expected that one measure of each of the four proposed mechanisms would be significant.

The present findings support the supposition that speakers perceive effort partially based on sensory feedback from the laryngeal and paralaryngeal structures, which can be quantified via these physiological measures. Although the three significant predictors accounted for 60% of the variance in self-ratings of vocal effort, there continued to be 40% of unaccounted for variance. It is quite possible that other physiological (e.g., respiratory, articulatory), acoustical (i.e., auditory feedback), and cognitive-emotional factors also contribute to the self-perception of vocal effort. For example, there is beginning evidence that personality factors may contribute to the development and persistence of voice problems (Roy & Bless, 2000; Roy, Bless, & Heisey, 2000) and that negative emotional valence (induced via visual imagery or stressful environments) directly increase perceived vocal effort in some speakers (Dietrich & Abbott, 2012; van Mersbergen, Patrick, & Glaze, 2008). In order to elucidate a more comprehensive profile of vocal effort, we suggest that future researchers consider a combined analysis of the significant physiological predictors identified in our study results with additional factors that may contribute to self-perceived vocal effort.

Physiological Measures of Vocal Effort

The working hypothesis of McCabe and Titze (2002) proposed that the perception of vocal effort is partially based on physiological changes that are supposed to

maintain or improve the intensity and quality of the voice. With that in mind, we suspect that each of the significant physiological predictors could be a manifestation of a strategy to improve the acoustical output of each speaker.

First, the physiological predictor with the largest effect size was our indirect measure of subglottal pressure. A laryngeal model by Zanartu et al. (2014) indicated that greater posterior glottal gap size yielded a reduction of energy transfer to the vocal folds and a simultaneous reduction in sound pressure level. When a compensatory model was created that specifically increased subglottal pressure, there was also an increase in sound pressure level. Based on this work, it is likely that subglottal pressure is a strategy to increase the amplitude of vocal fold vibration and thereby increase sound pressure level (i.e., vocal intensity).

It is then unsurprising that elevated subglottal pressure would be consistently documented across a wide variety of speakers with voice disorders, including those stemming from both structural and functional causes. Subglottal pressure may be elevated in speakers with vocal fold lesions such as phonotraumatic vocal hyperfunction (Dastolfo et al., 2016; Espinoza et al., 2017; Holmberg et al., 2003; Kuo et al., 1999) and glottic cancer (Friedman et al., 2013; Zietels et al., 2008), as well as in speakers without lesions, such as nonphonotraumatic vocal hyperfunction (Dastolfo et al., 2016; Espinoza et al., 2017; Hillman et al., 1989) and glottal incompetence (e.g., unilateral vocal fold paralysis, vocal fold atrophy; Dastolfo et al., 2016). It is likely that an increase in subglottal pressure is a universal strategy to compensate for a reduction in vocal abilities, regardless of reason for reduced vocal abilities (e.g., nodules, vocal fold paresis). Whether the increased amplitude of vibration then increases collision forces and results in additional contact stress of the vocal folds (possibly leading to phonotrauma) is still an area of continued investigation.

Table 4. Order of variable elimination during backward stepwise regression analysis.

Order of removal	Variable removed	p	Model adjusted R ²
1	Average percent activation of infrahyoids	.946	.61
2	Percent duration of suprahyoids	.889	.61
3	A-P compression	.710	.61
4	Average percent duration of infrahyoids	.171	.61
5	Kinematic stiffness ratios	.162	.61

Note. A-P = anterior-posterior.

In the same vein, M-L supraglottal compression may be a physiological strategy to increase vocal fold contact and the amplitude of the acoustic signal via effects on subglottal pressure and glottal resistance. This strategy has been reported elsewhere, specifically in speakers with glottal incompetence due to vocal fold paralysis or paresis. In a study by Bielamowicz, Kapoor, Schwartz, and Stager (2004), speakers with greater M-L compression had airflow, pressure, and acoustical measures that were within normal ranges. The authors concluded that M-L compression reduced glottal gap size and improved phonatory function.

Measures of M-L compression continue to be quantified subjectively, and although that allows for direct clinical translation, it also increases the possibility of rater variability and error. The interrater reliability of M-L compression ratings reported here are somewhat lower than other studies that used high-speed videoendoscopy (Parker, Kunduk, Fink, & McWhorter, 2017; Poburka et al., 2017); however, these other studies were completed using rigid endoscopy with a protruded tongue and sustained vowel, possibly impacting laryngeal configuration and improving rating reliability. Even with the increased variability of our ratings, M-L compression was a significant predictor of vocal effort with a medium effect size. These results suggest that the degree of M-L compression is a robust physiological indicator of vocal effort and that measures can be obtained during more natural speech contexts. M-L compression is, therefore, a promising metric of vocal effort and should continue to be investigated across speakers with primary symptoms of vocal effort.

Finally, percent activation of extrinsic suprahyoid muscles significantly predicted self-ratings of vocal effort in the combined analysis. Interpretation of these results in the context of the working hypothesis would indicate that increased extrinsic laryngeal tension is a compensatory behavior to change vocal output. In vocally healthy speakers, the suprahyoid muscles act to pull the hyoid bone anteriorly, which subsequently raises and rotates the thyroid cartilage (Honda, Hirai, Masaki, & Shimada, 1999). This configuration changes laryngeal muscle tension (both active and passive) and impacts vocal fold vibratory characteristics (Shipp, 1975; Sundberg & Askenfelt, 1981). Therefore, an increase in suprahyoid tension could result in vocal quality changes and be a strategy to change output in some speakers. However, it is thought that some voice disorders develop because speakers continue to use a compensatory voicing strategy when it is no longer needed (Hillman et al., 1989). The persistence of a strategy (in this case, tension of the suprahyoid muscles to increase vocal effort) could ultimately become maladaptive. For example, Lowell, Kelley, Colton, Smith, and Portnoy (2012) determined that speakers with nonphonotraumatic vocal hyperfunction exhibited significantly higher positioning of the hyoid bone and larynx during phonation when directly compared to vocally healthy speakers. More work is needed to determine when increased extrinsic laryngeal muscle tension is compensatory and how it may persist in some speakers with voice disorders.

Previous literature has shown the role of suprahyoid and infrahyoid muscle groups in the context of voice pathology to be variable. The fact that suprahyoid activation was a significant predictor of vocal effort whereas infrahyoid measures were not is an interesting finding in this study. This could be because sEMG sensors only provide information on active, phasic isotonic muscle contractions and are less apt at identifying passive tonic isometric muscle activation that can result from muscle stretch. As such, increased activation of the suprahyoid muscles could have passively stretched infrahyoid muscles, limiting the utility of sEMG to assess tension in these muscles. An alternative would be to use manual palpation of the laryngeal structures to assess extrinsic laryngeal tension (Aronson, 1990; M. Morrison, 1997; Roy, 2008; Roy et al., 1997, 1996; Roy & Leeper, 1993); however, it remains unclear whether manual palpation techniques are sensitive to small degrees of change in muscle tension (e.g., there was a 4% change noted from the regular speaking rate to the maximal effort condition for suprahyoid activation). For these reasons, manual palpation and sEMG in conjunction may provide the most accurate assessment of extrinsic laryngeal muscle tension in speakers with voice disorders and should be investigated further in vocal effort research.

Vocal Rate, Effort, and Intrinsic Laryngeal Tension

We designed this study to incorporate modulations of vocal rate and vocal effort in an attempt to create variation in self-perceived vocal effort. Examination of averaged data across the different voicing conditions revealed that self-ratings of effort were relatively consistent across variation in vocal rate (range of 13.14–17.22 on the 100-mm VAS). This was a surprising finding because previous studies reported that increased speech rate acts to increase tension in both oral articulators (Hertrich & Ackermann, 2000; Ostry & Munhall, 1985) and laryngeal muscles (Stepp, Hillman, et al., 2010), but it seems that the participants in this study did not perceive this tension as increased effort.

The fact that perceived effort did not increase during modulations of rate likely contributed to the nonsignificant findings between self-perceptual ratings of vocal effort and the measure of intrinsic laryngeal tension (kinematic stiffness ratio) leading to the rejection of both of the hypotheses laid out at the beginning of the study. As expected, kinematic stiffness ratios tended to increase as speech rate increased (Stepp, Hillman, et al., 2010), with the greatest stiffness ratios produced during the fast vocal rate, but kinematic stiffness ratios remained relatively unchanged during tasks that increased effort (range of 14.63–14.97 1/s). These ratios are inconsistent with the trends seen in ratings of vocal effort. We had theorized speakers would be able to perceive tension from sensorimotor feedback, possibly from muscle spindles in the intrinsic laryngeal muscles (S. Koike, Mukudai, & Hisa, 2016). However, the conscious perception of muscle tension via sensory feedback

from these spindles is still debated (Ludlow, 2005). It may be that intrinsic laryngeal muscle spindles affect unconscious, reflexive responses to muscle stretch and tension. To date, few studies have evaluated reflexive responses in the intrinsic (and extrinsic) laryngeal muscles in human subjects (Loucks, Poletto, Saxon, & Ludlow, 2005; Sapir, Baker, Larson, & Ramig, 2000), making it difficult to determine the role of sensory feedback from intrinsic laryngeal muscles in self-perceived voice symptoms.

Participant Variability

In order to evaluate individual variability of the participants in this study, per-participant Pearson product-moment correlations were calculated for all physiological variables. Appendix B provides a list of all individual correlations and the percentage of speakers who revealed strong relationships, per a cutoff criterion of $r \geq .70$. Only two of the 26 participants (approximately 8%) had strong correlations across all three physiological measures with their self-ratings of vocal effort. This provides evidence that these physiological events can act in isolation of one another and, furthermore, that the likelihood of an individual incorporating all three physiological strategies while increasing vocal effort is quite low.

Participant variability could be due to individual physiological preferences that result in a primary or “dominant” mechanism to increase effort. In a simple comparison of correlations within each speaker, 16 of the 26 participants (approximately 62%) exhibited the strongest correlations between effort ratings and the subglottal pressure measure (NSV_{Mag}), four speakers (approximately 15%) had their strongest correlations between effort ratings and M-L compression, and two speakers (approximately 8%) had the strongest correlations between effort ratings and percent activation of the suprahyoid muscles. Thus, the physiological profiles of the majority of speakers would fall into a category of subglottal pressure dominant as their primary contributor to vocal effort. Based on the findings of participant variability, we recommend that clinicians continue to investigate all three of the significant physiological predictors identified in this study (i.e., subglottal pressure, M-L compression, suprahyoid activation), as these were the dominant physiological mechanisms in approximately 88% of the speakers enrolled in the study.

Limitations and Future Directions

This study employed indirect estimation techniques for all physiological measures. It is widely believed that direct estimates may provide more accurate measurements of the mechanisms that underlie different physiological systems; however, direct physiological measures of the larynx often require techniques that are invasive (e.g., intramuscular EMG, tracheal puncture). The measures described in this article are less invasive and more clinically feasible, providing benefits for translation of the present work to the clinical setting. Still, it is possible that these indirect measures

may be more variable or less related to direct measurements taken during aberrant voice productions. For example, neck surface accelerometry is an exciting new prospect for estimation of specific aerodynamic parameters during more natural speech contexts (Mehta et al., 2015), but NSV_{Mag} has yet to be fully vetted against direct subglottal pressure estimates in speakers. Therefore, the methods described in this work, in relation to spontaneous speech and natural prosody, require further investigation before clinical application.

The present work was based on the presumption that vocal effort can be reliability quantified via a 100-mm VAS. A simple VAS with anchors at each end has been used to measure self-perceptual ratings of vocal effort, and those ratings have been associated with physiological measures (Solomon et al., 2003; Tanner et al., 2016). It should be noted that the perceptual literature is replete with examples of how the type of scale, the use of anchors and definitions, and the amount of training provided can influence perceptual ratings. Despite this knowledge, there is currently no universally accepted scale to subjectively quantify vocal effort, leading to ambiguity in choosing the correct scale for research and clinical purposes. To date, self-perceptual rating scales of vocal effort include a simple 0–10 equal interval scale (Hunter & Titze, 2009; McCabe & Titze, 2002), direct magnitude estimation (Dietrich & Abbott, 2012; Sivasankar & Fisher, 2002; Verdolini et al., 1994; Verdolini-Marston, Burke, Lessac, Glaze, & Caldwell, 1995), and the Borg Category Ratio 10 (Herndon, Sundarajan, Sivasankar, & Huber, 2017; Steinhauer, Grayhack, Smiley-Oyen, Shaiman, & McNeil, 2004; van Mersbergen et al., 2008). The Borg Category Ratio 10 is derived from a scale that was originally developed in the exercise physiology literature to characterize physical exertion (Borg, 1982; Neely, Ljunggren, Sylven, & Borg, 1992; Noble, Borg, Jacobs, Ceci, & Kaiser, 1983) and is undergoing psychometric evaluation as a perceptual measure of vocal effort (Baldner et al., 2015; van Leer & van Mersbergen, 2017). As these scales continue to be developed, the field would benefit from consensus on a self-perceptual scale of vocal effort in order to make effort ratings comparable across studies.

Finally, further work is needed to examine physiological manifestations of vocal effort in speakers with voice disorders. Although the physiological mechanisms reported in this study have been shown to be elevated in speakers with primary symptoms of vocal effort, it is not clear how vocal effort may manifest in different etiologic groups. Specifically, it remains unknown whether modulations of vocal effort in healthy speakers are applicable to those with structural changes (e.g., nodules), neurological-based voice disorders (e.g., spasmodic dysphonia), or functional-based dysphonia. Moreover, our study elicited increased vocal effort via instructions to do so, instead of employing a vocally fatiguing task, such as vocal loading. It is possible that the physiological measures associated with elevated vocal effort in healthy speakers may be different following a more challenging vocal task or during instances of concurrent vocal fatigue.

Conclusion

Vocal effort manifests as a combination of physiological mechanisms, including increases in indirect measurements of subglottal pressure, M-L supraglottal compression, and activation of extrinsic suprahyoid muscles. These mechanisms could be compensatory strategies to improve vocal fold vibration amplitude and improve the acoustical signal; however, exactly how these mechanisms play a role in individuals with voice disorders and how they interact with physiological, acoustical, and cognitive-emotional factors warrant further investigation. A better understanding of the mechanisms driving clinical presentations of voice disorders would improve diagnostic and therapeutic approaches to individuals with primary symptoms of excessive vocal effort.

Acknowledgments

This work was supported by National Institute on Deafness and Other Communication Disorders Grants R01DC015570 (awarded to C. E. S.), T32DC013017 (awarded to C. A. M.), and F31DC015752 (awarded to A. C. S.). It was also supported by a Sargent College Dudley Allen Research Grant (awarded to V. S. M.) and the Undergraduate Research Opportunity Grant (awarded to N. M. E.) from Boston University. We would like to thank Daniel Buckley, Jacob Noordzij, Lin Zhang, Jaime Kim, Hasini Weerathunge, and Dante Cilento for their assistance with data processing.

References

- Altman, K. W., Atkinson, C., & Lazarus, C. (2005). Current and emerging concepts in muscle tension dysphonia: A 30-month review. *Journal of Voice*, *19*(2), 261–267.
- Angsuwarangsee, T., & Morrison, M. (2002). Extrinsic laryngeal muscular tension in patients with voice disorders. *Journal of Voice*, *16*(3), 333–343.
- Aronson, A. E. (1990). *Clinical voice disorders: An interdisciplinary approach* (3rd ed.). New York, NY: Thieme.
- Bach, K. K., Belafsky, P. C., Wasylik, K., Postma, G. N., & Koufman, J. A. (2005). Validity and reliability of the glottal function index. *Archives of Otolaryngology—Head & Neck Surgery*, *131*(11), 961–964.
- Baldner, E. F., Doll, E., & van Mersbergen, M. R. (2015). A review of measures of vocal effort with a preliminary study on the establishment of a vocal effort measure. *Journal of Voice*, *29*(5), 530–541.
- Bielamowicz, S., Kapoor, R., Schwartz, J., & Stager, S. V. (2004). Relationship among glottal area, static supraglottic compression, and laryngeal function studies in unilateral vocal fold paresis and paralysis. *Journal of Voice*, *18*(1), 138–145.
- Borg, G. A. (1982). Psychophysical bases of perceived exertion. *Medicine & Science in Sports & Exercise*, *14*(5), 377–381.
- Cannito, M. P., Doiuchi, M., Murry, T., & Woodson, G. E. (2012). Perceptual structure of adductor spasmodic dysphonia and its acoustic correlates. *Journal of Voice*, *26*(6), 818.e5–818.e13.
- Chang, A., & Karnell, M. P. (2004). Perceived phonatory effort and phonation threshold pressure across a prolonged voice loading task: A study of vocal fatigue. *Journal of Voice*, *18*(4), 454–466.
- Colton, R. H. (1973). Some relationships between vocal effort and intraoral air pressure. *The Journal of the Acoustical Society of America*, *53*(1), 296.
- Cooke, A., Ludlow, C. L., Hallett, N., & Selbie, W. S. (1997). Characteristics of vocal fold adduction related to voice onset. *Journal of Voice*, *11*(1), 12–22.
- Dailey, S. H., Kobler, J. B., Hillman, R. E., Tangrom, K., Thananart, E., Mauri, M., & Zeitels, S. M. (2005). Endoscopic measurement of vocal fold movement during adduction and abduction. *Laryngoscope*, *115*(1), 178–183.
- Dastolfo, C., Gartner-Schmidt, J., Yu, L., Carnes, O., & Gillespie, A. I. (2016). Aerodynamic outcomes of four common voice disorders: Moving toward disorder-specific assessment. *Journal of Voice*, *30*(3), 301–307.
- de Alvear, R. M. B., Baron, F. J., & Martinez-Arquero, A. G. (2011). School teachers' vocal use, risk factors, and voice disorder prevalence: Guidelines to detect teachers with current voice problems. *Folia Phoniatrica Et Logopaedica*, *63*(4), 209–215.
- Dietrich, M., & Abbott, K. V. (2012). Vocal function in introverts and extraverts during a psychological stress reactivity protocol. *Journal of Speech, Language, and Hearing Research*, *55*(3), 973–987.
- Dworkin, J. P., Meleca, R. J., Simpson, M. L., & Garfield, I. (2000). Use of topical lidocaine in the treatment of muscle tension dysphonia. *Journal of Voice*, *14*(4), 567–574.
- Espinoza, V. M., Zanartu, M., Van Stan, J. H., Mehta, D. D., & Hillman, R. E. (2017). Glottal aerodynamic measures in women with phonotraumatic and nonphonotraumatic vocal hyperfunction. *Journal of Speech, Language, and Hearing Research*, *60*(8), 2159–2169.
- Friedman, A. D., Hillman, R. E., Landau-Zemer, T., Burns, J. A., & Zeitels, S. M. (2013). Voice outcomes for photoangiolytic KTP laser treatment of early glottic cancer. *Annals of Otology, Rhinology & Laryngology*, *122*(3), 151–158.
- Fryd, A. S., Van Stan, J. H., Hillman, R. E., & Mehta, D. D. (2016). Estimating subglottal pressure from neck-surface acceleration during normal voice production. *Journal of Speech, Language, and Hearing Research*, *59*(6), 1335–1345.
- Fujiki, R. B., Chapleau, A., Sundarajan, A., McKenna, V., & Sivasankar, M. P. (2017). The interaction of surface hydration and vocal loading on voice measures. *Journal of Voice*, *31*(2), 211–217.
- Fujiki, R. B., & Sivasankar, M. P. (2017). A review of vocal loading tasks in the voice literature. *Journal of Voice*, *31*(3), 388.e33–388.e39.
- Guzman, M., Ortega, A., Olavarria, C., Munoz, D., Cortes, P., Azocar, M. J., . . . Silva, C. (2016). Comparison of supraglottic activity and spectral slope between theater actors and vocally untrained subjects. *Journal of Voice*, *30*(6), 761–767.
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). *Multivariate data analysis* (3rd ed.). New York, NY: Macmillan.
- Hartl, D. M., Hans, S., Vaissiere, J., Riquet, M., & Brasnu, D. F. (2001). Objective voice quality analysis before and after onset of unilateral vocal fold paralysis. *Journal of Voice*, *15*(3), 351–361.
- Herndon, N. E., Sundarajan, A., Sivasankar, M. P., & Huber, J. E. (2017). Respiratory and laryngeal function in teachers: Pre- and postvocal loading challenge. *Journal of Voice*. Advance online publication. <https://doi.org/10.1016/j.jvoice.2017.11.015>
- Hertegard, S., Gauffin, J., & Lindestad, P. (1995). A comparison of subglottal and intraoral pressure measurements during phonation. *Journal of Voice*, *9*(2), 149–155.
- Hertrich, I., & Ackermann, H. (2000). Lip-jaw and tongue-jaw coordination during rate-controlled syllable repetitions. *The Journal of the Acoustical Society of America*, *107*(4), 2236–2247.
- Hillman, R. E., Holmberg, E. B., Perkell, J. S., Walsh, M., & Vaughan, C. (1989). Objective assessment of vocal hyperfunction: An experimental framework and initial results. *Journal of Speech, Language, and Hearing Research*, *32*(2), 373–392.

- Hirano, M. (1971). Laryngeal adjustment for different vocal onsets. An electromyographic investigation. *Journal of Otolaryngology Japan*, 74, 1572–1579.
- Hirose, H., & Gay, T. (1973). Laryngeal control in vocal attack. An electromyographic study. *Folia Phoniatrica et Logopaedica*, 25(3), 203–213.
- Holmberg, E. B., Doyle, P., Perkell, J. S., Hammarberg, B., & Hillman, R. E. (2003). Aerodynamic and acoustic voice measurements of patients with vocal nodules: Variation in baseline and changes across voice therapy. *Journal of Voice*, 17(3), 269–282.
- Holmberg, E. B., Hillman, R. E., & Perkell, J. S. (1988). Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice. *The Journal of the Acoustical Society of America*, 84(2), 511–529.
- Holmberg, E. B., Perkell, J. S., & Hillman, R. E. (1984). *Methods for using a noninvasive technique for estimating glottal functions from oral measurements*. Paper presented at the 107th Meeting: Acoustical Society of America.
- 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.
- Hunter, E. J., & Titze, I. R. (2009). Quantifying vocal fatigue recovery: Dynamic vocal recovery trajectories after a vocal loading exercise. *Annals of Otolaryngology, Rhinology & Laryngology*, 118(6), 449–460.
- Isetti, D., Xuereb, L., & Eadie, T. L. (2014). Inferring speaker attributes in adductor spasmodic dysphonia: Ratings from unfamiliar listeners. *American Journal of Speech-Language Pathology*, 23(2), 134–145.
- Koike, S., Mukudai, S., & Hisa, Y. (2016). Muscle spindles and intramuscular ganglia. In Y. Hisa (Ed.), *Neuroanatomy and neurophysiology of the larynx* (pp. 11–20). Tokyo, Japan: Springer.
- Koike, Y. (1967). Experimental studies on vocal attack. *Practica Otologica Kyoto*, 60, 663–688.
- Kotby, M. N., Kirchner, J. A., Kahane, J. C., Basiouny, S. E., & el-Samaa, M. (1991). Histo-anatomical structure of the human laryngeal ventricle. *Acta Oto-Laryngologica*, 111(2), 396–402.
- Kuo, J., Holmberg, E. B., & Hillman, R. E. (1999). *Discriminating speakers with vocal nodules using aerodynamic and acoustic features*. Paper presented at the IEEE International Conference on Acoustics, Speech, and Signal Processing, Phoenix, AZ.
- Lamarche, A., & Ternstrom, S. (2008). An exploration of skin acceleration level as a measure of phonatory function in singing. *Journal of Voice*, 22(1), 10–22.
- Lawrence, V. L. (1987). *Suggested criteria for fibre-optic diagnosis of vocal hyperfunction*. Paper presented at the Care of the Professional Voice Symposium, London.
- Lien, Y. A., Michener, C. M., Eadie, T. L., & Stepp, C. E. (2015). Individual monitoring of vocal effort with relative fundamental frequency: Relationships with aerodynamics and listener perception. *Journal of Speech, Language, and Hearing Research*, 58(3), 566–575.
- Loucks, T. M. J., Poletto, C. J., Saxon, K. G., & Ludlow, C. L. (2005). Laryngeal muscle responses to mechanical displacement of the thyroid cartilage in humans. *Journal of Applied Physiology*, 99(3), 922–930.
- Lowell, S. Y., Kelley, R. T., Colton, R. H., Smith, P. B., & Portnoy, J. E. (2012). Position of the hyoid and larynx in people with muscle tension dysphonia. *Laryngoscope*, 122(2), 370–377.
- Ludlow, C. L. (2005). Central nervous system control of the laryngeal muscles in humans. *Respiratory Physiology & Neurobiology*, 147(2–3), 205–222.
- McCabe, D. J., & Titze, I. R. (2002). Chant therapy for treating vocal fatigue among public school teachers: A preliminary study. *American Journal of Speech-Language Pathology*, 11(4), 356–369.
- McCall, G. N., Colton, R. H., & Rabuzzi, D. D. (1973). Preliminary EMG investigation of certain intrinsic and extrinsic laryngeal muscles in patients with spasmodic dysphonia. *The Journal of the Acoustical Society of America*, 53(1), 345.
- McKenna, V. S., Llico, A. F., Mehta, D. D., Perkell, J. S., & Stepp, C. E. (2017). Magnitude of neck-surface vibrations as an estimate of subglottal pressure during modulations of effort and intensity in healthy speakers. *Journal of Speech, Language, and Hearing Research*, 60, 3404–3416.
- McKenna, V. S., Murray, E. S. H., Lien, Y. A. S., & Stepp, C. E. (2016). The relationship between relative fundamental frequency and a kinematic estimate of laryngeal stiffness in healthy adults. *Journal of Speech, Language, and Hearing Research*, 59(6), 1283–1294.
- Mehta, D. D., Van Stan, J. H., Zanartu, M., Ghassemi, M., Guttag, J. V., Espinoza, V. M., . . . Hillman, R. E. (2015). Using ambulatory voice monitoring to investigate common voice disorders: Research update. *Frontiers in Bioengineering and Biotechnology*, 3, 155.
- Merrill, R. M., Roy, N., & Lowe, J. (2013). Voice-related symptoms and their effects on quality of life. *Annals of Otolaryngology, Rhinology & Laryngology*, 122(6), 404–411.
- Moon, J., & Alipour, F. (2013). Muscular anatomy of the human ventricular folds. *Annals of Otolaryngology, Rhinology & Laryngology*, 122(9), 561–567.
- Morrison, M. (1997). Pattern recognition in muscle misuse voice disorders: How I do it. *Journal of Voice*, 11(1), 108–114.
- Morrison, M. D., Rammage, L. A., Belisle, G. M., Pullan, C. B., & Nichol, H. (1983). Muscular tension dysphonia. *Journal of Otolaryngology*, 12(5), 302–306.
- Munhall, K. G., & Ostry, D. J. (1983). Ultrasonic measurement of laryngeal kinematics. In I. R. Titze & R. Scherer (Eds.), *Vocal fold physiology: Biomechanics, acoustics and phonatory control* (pp. 145–162). Denver, CO: Denver Center for the Performing Arts.
- Neely, G., Ljunggren, G., Sylven, C., & Borg, G. (1992). Comparison between the Visual Analogue Scale (VAS) and the Category Ratio Scale (CR-10) for the evaluation of leg exertion. *International Journal of Sports Medicine*, 13(2), 133–136.
- Noble, B. J., Borg, G. A., Jacobs, I., Ceci, R., & Kaiser, P. (1983). A category-ratio perceived exertion scale: Relationship to blood and muscle lactates and heart rate. *Medicine and Science in Sports and Exercise*, 15(6), 523–528.
- Ostry, D. J., & Munhall, K. G. (1985). Control of rate and duration of speech movements. *The Journal of the Acoustical Society of America*, 77(2), 640–648.
- Parker, L. A., Kunduk, M., Fink, D. S., & McWhorter, A. (2017). Reliability of high-speed videoendoscopic ratings of essential voice tremor and adductor spasmodic dysphonia. *Journal of Voice*, 33(1), 16–26.
- Pemberton, C., Russell, A., Priestley, J., Havas, T., Hooper, J., & Clark, P. (1993). Characteristics of normal larynges under flexible fiberoptic and stroboscopic examination—An Australian perspective. *Journal of Voice*, 7(4), 382–389.
- Pershall, K. E., & Boone, D. R. (1987). Supraglottic contribution to voice quality. *Journal of Voice*, 1(2), 186–190.
- Poburka, B. J., Patel, R. R., & Bless, D. M. (2017). Voice-Vibratory Assessment With Laryngeal Imaging (VALI) form: Reliability of rating stroboscopy and high-speed videoendoscopy. *Journal of Voice*, 31(4), e511–e513.
- Redenbaugh, M. A., & Reich, A. R. (1989). Surface EMG and related measures in normal and vocally hyperfunctional speakers. *Journal of Speech and Hearing Disorders*, 54(1), 68–73.

- Reidenbach, M. M.** (1996). The periepiglottic space: Topographic relations and histological organisation. *Journal of Anatomy*, 188(Pt. 1), 173–182.
- Reidenbach, M. M.** (1998). The muscular tissue of the vestibular folds of the larynx. *European Archives of Oto-Rhino-Laryngology*, 255(7), 365–367.
- Rosenthal, A. L., Lowell, S. Y., & Colton, R. H.** (2014). Aerodynamic and acoustic features of vocal effort. *Journal of Voice*, 28(2), 144–153.
- Roy, N.** (2008). Assessment and treatment of musculoskeletal tension in hyperfunctional voice disorders. *International Journal of Speech-Language Pathology*, 10(4), 195–209.
- Roy, N., & Bless, D. M.** (2000). Personality traits and psychological factors in voice pathology: A foundation for future research. *Journal of Speech, Language, and Hearing Research*, 43(3), 737–748.
- Roy, N., Bless, D. M., & Heisey, D.** (2000). Personality and voice disorders: A superfactor trait analysis. *Journal of Speech, Language, and Hearing Research*, 43(3), 749–768.
- Roy, N., Bless, D. M., Heisey, D., & Ford, C. N.** (1997). Manual circumlaryngeal therapy for functional dysphonia: An evaluation of short- and long-term treatment outcomes. *Journal of Voice*, 11(3), 321–331.
- Roy, N., Ford, C. N., & Bless, D. M.** (1996). Muscle tension dysphonia and spasmodic dysphonia: The role of manual laryngeal tension reduction in diagnosis and management. *Annals of Otolaryngology, Rhinology & Laryngology*, 105(11), 851–856.
- Roy, N., & Leeper, H. A.** (1993). Effects of the manual laryngeal musculoskeletal tension reduction technique as a treatment for functional voice disorders: Perceptual and acoustic measures. *Journal of Voice*, 7(3), 242–249.
- Roy, N., Merrill, R. M., Gray, S. D., & Smith, E. M.** (2005). Voice disorders in the general population: Prevalence, risk factors, and occupational impact. *Laryngoscope*, 115(11), 1988–1995.
- Rutkove, S. B.** (2007). Introduction to volume conduction. In A. S. Blum & S. B. Rutkove (Eds.), *The clinical neurophysiology primer* (pp. 43–53). Totowa, NJ: Humana Press.
- Sakakibara, K., Kimura, M., Imagawa, H., Niimi, S., & Tayama, N.** (2004). *Physiological study of the supraglottal structures*. Paper presented at the International Conference on Voice Physiology and Biomechanics, Marseille, France.
- Sama, A., Carding, P. N., Price, S., Kelly, P., & Wilson, J. A.** (2001). The clinical features of functional dysphonia. *Laryngoscope*, 111(3), 458–463.
- Sandage, M. J., Connor, N. P., & Pascoe, D. D.** (2013). Voice function differences following resting breathing versus submaximal exercise. *Journal of Voice*, 27(5), 572–578.
- Sapir, S., Baker, K. K., Larson, C. R., & Ramig, L. O.** (2000). Short-latency changes in voice F0 and neck surface EMG induced by mechanical perturbations of the larynx during sustained vowel phonation. *Journal of Speech, Language, and Hearing Research*, 43(1), 268–276.
- Sawashima, M., Kakita, Y., & Hiki, S.** (1973). Activity of the extrinsic laryngeal muscles in relation to Japanese word accent. *Annual Bulletin (Research Institute of Logopedics and Phoniatrics, University of Tokyo)*, 7, 19–25.
- Shipp, T.** (1975). Vertical laryngeal position during continuous and discrete vocal frequency change. *Journal of Speech and Hearing Research*, 18(4), 707–718.
- Sivasankar, M., & Fisher, K. V.** (2002). Oral breathing increases Pth and vocal effort by superficial drying of vocal fold mucosa. *Journal of Voice*, 16(2), 172–181.
- Smith, E., Gray, S. D., Dove, H., Kirchner, L., & Heras, H.** (1997). Frequency and effects of teachers' voice problems. *Journal of Voice*, 11(1), 81–87.
- Smith, E., Taylor, M., Mendoza, M., Barkmeier, J., Lemke, J., & Hoffman, H.** (1998). Spasmodic dysphonia and vocal fold paralysis: Outcomes of voice problems on work-related functioning. *Journal of Voice*, 12(2), 223–232.
- Smith, N. R., Rivera, L. A., Dietrich, M., Shyu, C. R., Page, M. P., & DeSouza, G. N.** (2016). Detection of simulated vocal dysfunctions using complex sEMG patterns. *IEEE Journal of Biomedical and Health Informatics*, 20(3), 787–801.
- Solomon, N. P., & DiMattia, M. S.** (2000). Effects of a vocally fatiguing task and systemic hydration on phonation threshold pressure. *Journal of Voice*, 14(3), 341–362.
- Solomon, N. P., Glaze, L. E., Arnold, R. R., & van Mersbergen, M.** (2003). Effects of a vocally fatiguing task and systemic hydration on men's voices. *Journal of Voice*, 17(1), 31–46.
- Stager, S. V., Bielamowicz, S. A., Regnell, J. R., Gupta, A., & Barkmeier, J. M.** (2000). Supraglottic activity: Evidence of vocal hyperfunction or laryngeal articulation? *Journal of Speech, Language, and Hearing Research*, 43(1), 229–238.
- Steinhauer, K., Grayhack, J. P., Smiley-Oyen, A. L., Shaiman, S., & McNeil, M. R.** (2004). The relationship among voice onset, voice quality, and fundamental frequency: A dynamical perspective. *Journal of Voice*, 18(4), 432–442.
- Steff, C. E.** (2012). Surface electromyography for speech and swallowing systems: Measurement, analysis, and interpretation. *Journal of Speech, Language, and Hearing Research*, 55(4), 1232–1246.
- Steff, C. E., Heaton, J. T., Jette, M. E., Burns, J. A., & Hillman, R. E.** (2010). Neck surface electromyography as a measure of vocal hyperfunction before and after injection laryngoplasty. *Annals of Otolaryngology, Rhinology & Laryngology*, 119(9), 594–601.
- Steff, C. E., Heaton, J. T., Stadelman-Cohen, T. K., Braden, M. N., Jette, M. E., & Hillman, R. E.** (2011). Characteristics of phonatory function in singers and nonsingers with vocal fold nodules. *Journal of Voice*, 25(6), 714–724.
- Steff, C. E., Hillman, R. E., & Heaton, J. T.** (2010). A virtual trajectory model predicts differences in vocal fold kinematics in individuals with vocal hyperfunction. *The Journal of the Acoustical Society of America*, 127(5), 3166–3176.
- Sundarrajan, A., Fujiki, R. B., Loerch, S. E., Venkatraman, A., & Sivasankar, M. P.** (2017). Vocal loading and environmental humidity effects in older adults. *Journal of Voice*, 31(6), 707–713.
- Sundberg, J., & Askenfelt, A.** (1981). *Larynx height and voice source: A relationship?* Paper presented at the Research Conference on Voice Physiology, Madison, WI.
- Sundberg, J., Titze, I., & Scherer, R.** (1993). Phonatory control in male singing—A study of the effects of subglottal pressure, fundamental-frequency, and mode of phonation on the voice source. *Journal of Voice*, 7(1), 15–29.
- Tanaka, S., & Gould, W. J.** (1983). Relationships between vocal intensity and noninvasively obtained aerodynamic parameters in normal subjects. *The Journal of the Acoustical Society of America*, 73(4), 1316–1321.
- Tanner, K., Fujiki, R. B., Dromey, C., Merrill, R. M., Robb, W., Kendall, K. A., . . . Sivasankar, M. P.** (2016). Laryngeal desiccation challenge and nebulized isotonic saline in healthy male singers and nonsingers: Effects on acoustic, aerodynamic, and self-perceived effort and dryness measures. *Journal of Voice*, 30(6), 670–676.
- Titze, I.** (1999). Toward occupational safety criteria for vocalization. *Logopedics Phoniatrics Vocology*, 24, 49–54.
- Van Houtte, E., Claeys, S., D'Haeseleer, E., Wuyts, F., & Van Lierde, K.** (2013). An examination of surface EMG for the assessment of muscle tension dysphonia. *Journal of Voice*, 27(2), 177–186.
- van Leer, E., & van Mersbergen, M. R.** (2017). Using the Borg CR10 physical exertion scale to measure patient-perceived

vocal effort pre and post treatment. *Journal of Voice*, 31(3), 389.e319–389.e325.

van Mersbergen, M., Patrick, C., & Glaze, L. (2008). Functional dysphonia during mental imagery: Testing the trait theory of voice disorders. *Journal of Speech, Language, and Hearing Research*, 51(6), 1405–1423.

Verdolini, K., Min, Y., Titze, I. R., Lemke, J., Brown, K., van Mersbergen, M., ... Fisher, K. (2002). Biological mechanisms underlying voice changes due to dehydration. *Journal of Speech, Language, and Hearing Research*, 45(2), 268–281.

Verdolini, K., Titze, I. R., & Fennell, A. (1994). Dependence of phonatory effort on hydration level. *Journal of Speech and Hearing Research*, 37(5), 1001–1007.

Verdolini-Marston, K., Burke, M. K., Lessac, A., Glaze, L. E., & Caldwell, E. (1995). Preliminary study of two methods of treatment for laryngeal nodules. *Journal of Voice*, 9(1), 74–85.

Witte, R., & Witte, J. (2010). *Statistics*. Hoboken, NJ: Wiley.

Yanagisawa, E., Estill, J., Kmucha, S., & Leder, S. (1989). The contribution of aryepiglottic constriction to “ringing” voice quality—A videolaryngoscopic study with acoustical analysis. *Journal of Voice*, 3(4), 342–350.

Zanartu, M., Galindo, G. E., Erath, B. D., Peterson, S. D., Wodicka, G. R., & Hillman, R. E. (2014). Modeling the effects of a posterior glottal opening on vocal fold dynamics with implications for vocal hyperfunction. *The Journal of the Acoustical Society of America*, 136(6), 3262–3271.

Zheng, Y. Q., Zhang, B. R., Su, W. Y., Gong, J., Yuan, M. Q., Ding, Y. L., & Rao, S. Q. (2012). Laryngeal aerodynamic analysis in assisting with the diagnosis of muscle tension dysphonia. *Journal of Voice*, 26(2), 177–181.

Zietels, S., Burns, J. A., Lopez-Guerra, G., Anderson, R., & Hillman, R. E. (2008). Photoangiolytic laser treatment of early glottic cancer: A new management strategy. *Annals of Otology, Rhinology & Laryngology*, 117(7), 1–24.

Appendix A (p. 1 of 5)

Data Acquisition and Processing

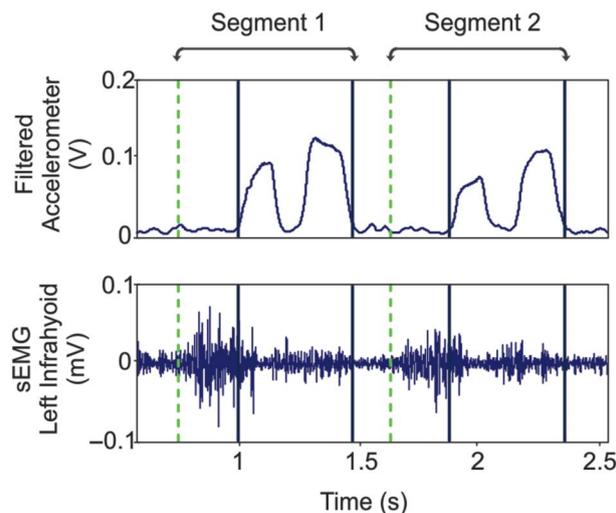
This appendix includes additional data acquisition and processing information for data acquired with sEMG sensors, neck surface accelerometry, the PAS, and high-speed videoendoscopy.

Neck sEMG

Following placement of the sEMG sensors, participants then completed a series of tasks to determine an MVC value for each sensor and to verify electrode placement over muscle groups. Participants completed three repetitions of each of the following tasks: saliva swallow, throat clear, neck flexion, and isometric contraction against resistance. The isometric contraction involved placing a dynamometer below the chin and countering the force of downward contractions. On average, participants produced a force of 14.2 lbf during this task. The maximal MVC was determined for each task for each of the three sensors via a sliding RMS window of 125 ms with 50% overlap (Stepp, 2012). The largest MVC value was determined for each sensor and used during data normalization of the neck sEMG signals.

The voicing onset and offset for each /ifi/ was determined from the accelerometer signal. Then, a 250-ms prephonatory time period was added at the voice initiation time point in order to account for muscle activity in the sEMG signal prior to the manifestation of voicing in the acoustic signal. If there was less than 250 ms between /ifi/ repetitions, the selected /ifi/ segment was only analyzed to the voicing offset of the previous /ifi/; this most often occurred during the fast rate productions. Figure A1

Figure A1. The upper panel shows a filtered accelerometer signal that was used to determine the onset and offset of voicing for each /ifi/ production, delineated by a solid dark line. The dashed line (---) represents a time period set to 250 ms prior to each phonation onset. Segment 1 and Segment 2 are the two /ifi/ segments that include the voicing segment plus the prephonatory segment. The lower panel is an example of the sEMG signal acquired from the sensor located at the left infrahyoid location. In this example, the analysis of Segment 1 and Segment 2 revealed a mean normalized activation amplitude of 3% and mean duration of activation of 100%.



provides two examples of /ifi/ segmentations during a fast rate recording in which there were large enough time blocks to segment a full 250-ms prephonatory segment. From here, two measures were determined for each recording: (a) percent activation and (b) percent duration.

Postrecording Task With the PAS

The post experimental recording task was completed using the PAS to verify a relationship between intraoral estimates of subglottal pressure and measures taken with the accelerometer. Participants were trained to produce a series of /pi/ syllables at a slow, steady rate of approximately 1.5 syllables per second (Hertegard, Gauffin, & Lindestad, 1995; Holmberg, Perkell, & Hillman, 1984). A single /pi/ string began with an /i/ vowel, followed by five /pi/ productions (i.e., /i pi pi pi pi pi/). Participants produced /pi/ strings at a comfortable pitch and speaking volume and then produced /pi/ strings with the addition of increasing levels of vocal effort in which they incrementally increased vocal effort at each /pi/ string. Participants utilized the visual feedback provided by the PAS display to view their increases in intraoral pressure and to maintain the same amount of intraoral pressure within a /pi/ string. Because of the instructions to monitor vocal effort via intraoral pressure feedback, the task needed to be completed after the experimental recordings, so as not to confound the strategies speakers used to increase vocal effort during the study.

The accelerometer signals gathered during the postrecording task were processed with the same semiautomated algorithm to extract the RMS of the vowel segments during experimental data processing of /ifi/ segments. As such, the algorithm determined the RMS of the vowel segments in each /pi/ string. Then, the maximum intraoral pressure was also determined for each /p/ production and the relationship between the two variables was assessed via Pearson product-moment correlation coefficients. The correlations were moderate to strong ($M = .86$, range .58–.97) for all 26 speakers, which met a prespecified cutoff criterion of $r > .50$ (McKenna et al., 2017) and verified a relationship between the accelerometer measure and intraoral estimate of subglottal pressure.

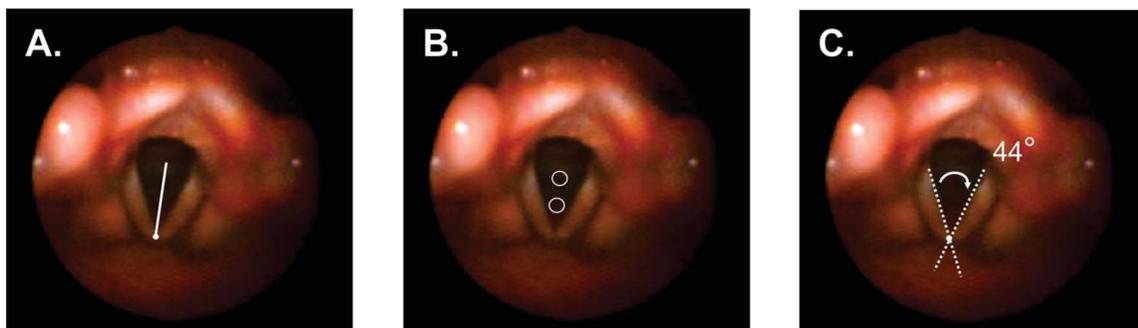
Training for Supraglottal Compression Ratings

Training was completed prior to experimental data processing. The first author and the certified SLP independently completed VALI ratings of M-L and A-P compression on 108 randomly extracted images, for a total of 216 ratings. The laryngeal images were from the experimental data in order to provide relevant examples of compression ratings. Following the independent ratings, any discrepancies greater than 1 point on the rating scale were discussed to consensus ($n = 18$ for M-L compression ratings and $n = 1$ for A-P compression ratings). Discrepancies were most commonly due to variation in endoscope viewing angle or image quality.

High-Speed Videoendoscopy: Kinematic Stiffness Ratios

This section describes the process with which kinematic estimations of stiffness were determined using a semiautomated algorithm. First, technicians underwent glottic angle identification training. This initial training was completed on flexible laryngoscopic images at a standard sampling rate (30 fps) and halogen light source that provided bright, unobstructed images of the vocal folds during /ifi/ utterances. The angle markings of the technicians were directly evaluated against angle markings made by the first author, meeting a two-way ICC analysis for consistency $\geq .80$. The technicians then completed training with a custom interactive algorithm. The interactive algorithm required the technician to center the glottis, identify the anterior commissure, identify pixels for shading differences, and then make judgments on the appropriateness of vocal fold edge detection, glottic angle tracking, and velocity curves. Once again, the technicians had to meet reliability standards of $ICC(2,1) \geq .80$. After these two steps were completed, the technicians could proceed to processing experimental data.

Figure A2. (a) Laryngoscopy image with glottic midpoint identified. (b) Laryngoscopy image with glottic space identified (circles) for algorithm pixel differentiation. (c) Regression lines (---) placed along the vocal fold edges to determine the glottic angle.



First, the technician identified the midline of the glottis, from the anterior commissure to the bilateral vocal processes (Figure A2a). The points along the line assisted in creating a restriction window for the possible locations of the anterior commissure during glottic angle tracking. Next, the technician identified the space within the glottis (Figure A2b) to provide representative pixel shades for differentiating the glottis from the vocal fold edge. From this information (i.e., the location of the anterior commissure and the edge points determined from the shading differences), two regression lines were fit to the vocal fold edge via a least squares regression model. Perpendicular error was minimized in order to yield more accurate vocal fold edge tracking. The intersection of the two lines created an angle that could be measured and used as the raw glottic angle for tracking glottic angle over time.

The raw glottic angles were plotted over time from the vibration during the initial /i/ vowel, through the abduction and adduction behavior of the /f/, and finally through the oscillations of the following /i/. In order to filter out vocal fold vibration prior to and after the /f/ segment, upper and lower envelopes were generated from the vocal fold data. A custom function tracked the local maximum and minimum angles related to the opening and closing behavior of vocal fold vibration. The raw vocal fold angles were kept for further analysis at the point in time at which the low and high envelopes converged, which indicated onset and offset of the /f/ segment. Outside the /f/ segment, raw values of the lower envelope (the minimum angle during vocal fold vibration) were used for further analysis and are referred to as “pruned angles.” Figure A3 provides an example of raw angles as well as low and high envelopes with arrows pointing to the convergence zones.

The pruned vocal fold angles (during the vibration) and the raw vocal fold angles during the /f/ phoneme were then zero-phase filtered using a low-pass FIR filter of order $n = 15$, with a cutoff frequency of 25 Hz. Adduction was then determined as the time from the maximum angle (from the low-pass filtered data) to the point where vocal fold angle dropped to less than 20% of the maximum vocal fold angle. To account for the quick drop in vocal fold angles at the onset of vocal fold vibration, an empirically derived filter window was applied to the data. This window prevents large velocities associated with vocal fold vibrations at the onset of voicing from being extracted in the final determination of gross vocal fold adductory velocity. Thus, the maximum angle velocity was identified as the minimum derivative value (i.e., the most negative slope) of the low-passed vocal fold angles within the adduction window. Figure A4 provides the smoothed data and the range of points from which the maximum angular velocity was determined.

In approximately 28% of samples, the algorithm did not track vocal fold edges due to the images being too dark or because supraglottic structures were covering the view of the true vocal folds (e.g., epiglottis, false vocal folds). In these cases, the user was able to manually mark glottic angles at a reduced sampling rate of 50 frames per second; the algorithm then incorporated the new anterior commissure information to create new restrictions for solving the vocal fold edge detection. From here, the angles are plotted, pruned, and analyzed with the same methodology as with the algorithm-generated angles. Following the manual markings, the technicians accepted 75% of these /ifi/ productions (of the 28% that required manual marking). Finally, the technicians discarded any /ifi/ productions that could not be determined by the algorithm or by manual-assisted angle estimations, which accounted for only 7% of all individual /ifi/ productions in this study.

This same process was repeated for every /ifi/ production, resulting in kinematic stiffness ratios that could be averaged across a single voice recording. Users were able to see the results of the automated algorithm and intervene if they suspected the algorithm did not track vocal fold edges accurately. The visual screen they used to determine this incorporated the raw video information, the microphone and accelerometer signals, the raw glottic angles over time, and the angular velocity estimations (see Figure A5).

Figure A3. Raw vocal fold angles with low and high envelopes. Arrows indicate point of convergence of the envelopes. The raw angles were used within the space between the arrows and the low angles were used outside the arrows for further processing.

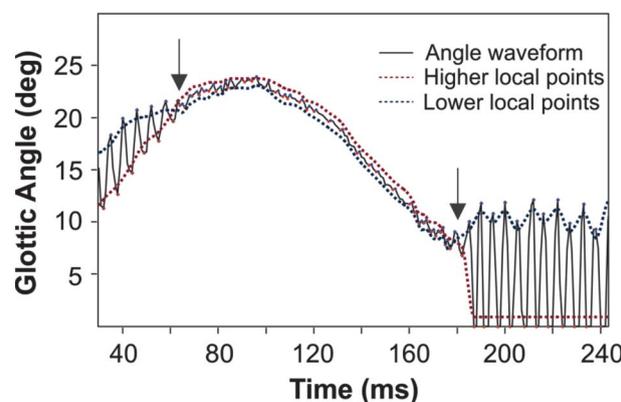


Figure A4. Image of the smoothed angle data. The maximum abductory angle and the maximum angular velocity have been determined in the range of 20%–80%, with consideration of the filter window.

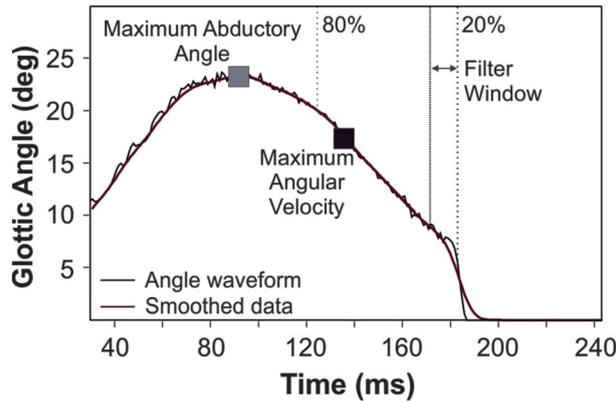
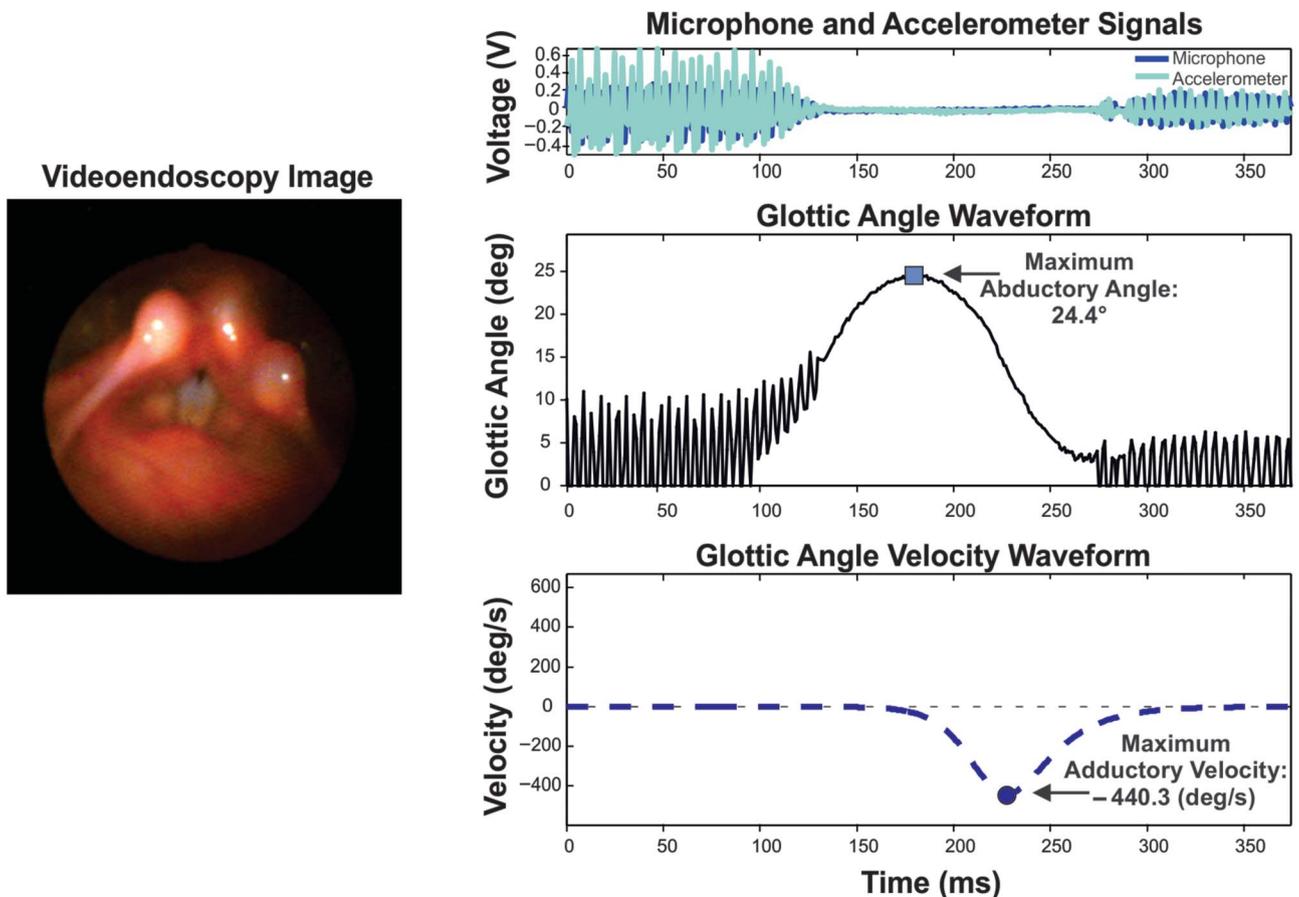


Figure A5. Schematic of the interactive screen during data processing. The user was able to see the videoendoscopic image, microphone and accelerometer signals, raw angle waveform (here, the angles have been smoothed during the /f/ segment), and the angular velocity waveform derived from the processed angle data.



Data Acquisition and Processing

The initial data processing was rechecked by a second trained technician, with a total processing time of approximately 6–8 hr per participant. To determine the validity of the algorithm, we directly compared the smoothed angle data from the algorithm to two additional trained technicians who were blind to the data set. Two individual /ifi/ productions per participant were randomly extracted (one from a speed condition and one from an effort condition). These extractions were only in the 72% of the data that were automatically calculated via the algorithm. The glottic angles were manually identified by the additional technicians at a down-sampled rate of 100 fps. Table A1 provides a summary of ICC results for the technicians. A final two-way ICC analysis for consistency was calculated between the smoothed angle data and an average of the additional technicians' manual makings resulting in an ICC(2, 1) = .85 (95% CI [.81, .89]).

Table A1. Two-way intraclass correlation coefficients (ICCs) and 95% confidence intervals between the trained technicians and the semiautomated glottic angle algorithm.

Comparison	Reliability
Technician 1 vs. Algorithm	.82 [.77, .86]
Technician 2 vs Algorithm	.84 [.80, .88]
Technician 1 vs. Technician 2	.89 [.86, .91]
Averaged technicians vs. Algorithm	.85 [.81, .89]

Appendix B

Individual Correlations (*r*) Between the Physiological Measures and Self-Perceptual Ratings of Vocal Effort

Participant	Correlations (<i>r</i>) with self-ratings of vocal effort							
	NSV _{Mag}	Kinematic Stiffness Ratio	Percent activation of suprahyoids	Percent duration of suprahyoids	Average percent activation of infrahyoids	Average percent duration of infrahyoids	M-L compression	A-P compression
P1	.83	.01	.78	.73	.81	.55	.51	.26
P2	.92	.72	.88	.53	.75	.63	.04	-.48
P3	.67	-.17	.87	.67	.83	.75	-.19	-.35
P4	.97	-.13	-.46	-.71	-.01	-.11	.85	.80
P5	.94	.01	.80	.67	.77	.03	-.30	-.60
P6	.02	.26	.95	.74	.82	.72	.47	.07
P7	.71	.33	.60	.54	.67	.47	.37	.26
P8	.63	.22	.71	.38	.21	.66	.72	-.84
P9	.67	.57	.07	.26	-.53	-.41	-.12	.19
P10	-.70	.35	.28	-.03	-.53	.02	.74	.69
P11	.86	-.16	.05	.10	.58	.76	.77	.64
P12	.96	.69	-.45	.62	.33	-.70	-.24	-.60
P13	.49	.40	.24	.42	.38	.63	.45	.18
P14	.86	-.17	-.78	-.86	-.43	-.81	.18	.32
P15	.94	-.34	.94	.80	-.47	.80	.10	.57
P16	.93	-.18	.44	.05	.87	.42	.74	.29
P17	.93	-.32	.59	.39	-.55	-.63	-.48	.38
P18	.94	.34	.86	.86	.65	.42	.88	.96
P19	.96	.01	.90	.76	-.57	-.40	.30	-.50
P20	.96	.02	.88	.70	.61	.83	.10	.11
P21	-.66	-.72	.71	.35	.48	.77	.87	.79
P22	.62	.22	.42	.18	.66	.66	.75	.71
P23	.96	.57	-.25	-.29	-.77	-.65	.92	.82
P24	.97	-.25	.94	.86	.54	.39	.82	.79
P25	.95	-.12	.52	.67	.05	.63	.85	.77
P26	-.02	.62	.48	.06	.39	.04	.61	.34
Number of speakers with <i>r</i> ≥ .70	17 (65%)	1 (4%)	12 (46%)	7 (27%)	6 (23%)	6 (23%)	11 (42%)	7 (27%)

Note. Correlations that met the criterion of $r \geq .70$ are bolded. NSV_{Mag} = magnitude of neck surface vibration; M-L = mediolateral; A-P = anterior-posterior.