The relationship between acoustical and perceptual measures of vocal effort

Victoria S. McKenna[a] and Cara E. Stepp[b]
Department of Speech, Language, and Hearing Sciences, Boston University, 677 Beacon Street, Boston, Massachusetts 02215, USA

(Received 17 May 2018; revised 15 August 2018; accepted 6 September 2018; published online 27 September 2018)

Excessive vocal effort is a common clinical voice symptom, yet the acoustical manifestation of vocal effort and how that is perceived by speakers and listeners has not been fully elucidated. Here, 26 vocally healthy adults increased vocal effort during the production of the utterance /ifi/, followed by self-ratings of effort on a 100 mm visual analog scale. Twenty inexperienced listeners assessed the speakers' vocal effort using the visual sort-and-rate method. Previously proposed acoustical correlates of vocal effort were calculated, including: mean sound pressure level (SPL), mean fundamental frequency (f_o), relative fundamental frequency (RFF) offset cycle 10 and onset cycle 1, harmonics-to-noise ratio (HNR), cepstral peak prominence and its standard deviation (SD), and low-to-high (L/H) spectral ratio and its SD. Two separate mixed-effects regression models yielded mean SPL, L/H ratio, and HNR as significant predictors of both speaker and listener ratings of vocal effort. RFF offset cycle 10 and mean f_o were significant predictors of listener ratings only. Therefore, speakers and listeners attended to similar acoustical cues when making judgments of vocal effort, but listeners also used additional time-based information. Further work is needed to determine how vocal effort manifests in the speech signal in speakers with voice disorders.
 https://doi.org/10.1121/1.5055234

[AKCL] Pages: 1643–1658

I. INTRODUCTION

Excessive vocal effort is a common clinical symptom of speakers with voice disorders (Altman *et al.*, 2005; Bach *et al.*, 2005; Cannito *et al.*, 2012; Roy *et al.*, 2005; Smith *et al.*, 1998). It has also been reported in individuals with high occupational voice demands, such as teachers and singers (de Alvear *et al.*, 2011; Smith *et al.*, 1997), and approximately 10% of vocally healthy older adults (Merrill *et al.*, 2013).

The study of vocal effort is multidisciplinary, with contributions from exercise physiology, speech-language pathology, psychology, occupational health, and otolaryngology (to name a few). Vocal effort has been described as an "exertion of the voice" (Baldner *et al.*, 2015) and "perceived effort in producing speech" (Eadie *et al.*, 2010; Eadie *et al.*, 2007; Isetti *et al.*, 2014; Verdolini *et al.*, 1994). Other definitions have stated that the vocal exertion can be "quantified objectively by the A-weighted speech level at 1 m distance in front of the mouth and qualified subjectively by a description" (ISO, 2002). This definition provides an objective indicator of effort (solely that of the amplitude of the speech signal) and has been used as a basis for research focused on how the environment impacts the perception of vocal effort (i.e., background noise, room acoustics; Bottalico *et al.*, 2016). Although the definition provides a promising metric of vocal effort, it is likely that excessive vocal effort is not related to the amplitude of the signal alone. To date, multiple acoustical measures have been associated with increasing vocal effort in vocally healthy speakers and speakers with voice disorders. These acoustical changes include time-, spectral-, and cepstral-based measures; yet, a comprehensive analysis of all of these measures is lacking from the literature.

The present work was based on the working hypothesis proposed by McCabe and Titze (2002), which assert that the sensation of vocal effort stems from a "miscalibration" between the effort needed to initiate and maintain voicing to the quality or intensity of the resultant speech signal. Quantifying that mismatch in the clinical setting has proved challenging, with perceptual ratings between speakers and expert clinical judgements not always aligning. It is hypothesized that speakers and listeners may be attending to separate acoustical cues when making these judgements, but this has not been tested on a large set of acoustical measures. As such, the purpose of this study was to evaluate the relationship between previously hypothesized acoustical predictors of vocal effort and perceptual judgments of vocal effort.

A. Perceptual measures of vocal effort

Auditory-perceptual ratings are considered the gold-standard for evaluating voice disorders and assessing treatment progress (Oates, 2009; Selby *et al.*, 2003). Perceptual ratings include self-reports by speakers, as well as listener ratings completed by clinical staff (e.g., speech-language pathologist; SLP) and familiar listeners (e.g., family members, caregivers). These perceptual ratings provide insight into the voice

[a]Electronic mail: vmckenna@bu.edu
[b]Also at: Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA.

impairment and can be used to help define therapeutic goals in voice therapy.

Two types of speaker self-perceptual ratings are employed clinically. The first type of rating provides estimates of the frequency, severity, and duration of vocal symptoms, as well as the impact voice problems have on the quality of life of the speaker (Hogikyan and Sethuraman, 1999). Many psychosocial questionnaires employ this first type of rating to explore the incidence of vocal effort and the extent to which vocal effort affects daily life (e.g., Vocal Handicap Index, Glottal Function Index; Bach *et al.*, 2005; Jacobson *et al.*, 1997). The second type of self-perceptual rating is reported immediately following a specific voice task to provide an instantaneous rating of current voice symptoms. An example of this is the Inability to Produce a Soft Voice (IPSV), which immediately evaluates how difficult it is to vary pitch and loudness. High levels of difficulty with IPSV tasks are associated with physical changes in the vocal folds, such as vocal fold swelling (Bastian *et al.*, 1990), whereas improvements in ISPV ratings are predictive of vocal recovery following vocal fatigue (Hunter and Titze, 2009). Scales such as the Borg Category Ratio 10 (Borg CR10; Borg, 1982; Neely *et al.*, 1992) and the 100 mm visual analog scale (VAS) are used to assess the instantaneous sensation of vocal effort severity at the time of the rating (e.g., Sundarrajan *et al.*, 2017). Both are versatile scales that can be used to capture self- and listener-perceptual ratings of vocal effort, allowing for a direct comparison between ratings made by both groups (Eadie *et al.*, 2010; Eadie and Stepp, 2013; Isetti *et al.*, 2014; Stepp *et al.*, 2012).

Many clinical tools have been developed to quantify listeners' perceptions of voice as well. Listener-perceptual ratings provide information on the impact that the voice has on the communication partner and provide another perspective on the speaker's vocal impairment (Eadie *et al.*, 2013; Isetti *et al.*, 2014). Listener-perceptual ratings are especially important for speakers who may not have an accurate perception of their own voices or when speakers have become accustomed to their own voices. For example, speakers with Parkinson's disease can exhibit reduced vocal loudness and reduced pitch variation, but often report no problems in their speech and voice (Kwan and Whitehill, 2011).

Common clinical tools used to assess listener-perceptions include the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V; Kempster *et al.*, 2009) and the Grade, Roughness, Breathiness, Asthenia, Strain (GRBAS; Hirano, 1981) scale, which both assess the perception of pitch, loudness, and quality of the voice. On each scale, the perceptual category of "strain" is defined as "the perception of excessive vocal effort." The terms *effort* and *strain* are often used interchangeably when describing listener-perceptual ratings (e.g., Lien *et al.*, 2015; Rosenthal *et al.*, 2014) as it is believed that a strained voice is produced with the addition of vocal effort. However, strain encompasses the additional perceptual qualities of breathiness and roughness (in upwards of 50% of speakers; Lowell *et al.*, 2012), and it seems that speakers can report elevated levels of vocal effort without dysphonic changes to their voices that are associated with strain. For example, studies focused on increased vocal effort from vocal fatigue have reported consistent changes to pitch and intensity of the voice, but inconsistent changes to other acoustical measures (Laukkanen *et al.*, 2008; Vilkman *et al.*, 1999; Xue *et al.*, 2018). The overlap between vocal effort and vocal strain could contribute to some of the discrepancies between self- and listener-perceptual ratings reported in the literature.

To date, the relationships between self- and listener-perceptual ratings of vocal effort have shown that they are, at best, only moderately related to one another (Eadie *et al.*, 2010; Eadie *et al.*, 2007; Johnson, 2012). In a study by Lee *et al.* (2005), the relationships between speaker and expert listener ratings of vocal strain (defined in this study as the perception of excessive vocal effort) were considered weak with a correlation of $r = 0.19$. Results showed that the speakers with dysphonia rated their amount of vocal strain consistently greater than the expert clinical judgments.

Various hypotheses have been proposed to try to explain the discrepancies between self- and listener-perceptual ratings of voice. One hypothesis is that speakers use both somatosensory and auditory feedback when making ratings, whereas listeners only have auditory information of which to make judgments. This hypothesis is supported by the results of a study that examined self- and listener-perceptual ratings of vocal effort when speakers had unaffected auditory feedback, and then altered auditory feedback via masking noise (Lane *et al.*, 1961). Results indicated that in both experiments, speakers rated their own vocal effort consistently greater than listener ratings of the same stimuli. The authors hypothesized that speakers may have used somatosensory feedback and bone conduction (both speaker-specific sensory information) when making their own self-ratings of vocal effort. Furthermore, studies have investigated how altered laryngeal sensory feedback may impact vocal control. Investigations have shown that speakers had a reduced ability to control pitch when a numbing agent was applied to the vocal fold mucosa (Kleber *et al.*, 2013; Sundberg *et al.*, 1995). These results provide evidence that speakers use somatosensory feedback from sensory receptors (e.g., mechanoreceptors) in the larynx during fine-tuned adjustments of voice, even when auditory feedback is unaffected. It has since been hypothesized that speakers have more reliable estimates of their own vocal effort compared to listeners since they have access to both sensory modalities during their self-perceptual ratings (Lee *et al.*, 2005).

In order to assess the differences between speaker- and listener-perceptual ratings, researchers have turned to acoustical analysis of the speech signal. With the advent of clinically accessible software and algorithms, acoustical measures are now a standard of clinical care (Patel *et al.*, 2018). Understanding the acoustical manifestation of vocal effort would provide more information to clinicians and assist in their ability to identify and then remediate vocal effort in the clinical setting.

B. Acoustical measures of vocal effort

Acoustical measures may provide a quantitative way to examine the discrepancy between self- and listener-perceptual ratings of vocal effort. At present, multiple

acoustical measures have been associated with the perception of vocal effort, including amplitude-, time-, spectral-, and cepstral-based measures. However, there seems to be no single acoustical measure predictive of perceptual ratings of vocal effort (speaker or listener), indicating that the perception of effort is likely related to multiple acoustical changes in the speech signal.

1. Amplitude-based

The amplitude of the speech signal can be quantified in sound pressure level (dB SPL) and is perceived as loudness. A study by Rosenthal *et al.* (2014) examined a series of acoustical and aerodynamic measures during modulations of vocal effort in healthy speakers. The authors reported a positive association between mean SPL and vocal effort, with an increase of 3 dB SPL from comfortable speaking effort to a maximal vocal effort. It has been suggested that increased subglottal pressure, the pressure that is crucial to initiating and maintaining vocal fold oscillation, is a physiological manifestation of vocal effort (Verdolini *et al.*, 1994), which simultaneously acts to increase mean SPL. Therefore, the perception of vocal effort may be related to an increase in the amplitude of the speech signal; however, some speakers with voice disorders exhibit elevated subglottal pressure without the same degree of change in mean SPL (Espinoza *et al.*, 2017; Friedman *et al.*, 2013). Importantly, listeners are still able to perceive vocal effort in patient populations that have these symptoms (e.g., vocal hyperfunction; Stepp *et al.*, 2012). Thus, it is unlikely that speakers and listeners depend on the amplitude of the speech signal alone; rather, it is more likely that the perception of vocal effort is a combination of changes in mean SPL and other acoustical parameters in the speech signal.

2. Time-based

Time-based measures include mean fundamental frequency (f_o), harmonics-to-noise-ratio (HNR), and relative fundamental frequency (RFF). These acoustical measures have been shown to be correlated with auditory-perceptual judgments of vocal effort in speakers with voice disorders (Eadie and Stepp, 2013; Stepp *et al.*, 2012) and to change when vocally healthy speakers purposefully increase vocal effort and strain (Lien *et al.*, 2015; McKenna *et al.*, 2016).

Increases in mean f_o are attributed to increased tension of the intrinsic laryngeal muscles, such as the cricothyroid (Lofqvist *et al.*, 1989; Shipp, 1975) as well as increased subglottal pressure (Titze, 1989). Increased mean f_o has been reported during instances of vocal fatigue (Rantala *et al.*, 1998; Vilkman *et al.*, 1999), which is hypothesized to be due to compensatory increases in laryngeal tension and vocal effort. For example, Ghassemi *et al.* (2014) monitored mean f_o in healthy speakers and in speakers with vocal hyperfunction. Results showed that only speakers with vocal hyperfunction exhibited an increase in mean f_o over the duration of the day, which the authors related to increased vocal effort and increased laryngeal tension from vocal fatigue. As such, vocal effort may manifest as increases in mean f_o.

HNR is an acoustical measure that characterizes the periodicity of the speech signal (Murphy *et al.*, 2008). Although HNR can be calculated in the frequency domain (Qi and Hillman, 1997), we are referring to HNR as a time-based acoustical measure in the present study due to our calculation of the measure in the time domain. HNR is a ratio of periodic energy to aperiodic noise in the signal and is affected by aperiodic vocal fold vibration. As such, HNR can be reduced in speakers with dysphonia due to aperiodic vocal fold vibration. For example, HNR is reduced in speakers with vocal fold lesions, such as those with glottic cancer (Friedman *et al.*, 2013) and vocal fold nodules (Schindler *et al.*, 2009). Speakers with vocal fold nodules frequently report increased vocal effort (Hillman *et al.*, 1989; Holmberg *et al.*, 2003), which may be related to vocal fold vibratory function. To date, there seems to have been no specific study that directly examines the relationship between HNR and perceptual ratings of vocal effort. The evidence that HNR is affected in speakers with voice disorders who have primary symptoms of vocal effort makes it a promising acoustical indicator of vocal effort.

Unlike mean f_o and HNR, which are determined from steady-state voicing segments, RFF is calculated during voicing transitions surrounding a voiceless consonant (e.g., /ifi/). The offset of voicing and re-onset of voicing in these vowel segments are referred to as RFF offset cycles and RFF onset cycles, respectively. The two cycles closest to the voiceless consonant, RFF offset cycle 10 and RFF onset cycle 1, are reported to be reduced when healthy speakers purposefully increase vocal effort and strain (Lien *et al.*, 2015; McKenna *et al.*, 2016) and when speakers have voice disorders, such as spasmodic dysphonia and vocal hyperfunction (Eadie and Stepp, 2013; Heller Murray *et al.*, 2017). Furthermore, these cycles have shown moderate associations with listener-perceptual ratings of vocal effort in some speakers (Eadie and Stepp, 2013; Lien *et al.*, 2015). The hypothesized mechanisms underlying changes to RFF values include aerodynamic forces, vocal fold abduction, and intrinsic laryngeal tension (Heller Murray *et al.*, 2017; McKenna *et al.*, 2016; Stepp *et al.*, 2011). Accordingly, RFF values may be related to tension in the vocal mechanism and perceived as increased vocal effort.

3. Spectral- and cepstral-based

Unlike time-based measures, spectral- and cepstral-based measures do not require calculation of specific time-based information (e.g., periods, periodicity) of the voicing segments in the speech signal. As such, spectral- and cepstral-based measures may be more appropriate for speakers with moderate-severe dysphonia when aperiodic voice signals preclude estimation of time-based measures.

The spectrum is determined via calculation of the fast Fourier transform (FFT) of the time-based signal, and provides information on the strength of energy across different frequencies. The proportion of low (below 4000 Hz) to high (above 4000 Hz) frequency information is referred to as the low-to-high (L/H) ratio and has been shown to be reduced in speakers with dysphonia (Awan *et al.*, 2010; Lowell *et al.*,

2013). Spectral energy above 4000 Hz can be due to increased high frequency aspiration noise, suspected to be due to larger posterior glottal gap sizes (Klatt and Klatt, 1990; Zanartu *et al.*, 2014) and perceived as excessive breathiness (Hillenbrand and Houde, 1996). Although vocal effort has also been associated with the percept of breathiness in speakers with phonotraumatic vocal hyperfunction (e.g., vocal nodules; Holmberg *et al.*, 2003), it is unclear whether breathiness causes a compensatory effortful vocal response, or conversely, whether excessive vocal effort has a breathy perceptual quality to it.

A cepstrum is calculated as the FFT of the logarithm of the power spectrum (Bogert *et al.*, 1963; Noll, 1964, 1967). Cepstral peak prominence (CPP) is the magnitude of the dominant rahmonic, thought to be equivalent to f_o, when compared to the predicted cepstral energy (Awan and Roy, 2005). A study by Rosenthal *et al.* (2014) evaluated CPP in healthy speakers across three levels of vocal effort: comfortable, minimal, and maximal. A significant difference was found between the comfortable and the maximal effort conditions, with greater CPP values reported during increased vocal effort. The authors suggested that CPP may have increased due to simultaneous increases in mean SPL. This is consistent with prior research that showed a positive association between CPP and SPL (Awan *et al.*, 2012). Although it seems promising that a cepstral-based measure may be sensitive to changes in vocal effort, further information is needed to determine how SPL may change or influence that prediction. A comprehensive analysis of multiple acoustical measures at the same time would provide information on which acoustical changes are contributing to the perception of vocal effort.

C. Aim and hypothesis

The purpose of this study was to further understand how vocal effort manifests in the speech signal and how it is interpreted perceptually. Therefore, we evaluated the relationship between perceptual ratings of vocal effort and a set of acoustical measures previously correlated with vocal effort. The acoustical measures examined in this study included amplitude-, time-, spectral-, and cepstral-based measures. We hypothesized that the acoustical measures that significantly predicted self-perceptual ratings of vocal effort would be different than those that significantly predicted listener-perceptual ratings.

II. METHOD

A. Speaker recordings

We enrolled 26 healthy young adults [18–29 years, Mean (M) = 20.9 years, standard deviation (SD) = 2.8 years] who were speakers of Standard American English. Participants had no history of speech, language, hearing, neurological, pulmonary, or voice disorders, and were nonsmokers. We enrolled 10 men and 16 women (~60% women), which is consistent with the sex distribution of speakers with voice disorders (Brinca *et al.*, 2015). Participants were screened for normal vocal function by a certified SLP via auditory-perceptual screening and flexible laryngoscopy. Speakers provided informed consent with approval of the Boston University Institutional Review Board prior to beginning the study.

Participants were trained to produce iterations of the utterance /ifi/. Each /ifi/ set consisted of eight consecutive /ifi/ productions, with a pause in the middle (e.g., /ifi ifi ifi ifi/, pause, /ifi ifi ifi ifi/). The combination of the phonemes in the utterance /ifi/ provided a stimulus that met all criteria for the acoustical processing planned in this study.

Four different voice conditions were elicited across speakers: typical speaking voice, mild effort, moderate effort, and maximal effort. Effort was elicited via the following instructions: "Increase your effort during your speech by 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." These instructions were specifically chosen to elicit effort from the laryngeal structures instead of a free interpretation of effort which could include other physiological (e.g., respiratory, articulatory) or cognitive contributions. Furthermore, the goal of these instructions was to increase vocal effort in a way that speakers may increase effort during conversational speaking conditions. 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." Each condition was recorded two times and had a range from six to ten /ifi/ productions, with the target of eight productions.

Following each recording, speakers completed ratings of their self-perceived vocal effort on a 100 mm VAS. The VAS has the benefits of being a continuous scale that allows for explicit anchors (Gerratt *et al.*, 1993). Zero was anchored as "No Effort" and 100 was anchored as "The Most Effort."

Speaker recordings were made with a directional headset microphone (Shure SM35 XLR) placed 45° from midline of the vermilion of the lips and 7 cm from the corner of the mouth. A neck-surface accelerometer (BU series 21771; Knowles Electronic, Itasca, IL) was placed with double sided adhesive at midline of the anterior neck, superior to the sternal notch and inferior to the cricoid cartilage. In order to determine mean SPL during processing of the speech signal, a calibration procedure was performed. The calibration included three electrolaryngeal pulses at the midline of the lips and readings of known dB SPLs from a sound pressure level meter (CM-150, Galaxy Audio; A-weighted) held at the microphone (7 cm away from, and directed toward, the mouth). The known dB SPLs of the electrolaryngeal pulses were later used to calibrate speech recordings to mean SPL (see Acoustical Data Processing for further information). The microphone and accelerometer signals were preamplified (Xenyx Behringer 802 Preamplifier) and then digitized at 30 kHz with a data acquisition board (National Instruments 6312 USB). The signals were acquired via a MATLAB algorithm and converted to wave files for further processing.

The voice recordings in this study were made with concurrent high-speed flexible laryngoscopy recordings, which are discussed in a separate study. No laryngeal numbing agent was provided so as not to affect laryngeal feedback or sensitivity (Dworkin *et al.*, 2000). Due to the recording limitations of the high-speed flexible nasendoscopic equipment, each speech recording was only eight seconds in duration. Inadvertently, some of the final /ifi/ productions in a recording were cut-off in the middle of the production. These incomplete /ifi/ productions were discarded during acoustical and perceptual processing.

B. Perceptual stimuli preparation

Stimuli sets were created for the visual sort-and-rate (VSR) method (Granqvist, 2003). The VSR method provides multiple voice samples in a single listening set for direct comparison against one another. The VSR method has higher intra- and inter-rater reliability when compared to listener-perceptual ratings using the VAS technique (Granqvist, 2003). In the present study, the number of voice samples chosen within a stimuli set, as well as the number of total sets for auditory-perceptual ratings, were comparable to previous studies using the VSR method to rate vocal effort (Heller Murray *et al.*, 2016; Lien *et al.*, 2015).

Different stimuli sets were generated for each listener. Each set consisted of nine different voice recordings from nine different speakers. Within the nine recordings, eight of the positions were filled with two recordings from each voice condition (i.e., two typical, two mild, two moderate, two maximal, for a total of eight recordings). Since three speakers had extra voice recordings, these three instances were then placed into the ninth position of the set. Finally, the remaining position in each set was filled with a randomly selected recording, which was later used for intra-rater reliability calculations. This randomization scheme resulted in 26 randomized stimuli sets, each with nine recordings. Twenty-three recordings (approximately 10% of the sample) were repeated for reliability. The randomization was completed for every listener, resulting in different stimuli sets for each listener.

C. Participants (listeners)

Twenty adults (11 female; $M = 20.7$ years, $SD = 2.8$ years) were recruited as inexperienced listeners for the study. Inexperienced listeners were chosen since previous studies reported no effect of listener experience on ratings of vocal effort when training is provided (Eadie *et al.*, 2010). Listeners were speakers of Standard American English with no reported history of speech, language, hearing, or voice disorders, as well as no prior experience with voice disorders. All listeners passed a hearing screening of pulsed pure tones (Burk and Wiley, 2004) at 25 dB hearing level (HL) at frequencies of 125, 250, 500, 1000, 2000, 4000, and 8000 Hz (Schlow, 1991) with over-the-ear headphones. With the approval of the Boston University Institutional Review Board, informed consent was obtained from each participant prior to participation in the study.

D. Listener training and protocol

Listeners were seated in a sound-treated room for the duration of the study. Prior to the experimental auditory-perceptual ratings, listeners were provided with a definition of vocal effort via the script: "You are going to hear a series of voice samples. Some will be of typical speaking voices and some will have increased vocal effort. Vocal effort is considered an exertion of the voice. It may sound like the speakers are trying to push their air out and strain to produce voice." Next, listeners were provided with familiarity samples of two different speakers (one male, one female) reading the second sentence of the Rainbow Passage. The familiarity samples included a voice recording at a typical speaking voice, and then the same speaker repeating the sentence in an effortful voice. The voice samples were not anchored to an effort scale, as their sole purpose was to provide an auditory example of vocal effort.

Participants completed a single VSR training module with /ifi/ recordings of various vocal effort levels, recorded separately from the experimental data set. The training module allowed the listeners to familiarize themselves with the interactive computer program as well as rating vocal effort on non-word productions (e.g., /ifi/). The listeners were trained to interact with a custom MATLAB VSR interface. The interface had nine voice samples located at the same horizontal level on the screen and a vertical axis to rate vocal effort. The top of the vertical axis was anchored at "100" and described as "The Most Effort," while the bottom of the axis was anchored at "0" and described as "No Effort" (see Fig. 1). First, participants were instructed to listen to the voice stimuli and sort the stimuli vertically so that stimuli of similar vocal effort were near the same vertical level. Then, participants were instructed to re-listen to the stimuli and rate the stimuli against each other to make small adjustments to the amount of vocal effort perceived in each recording.

Following the familiarity samples and VSR interface training, listeners progressed to the experimental VSR paradigm. Each participant wore over-the-ear headphones (Sennheiser HD 280 Pro) and the set-up was calibrated to a presentation level of an average of 76 dB SPL. The calibration procedure did not eliminate variation in dB SPL within or between samples, but set an average listening level. Listeners were allowed to listen to each recording as many times as they wished. Rest breaks were built into each session at 20 min increments. In general, participants were able to complete 8–10 sets every 20 min. The entire session, including consent, hearing screening, training, and auditory-perceptual ratings, lasted approximately 1.5 h.

E. Acoustical data processing

1. Mean SPL

In order to calculate mean SPL for the voicing segments of each /ifi/ production, the onset and offset of each vowel was determined via an algorithm developed for the neck-surface accelerometer signal captured concurrently with the

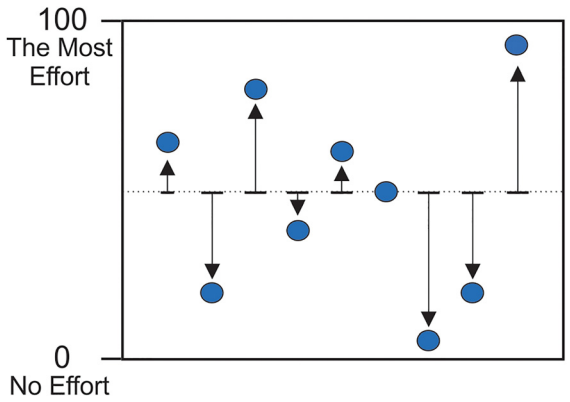

FIG. 1. (Color online) The listeners were presented with an interface that had nine different voice recordings (circles) at the midpoint of the screen, designated here as the dotted line. After listening to the stimuli, listeners moved each stimulus up or down (arrows) from the midline of the screen to sort them, and then made ratings against stimuli in the same area of the screen. The dotted line and arrows were not seen by the listeners, but are used in this image to depict the range of movement on the screen.

microphone signal. The accelerometer signal was full-wave rectified and filtered using a first-order low-pass Butterworth filter at 12 Hz. Then, to establish voicing onset and offset, a threshold was determined as four times the mean of 500 ms of quiet rest in the filtered signal for each recording. The threshold was determined empirically and verified via visual inspection of all waveforms. The root-mean-square (rms) was calculated in the time-aligned segments of the microphone signal that corresponded to the vowel segments in the accelerometer signal.

Once the rms of each vowel was determined (rms_{mic}), the rms_{mic} was converted to dB SPL based on the known dB SPLs from the calibration procedure. First a regression formula was created between the rms of the electrolaryngeal pulses made at the lips to the known dB SPL acquired from the sound level meter at the microphone. Then, the slope and intercept of that regression line ($Slope_{ref}$ and $Intercept_{ref}$) were used to predict mean SPL for each rms_{mic} [see Eq. (1)]

$$Mean\ SPL\ (dBSPL) = Slope_{ref} \times (20log_{10}(rms_{mic})) + Intercept_{ref}. \quad (1)$$

2. Mean f_o

An autocorrelation function in Praat (v.5.4.04; Boersma, 2001) was used to determine the mean f_o for each vowel (Boersma, 1993). Prior to analysis, the pitch range was adjusted to 60–300 Hz for male speakers and 90–500 Hz for female speakers (Vogel *et al.*, 2009). Mean f_o values were verified by visually examining the autocorrelation pulses provided in the acoustic waveforms in Praat. Each of these values were averaged for each voice recording. The mean f_o during the typical speaking conditions were averaged together as a reference for each speaker. Then, each mean f_o (measured in Hz) for each condition was converted to semitones (ST) relative to the speaker's average from the typical condition. The conversion to ST allows for comparison across speakers who may have different mean f_o values (e.g., pitch differences between men and women). This final mean f_o was considered representative of a change in ST from each speaker's typical vocal production.

3. HNR

HNR (dB) was determined for each vowel via an algorithm implemented in Praat (Boersma, 1993; Severin *et al.*, 2005). HNR was calculated from the harmonicity function, which is a forward cross correlation that uses the time-domain to determine the strength of the energy in the first harmonic (H1) relative the energy in the rest of the signal [see Eq. (2)]. HNR values were averaged over each voice recording. Of note, the choice to use the entire vowel segment during HNR calculations could increase variability in the HNR measure due to inclusion of the onset and offset voicing cycles (instead of just vowel steady-state). This decision was made due to the relatively short vowel segments in the /ifi/ utterance (compared to that of a sustained vowel which allows for identification of longer durations of the steady-state portion of the signal). The analysis was implemented consistently across all speakers in the study, making the measurements directly comparable to one another,

$$HNR\ (dB) = 10 \times log_{10}\left(\frac{Energy\ in\ H1}{1 - Energy\ in\ H1}\right). \quad (2)$$

4. RFF

RFF values were determined for each of the last ten cycles from the initial vowel, known as *offset cycles*, and then for the first ten voicing cycles of the following voiced segment, referred to as *onset cycles*. Each RFF cycle value is calculated by determining the instantaneous f_o of the cycle (the inverse of the period), normalizing that to the instantaneous f_o of a reference cycle that is closest to the midpoint of each vowel (i.e., offset cycle 1, onset cycle 10), and then converting to ST [see Eq. (3)]. Therefore, each RFF cycle reflects a change in ST from the instantaneous f_o of the vowel steady-state and can only be compared to other cycles in the same position (i.e., offset cycle 10 should only be compared to another offset cycle 10).

$$RFF\ (ST) = 39.86 \times log_{10}\left(\frac{cycle\ f_o}{reference\ f_o}\right). \quad (3)$$

RFF offset and onset values were calculated for each /ifi/ production via a custom MATLAB algorithm (Lien *et al.*, 2017). RFF offset cycle 10 and onset cycle 1 (the cycles closest to the fricative /f/) were targeted for further analysis due to their hypothesized relevance to laryngeal tension and vocal effort (Eadie and Stepp, 2013; Heller Murray *et al.*, 2017; Lien *et al.*, 2015; McKenna *et al.*, 2016; Stepp *et al.*, 2011). RFF offset cycle 10 and onset cycle 1 were individually averaged across the productions in each voice recording. The present study required a minimum of two cycle values

for averaging across each recording for further inclusion in the statistical analysis.

5. CPP and Cepstral peak standard deviation (CPP SD)

Cepstral analyses were completed using Analysis of Dysphonia in Speech and Voice (ADSV) software (model 5109, V. 3.4.2). Prior to analysis, each /ifi/ production was cropped to eliminate any non-speech segments in the sample by visual inspection of the acoustical signal. The program further used vocalic detection to eliminate voiceless /f/ segments (Awan, 2011). The software downsamples the acoustic signal to 25 kHz and determines the cepstrum of the signal (i.e., the FFT of the logarithm power spectrum) using a series of Hamming windows with a window length of 1024 samples and 75% overlap. CPP and CPP SD were then calculated from a smoothed cepstrum (averaged over seven frames) with peak extraction ranges pre-specified to quefrency ranges that corresponded with 60–300 Hz for male speakers and 90–500 Hz for female speakers. CPP was calculated as the amplitude of the highest rahmonic peak (dB) compared to the amplitude of the quefrency point on the regression line of the averaged power cepstrum (Awan and Roy, 2005; Awan *et al.*, 2010). In order to verify that CPP extraction was within a quefrency range that corresponded to a reasonable mean f_o, the mean CPP f_o was compared to the mean f_o values determined in Praat. For any instances in which CPP f_o varied more than 10% of the mean f_o from Praat, the sample was re-checked and excluded if suspected to be inaccurate. CPP and CPP SD were each averaged for every voice recording.

6. L/H Ratio and L/H SD

The L/H ratio, a ratio of low to high spectral energy, and L/H SD were calculated for each /ifi/ production using ADSV software. The software downsamples the time-domain signal to 25 kHz, creates a series of Hamming windows (1024 samples, 75% overlap), and uses the FFT to convert the original signal to the frequency domain (Awan, 2011). The L/H ratio was calculated from the spectrum (Awan *et al.*, 2010) with a ratio cut-off of 4000 Hz (Hillenbrand and Houde, 1996; Lowell *et al.*, 2013). L/H ratio and L/H SD were averaged for each participant for each speaking condition.

F. Statistical analysis

1. Listener reliability

Intra-rater reliability was calculated from the repeated stimuli (10% randomly selected voice samples) using a two-way intraclass correlation coefficient (ICC). Inter-rater reliability was analyzed on all samples across all listeners with an ICC two-way analysis for consistency, as well as an analysis of means for the group of listeners. Reliability analyses were completed with the statistical package R (ver. 3.2.2). Following reliability analyses, raw values (0–100) were averaged across listeners, resulting in a single averaged value for each voice recording.

2. Statistical models

Statistical analyses were completed in Minitab statistical software (ver. 18). Per-speaker Pearson product-moment correlation coefficients (r) were determined between self-ratings of vocal effort and the averaged listener ratings. To compare the findings of the present study to previous studies that analyzed individual acoustical predictors to ratings of vocal effort, a series of mixed effect regression models were analyzed for each acoustical variable separately against speaker and averaged listener ratings [see Eq. (4)]. The coefficient of determination (adjusted R^2) was calculated for each model to provide information on the amount of variance each acoustical measure contributed to the ratings on an individual basis.

$$Acoustical\ measure + Speaker(random) = Perceptual\ Rating. \quad (4)$$

A comprehensive analysis of all acoustical predictors was completed using two separate mixed-effect linear regression models. The predictor variables were the acoustic measures and "speaker" (random factor). The outcome measure of the first model was the self-perception of vocal effort as rated on the 100 mm VAS. The outcome of the second model was the averaged listener-perceptual ratings completed during the VSR task. The significance level was first set to $p < 0.05$, but because the same acoustical measures were used as predictors in both models, the p-value was reduced to $p < 0.025$ to minimize type I error. The coefficient of determination (adjusted R^2) was calculated for each model and the beta coefficients were examined. A subsequent analysis of the adjusted sum of squares of the predictors allowed for the calculation of predictor effect sizes (η_p^2), which partial out the contribution of each predictor with respect to the other predictors in the model. Effect sizes are reported for significant predictors only.

III. RESULTS

Speakers produced a total of 211 voice recordings for analysis. Across the nine acoustical measures, only data points from RFF offset cycle 10 and onset cycle 1 were missing from the data set (when less than two values were available for averaging). Missing RFF values may be due to instances of excessive glottalization or when there are fewer than ten vocal cycles in the vowel segment (Lien and Stepp, 2014). The missing data points accounted for 12% of the possible RFF values and were evenly distributed between offset and onset values, with 26 and 27 missing values, respectively. A total of 13 participants (half of the sample) had a missing RFF value and missing values did not appear to be affected by the voicing condition since 21% of the missing values occurred during the typical voicing condition, 28% occurred during productions of mild effort, 25% during the moderate effort condition, and the remaining 26% in the maximal effort condition. In total, 1846 acoustical data points were analyzed (211 recordings × 9 variables – 53

TABLE I. Summary of mean and SD for acoustical and perceptual measures for all voice conditions. Note: SPL = sound pressure level; RFF = relative fundamental frequency; ST = semitone; CPP = cepstral peak prominence; L/H = low-to-high; HNR = harmonics-to-noise ratio; f_o = fundamental frequency.

	Voice Condition Mean (SD)			
Measure	Typical	Mild Effort	Moderate Effort	Maximal Effort
Mean SPL (dB SPL)	80.27 (3.15)	81.24 (3.16)	83.03 (3.42)	85.05 (4.13)
RFF Offset 10 (ST)	−0.51 (0.99)	−0.64 (1.08)	−1.07 (0.91)	−1.50 (1.29)
RFF Onset 1 (ST)	2.45 (0.81)	2.21 (0.60)	2.23 (0.86)	2.00 (0.77)
CPP (dB)	5.48 (1.18)	5.39 (0.92)	5.70 (0.83)	5.76 (0.83)
CPP SD (dB)	1.38 (0.49)	1.44 (0.45)	1.62 (0.42)	1.72 (0.37)
L/H Ratio (dB)	38.72 (2.42)	37.22 (2.93)	35.69 (3.22)	34.47 (3.78)
L/H SD (dB)	9.78 (1.94)	10.13 (2.00)	10.80 (2.07)	10.73 (2.17)
HNR (dB)	16.35 (3.78)	15.39 (3.96)	15.81 (3.99)	15.34 (4.02)
Mean f_o (ST)	0.00 (0.00)	0.33 (1.66)	1.20 (1.53)	2.32 (2.41)
Speaker Rating (0–100)	14.39 (11.74)	26.09 (10.34)	43.77 (12.96)	68.63 (19.61)
Listener Rating (0–100)	20.60 (7.04)	35.92 (15.74)	50.19 (19.60)	64.25 (19.97)

missing RFF values). Table I provides the mean and SD of the acoustical measures across each voice condition. Correlations between all acoustical measures can be found in the Appendix.

A. Listener reliability

The average intra-rater reliability across all twenty listeners was ICC (2,1) = 0.82 (SD = 0.08) and range of ICC = 0.62–0.93. Inter-rater reliability analysis resulted in ICC (2,1) = 0.73 [95% confidence interval (CI) = 0.69–0.77]. When calculating an ICC(2,20) for means, the inter-rater increased to a coefficient of 0.98 (95% CI = 0.98–0.99). These reliability findings are remarkably similar to previous reports on the inter- and intra-rater reliability of inexperienced listeners rating vocal effort, with intra-rater reliability or $r = 0.84$, and interrater agreement of 77% (Eadie *et al.*, 2010).

B. Speaker and listener ratings

Pearson product-moment correlations were calculated for each speaker to determine the relationship between self- and listener-perceptual ratings. The ratings met the assumptions of the parametric testing (e.g., absence of outliers, linearity, normality). The average correlation across speakers was $r = 0.86$ (*median* = 0.92, range = 0.20–0.99). Twenty of the speakers appeared to follow roughly the same linear trend with a strong correlation across the subset of speakers ($r = 0.85$; Panel D of Fig. 2). Figure 2 provides a visualization of the relationships between speaker and averaged listener ratings.

C. Mixed-effects regression models

Individual mixed-effects regression models were analyzed for each acoustical measure and the two ratings of vocal effort. Table II provides a list of adjusted R^2 for each model to the separate outcome variables of speaker rating and averaged listener rating. For both speaker and listener models, mean SPL accounted for a substantial portion of the variance with adjusted R^2 equal to 0.64 and 0.70, respectively. L/H ratio, mean f_o, and RFF offset cycle 10 were moderate predictors of listener ratings (adjusted $R^2 = 0.46 - 0.57$), while only the L/H ratio accounted for a moderate amount of variance in the model for speakers (adjusted $R^2 = 0.42$). All other acoustical variables accounted for less than 40% of the variance in the models and are considered weak predictors of vocal effort for both speakers and listeners when analyzed in isolation.

Two separate mixed-effects regression models were calculated to analyze the relationship between all of the acoustical measures and the speaker and listener ratings of vocal effort. All acoustical variables were normally distributed. Each was determined to be linearly related to the outcome variables via visual inspection of the distribution of the residuals in the individual mixed-effects models described above. CPP SD, however, revealed a high variance inflation factor, indicating a violation of multicollinearity (Hair *et al.*, 1995). Consequently, CPP SD was removed from the models, resulting in a reduction of acoustical predictors to eight, instead of the original nine acoustical measures.

The first mixed-effects regression model determined the relationship between the remaining eight acoustical measures and the self-perceptual ratings of vocal effort. Results revealed that mean SPL, L/H ratio, and HNR were significant predictors of vocal effort. The acoustical measures accounted for 72% variance in the model (adjusted $R^2 = 0.72$). Mean SPL had a large effect size of $\eta_p^2 = 0.36$, whereas L/H ratio and HNR both had medium effect sizes (Witte and Witte, 2010). Examination of the beta coefficients revealed that mean SPL increased as the self-perception of effort increased, whereas L/H ratio and HNR decreased with increased ratings of the self-perception of vocal effort.

The second model determined the relationship between averaged listener-perceptual ratings of vocal effort and the same eight acoustical predictors. Mean SPL, HNR, L/H ratio, mean f_o, and RFF offset cycle 10 were significant predictors of listener ratings of vocal effort and accounted for 82% of the variance in the model (adjusted $R^2 = 0.82$). Mean SPL and mean f_o had positive relationships with listener ratings, while L/H ratio, HNR, and RFF offset cycle 10 all decreased as vocal effort increased. Table III provides a list

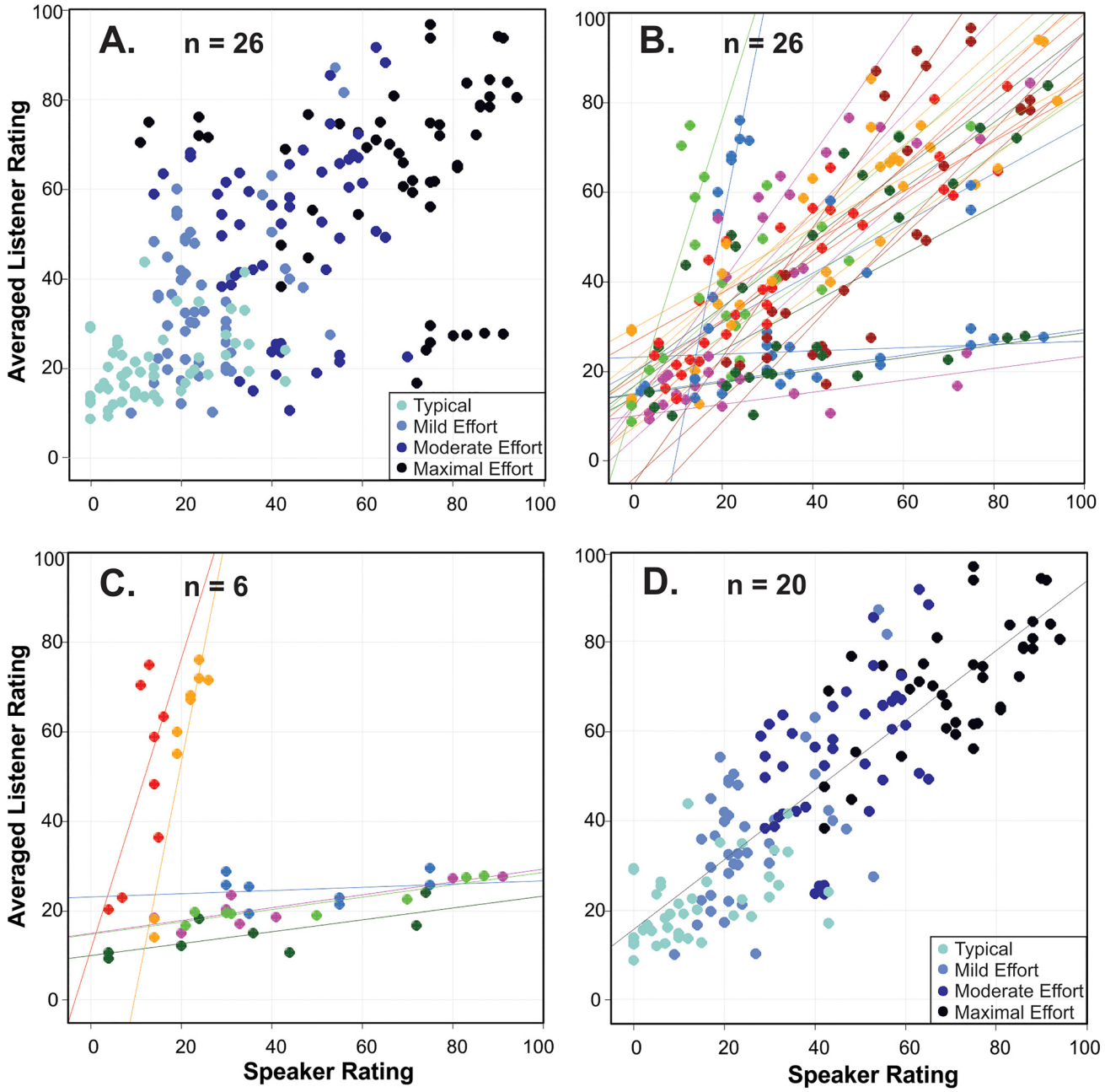

FIG. 2. (Color online) Scatterplot of speaker self-perceptual ratings to averaged listener ratings of vocal effort. Plot A provides a visualization of all the raw data. Plot B provides the raw data with lines of best fit for each speaker. Plot C provides a separate visualization of six participants who do not follow the same linear trends as the main group of speakers. Plot D provides a visualization of 20 speakers who appear to all follow a similar linear trend between the two ratings.

of statistical values with effect sizes calculated for the significant acoustical predictors.

IV. DISCUSSION

The aim of the present study was to examine the acoustical manifestation of vocal effort and determine the relationship between the speech signal and perceptual ratings of vocal effort. We examined a large set of acoustical predictors since many acoustical measures have been proposed to be related to the perception of vocal effort for both speakers and listeners. We hypothesized that speakers and listeners would make judgments of vocal effort based on separate acoustical cues. Our hypothesis was supported when there were different acoustical predictors for listener ratings (mean f_o and RFF offset cycle 10) that were not significant predictors of speaker ratings.

A. Acoustical correlates of vocal effort

When analyzed on an individual basis in separate mixed-effects models, the acoustical predictors behaved as expected. There was a wide range of predictive strength and many acoustical predictors revealed moderate-to-strong relationships with perceptual ratings. For the listener models, the adjusted R^2 values ranged from 0.23 to 0.70,

TABLE II. Adjusted coefficient of determination (R^2) for each mixed-effects regression model between individual acoustical predictors and ratings of vocal effort. Note: SPL = sound pressure level; L/H = low-to-high; f_o = fundamental frequency; ST = semitone; CPP = cepstral peak prominence; SD = standard deviation; RFF = relative fundamental frequency; HNR = harmonics-to-noise-ratio.

	Adjusted R^2	
Acoustical Measure	Speaker Rating	Listener Rating
Mean SPL (dB SPL)	0.64	0.70
L/H Ratio (dB)	0.42	0.57
Mean f_o (ST)	0.39	0.54
CPP SD (dB)	0.30	0.39
RFF Offset 10 (ST)	0.27	0.46
RFF Onset 1 (ST)	0.13	0.29
L/H SD (dB)	0.11	0.28
CPP (dB)	0.11	0.23
HNR (dB)	0.08	0.25

and four of the nine acoustical measures accounted for more than 40% of the variance in each model. Conversely, the acoustical measures did not account for the same amount of variance when predicting speaker perceptual ratings, lending some initial support to potential differences between speaker and listener perceptions of vocal effort. The speaker models had a smaller range of R^2 values (0.08–0.64) and only two predictors with R^2 values greater than 0.40. These results further highlight the need for combined models to evaluate multiple acoustical variables concurrently to understand which are the most salient to the perceptual ratings, and to tease out how speaker and listener perceptual judgements may be influenced by different features of the acoustical signal.

In the combined acoustical models, the acoustical measures of mean SPL, L/H ratio, and HNR were significant predictors of both self- and listener-perceptual ratings of vocal effort. The speakers in this study were instructed to increase vocal effort while maintaining the same vocal volume in order to simulate increased vocal effort in a comfortable speaking environment. Despite this instruction, the speakers increased their vocal intensity by an average of 5 dB SPL across all vocal conditions. This is slightly greater than a prior report of an increase of 3 dB SPL during modulations of vocal effort in the study by Rosenthal *et al.* (2014). However, a typical speaking voice can easily produce a vocal intensity range of up to 6–7 dB SPL (Schmidt *et al.*, 1990). Thus, the speakers in the present study appeared to use a functional range of mean SPL comparable to that of conversational speech. Results confirm that mean SPL is a strong acoustical cue to indicate vocal effort for both speakers and listeners, even when kept within a functional intensity range. It is likely that these increases in mean SPL were perceived in combination with other changes to the acoustical signal, assisting in cueing the speakers and listeners to the perception of vocal effort.

The L/H ratio is reflective of an overall proportion of low-to-high frequency information, but the ratio does not provide information about the periodicity of the energy in the signal. It is generally assumed that increased high frequency energy is due to aspiration noise, supported by prior studies examining the energy in different frequency bands and simultaneous changes to glottal configuration (Klatt and Klatt, 1990). If the changes in high frequency energy were due to aperiodic noise, the L/H ratio would decrease and there would be a concurrent reduction in HNR values as well (i.e., the results of the present study). Therefore, we can infer that increased vocal effort acts to increase aperiodic high frequency energy in the acoustical signal. The physiological basis of this change may be due to adjustments to glottal configuration and/or reduced periodicity of vocal fold vibration (Boone *et al.*, 2014). These could be due to increased or imbalanced laryngeal muscle activity, which has been reported in specific patient populations with vocal effort (e.g., vocal hyperfunction; Hillman *et al.*, 1989).

TABLE III. Statistical outcomes for each mixed-effects regression model. Effect sizes and interpretations are placed for significant predictors only. Note: Coef. = Coefficient; SE = standard error; SPL = sound pressure level; L/H = low-to-high; HNR = harmonics-to-noise-ratio; SD = standard deviation; RFF = relative fundamental frequency; f_o = fundamental frequency; CPP = cepstral peak prominence.

Model	Acoustic measure	Coef.	SE Coef.	t-value	*p*-value	Effect Size (η_p^2)	Effect Size Interpretation
Speaker	Mean SPL	6.76	0.79	8.59	<0.001	0.36	Large
	L/H Ratio	−2.21	0.61	−3.64	<0.001	0.09	Medium
	HNR	−2.80	0.79	−3.54	0.001	0.08	Medium
	L/H SD	−1.98	0.88	−2.25	0.026	–	–
	RFF Offset Cycle 1	−2.02	1.53	−1.32	0.188	–	–
	Mean f_o	1.51	1.26	1.20	0.231	–	–
	CPP	2.45	2.20	1.12	0.267	–	–
	RFF Offset Cycle 10	1.76	1.90	0.93	0.355	–	–
Listener	Mean SPL	4.27	0.56	7.64	<0.001	0.31	Large
	HNR	−2.66	0.56	−4.73	<0.001	0.15	Medium
	L/H Ratio	−1.62	0.43	−3.76	<0.001	0.09	Medium
	Mean f_o	2.22	0.89	2.49	0.014	0.05	Small
	RFF Offset Cycle 10	−3.29	1.35	−2.44	0.016	0.04	Small
	L/H SD	1.25	0.63	2.00	0.048	–	–
	CPP	−2.97	1.56	−1.91	0.059	–	–
	RFF Onset Cycle 1	−1.06	1.08	−0.98	0.330	–	–

It is somewhat surprising that HNR was a significant predictor for speakers and listeners in the combined model. When examined alone as the only acoustical predictor, the relationships reported between HNR and self- and listener-perceptual ratings were weak (i.e., $R^2 = 0.08$ and 0.25, respectively). In order for this to occur statistically, the other regressors must be correlated with one another, reducing their overall importance in the final model (measured via effect size). Review of per-speaker correlations between each acoustical measure and HNR (see the Appendix for a complete list) revealed weak correlations of average $r = 0.01$–0.17. These are considerably lower than some of the other reported within-speaker correlations between the other acoustical predictors (i.e., mean SPL and mean f_o were correlated an average of $r = 0.66$). The independence of this measure from the other acoustical variables contributed to its medium effect size in both of the combined models.

Although the results of the two statistical models revealed similar significant acoustical predictors, two time-based measures (mean f_o and RFF offset cycle 10) were significant predictors of only listener ratings. Mean f_o increased as listener ratings of vocal effort increased with a small effect size. Previous work on pitch discrimination has shown that listeners are able to distinguish a change between two presented tones (just noticeable difference task) at about 0.5 ST (Nikjeh *et al.*, 2009). The change in mean f_o, an average increase of 2.3 ST from typical to maximal vocal effort, would have been perceptible to the listeners and provided an additional acoustical cue for judgments of vocal effort.

The findings that mean f_o was significant to listener ratings, but not speaker ratings, could be due to a shared acoustical representation between vocal effort and vocal fatigue. Researchers have proposed that vocal effort is proportional to vocal fatigue in which increasing fatigue produces simultaneous changes in vocal effort (Chang and Karnell, 2004; Somodi *et al.*, 1995). As such, it follows that the acoustical representation of vocal effort and fatigue may be similar. Evidence shows that mean f_o increases following vocal loading and vocally fatiguing tasks (Laukkanen *et al.*, 2008; Rantala *et al.*, 1998; Stemple *et al.*,1995; Vilkman *et al.*, 1999). We propose that listeners may have focused on increases in mean f_o due to this relationship. Since the speakers in the present study were not likely to be experiencing vocal fatigue as they were healthy speakers and had not completed a vocal loading task, we suspect the speakers did not use this acoustical cue when rating their own vocal effort. This may have led to a discrepancy between the acoustical predictors in each model.

RFF offset cycle 10 was also a significant predictor of listener ratings of vocal effort, albeit with a small effect. These results are consistent with previous reports of weak-to-moderate relationships between RFF offset 10 and listener-perceptual ratings of vocal effort (Lien *et al.*, 2015). RFF offset cycles are hypothesized to be affected by abduction of the vocal folds and intrinsic laryngeal tension during the offset of voicing. A study by Heller Murray *et al.* (2017) proposed that increased intrinsic laryngeal tension results in a reduction of abductory behavior, causing longer vocal fold contact time at the offset of voicing. This results in slower vibrational cycles and lower RFF offset 10 values. In that study, speakers with non-phonotraumatic vocal hyperfunction (i.e., muscle tension dysphonia) had RFF offset cycle 10 values equal to −1.35 ST and those with phonotraumatic vocal hyperfunction had slightly lower RFF offset cycles 10 values of −1.76 ST. In the present study, the maximal effort condition had an average RFF offset cycle 10 values of −1.5 ST, which is markedly similar to the results of Heller Murray and colleagues. Thus, it is possible that the reduction of RFF offset cycle 10 values in the present study are due to similar mechanisms between speakers with vocal hyperfunction and vocally healthy speakers who are purposefully increasing vocal effort. Why the perception of offset cycle 10 was significant predictor for listener ratings and not speaker ratings is a question that warrants further investigation.

Many of the acoustical measures calculated in the present study were not significant predictors of vocal effort in either model. For example, CPP was not predictive of changes in vocal effort for speakers or listeners. Previous studies are equivocal as to whether instances of dysphonia and vocal effort act to increase, or decrease, CPP values. Numerous studies have found associations between CPP and overall dysphonia, with decreases in the relative strength of the first rahmonic in dysphonic voices (Awan *et al.*, 2014b; Awan *et al.*, 2010; Lowell *et al.*, 2012). When a study by Rosenthal *et al.* (2014) specifically examined the impact of vocal effort on CPP, the results determined that CPP values *increased* during effortful voice productions. Other work has determined that increased mean SPL may result in a stronger, more steady rahmonic energy (Awan *et al.*, 2012). Examination of CPP values in the present study did not reveal any trends across voice conditions and furthermore, average CPP values did not meet the cut-off criterion indicating a dysphonic vocal quality (e.g., 4 dB; Heman-Ackah *et al.*, 2014). These findings have significant implications for future work as CPP has been the focus of many studies investigating the relationship between speech acoustics and vocal effort following vocal loading tasks (Fujiki *et al.*, 2017; Sundarrajan *et al.*, 2017). The findings here would indicate that CPP is not an acoustical variable salient to the perception of vocal effort for speakers or listeners.

B. Listener vs speaker ratings of vocal effort

Results showed that listener intra-rater reliability measures were considered moderate-to-excellent (ICC = 0.62 – 0.93) and inter-rater reliability was deemed moderate as well (Koo and Li, 2016). The VSR technique may have improved reliability by allowing the listeners to directly compare voice samples instead of only rating a single voice sample at a time (e.g., VAS tasks). Furthermore, the listeners in the present study were provided familiarity samples of vocal effort, which could have assisted in cueing the listeners to the perceptual qualities of vocal effort. The samples may have also acted to confirm a previously established internal

auditory representation of vocal effort, improving listener reliability and confidence.

Researchers have also reported concerns that listeners may have difficulty distinguishing vocal effort from overall dysphonia severity (Stepp *et al.*, 2012). The findings in the present study do not appear to support that hypothesis. CPP, a strong correlate to overall dysphonia (Awan *et al.*, 2014b; Awan *et al.*, 2010; Lowell *et al.*, 2012), was not a significant predictor of listener ratings of vocal effort. When evaluated on an individual acoustical basis, CPP only accounted for a small amount of variance in listener ratings ($R^2 = 0.23$). Furthermore, the lack of change in CPP values across voice tasks and its weak relationship with listener ratings provides evidence that the speakers and listeners were judging vocal effort instead of vocal strain. CPP is consistently a significant predictor of vocal strain (e.g., Anand *et al.*, 2018; Lowell *et al.*, 2012), which is in direct opposition to the findings here.

The average Pearson product-moment correlation coefficients between self- and listener-perceptual ratings were very strong (mean $r = 0.86$, median $r = 0.92$), indicating that speakers and listeners have similar acoustical representations of vocal effort. These relationships exceed those of previous studies that report weak-to-moderate relationships between speaker and listener perceptual ratings of vocal effort (Eadie *et al.*, 2010; Eadie *et al.*, 2007). It may have been that vocal effort is easier to perceive in vocally healthy speakers who do not present with other conflating percepts of voice compared to speakers with voice disorders. It is also possible that the strong relationship between ratings was due to the parallel instructions provided to both groups during the production and perception tasks.

Prior work has shown that speakers report greater degrees of vocal effort when directly compared to listener ratings (Lane *et al.*, 1961). In the present study, there were no consistent trends of which to conclude that one rating was greater than the other. Inspection of the relationship between the speaker and listener ratings revealed that 20 of the 26 speakers had a similar linear trend with a slope of $\beta = 0.79$ and a correlation of $r = 0.85$ (refer to Panel D of Fig. 2). The other six speakers did not appear to display the same relationship between self- and listener-perceptual ratings. Four speakers reported changes in self-perception of vocal effort that were not reflected in the listener-ratings. These speakers exhibited much shallower slopes ($\beta = 0.04$–0.14) compared to the larger group of 20 speakers. Review of their data revealed that two of these participants tended to *decrease* mean f_o while increasing vocal effort, another exhibited almost no change in mean SPL across all productions (range = 2 dB), and the last exhibited *positive* RFF offset cycle 10 values. All of these acoustical differences could have influenced listener-perceptual ratings of these speakers and led to the discrepancy between ratings.

Conversely, two speakers reported lower variation in their vocal effort, whereas the listeners perceived the speakers' vocal effort as much greater ($\beta = 3.36$ and 5.12). Review of these participants' data did not reveal any trends in their acoustical measures that may have contributed to perceptual ratings. Thus, based on the evidence in this study, we hypothesize that these speakers may have relied more on somatosensory feedback than auditory feedback during their self-ratings, which may not have been captured in the acoustical signal.

Prior work in articulatory motor control has identified sensory preferences for different speakers. A study by Lametti *et al.* (2012) evaluated the degree of compensatory response to simultaneous perturbations in sensory (jaw) and auditory (first formant) feedback during speech. Results indicated that speakers who compensated more for perturbations in auditory feedback responded less to perturbations in sensory feedback. A review of speaker sensory preferences revealed an uneven distribution in which 53% responded only to auditory perturbations, 26% responded to both auditory and somatosensory perturbations, and 21% responded only to somatosensory perturbations. It is currently unknown how many speakers may rely solely on auditory feedback, solely on somatosensory feedback, or both, when making judgments of vocal effort. Auditory perturbation paradigms have identified individuals who are reliant on auditory feedback, by responding to perturbations of pitch and intensity (Bauer *et al.*, 2006; Behroozmand *et al.*, 2012; Burnett *et al.*, 1998). Still, there continues to be a small proportion of speakers who show no vocal compensation to changes in auditory feedback (Larson *et al.*, 2007). A few studies have evaluated the impact of direct sensory perturbations to the larynx (Loucks *et al.*, 2005; Sapir *et al.*, 2000), yet no study has evaluated concurrent sensory and auditory feedback perturbations to determine sensory preference in vocal control. Our results indicated that 6 of the 26 speakers (approximately 23%) reported self-perceptual ratings of vocal effort that were not consistent with listener-perceptual ratings. This proportion is similar to the 21% of speakers in the study by Lametti *et al.* (2012) who preferred to only respond to somatosensory feedback perturbations. We suspect that vocal motor control may be driven by similar feedback systems as speech motor control in which speakers have sensory preferences affecting their vocal behavior and self-perception.

C. Limitations and future directions

This study analyzed acoustical recordings from vocally healthy speakers who were purposefully increasing vocal effort. Although healthy speakers, especially individuals with high voice use, have reported increased vocal effort during daily tasks, these are not speakers with diagnosed voice disorders. It is possible that speakers who exhibit vocal fatigue and vocal effort to the point of dysphonic voice changes may exhibit different acoustical manifestations of vocal effort. However, we do not think that the results described in the present study are completely irrelevant to those with voice disorders, since prior work comparing modulations in vocal quality in healthy speakers to those with voice disorders have reported similarities between acoustical measures. For example, Hillenbrand *et al.* (1994) examined the acoustical correlates of breathiness in vocally healthy speakers. The researchers then completed a follow-up study on speakers with voice disorders and found strikingly similar acoustical

manifestations of breathiness between the healthy speakers modulating vocal quality and those with voice disorders (Hillenbrand and Houde, 1996). Therefore, it is possible that the findings in the present study may overlap with the acoustical manifestations of vocal effort in some speakers with voice disorders; specifically, we suggest future work first investigate speakers with high voice use, non-phonotraumatic vocal hyperfunction, and glottal incompetence, because these speakers do not have structural changes to the vocal folds. The direct translation of the present findings to speakers with vocal fold lesions (e.g., nodules, polyps) or neurologically-based voice disorders (e.g., spasmodic dysphonia) requires further consideration and investigation.

All speaker recordings were completed in the same order: typical voice, mild effort, moderate effort, and then maximal effort. Preferably, a randomized elicitation technique would have mitigated the possibility of an order effect; however, we suspect that the elicitation order did not impact the results of the study. For the listening task, the stimuli were randomized within each set and between all listeners, limiting the possibility of an order effect for these ratings. The statistical results revealed a strong relationship between speaker and listener ratings of vocal effort and an overlap in the acoustical representations of vocal effort between the two groups, leading to the conclusion that the elicitation order did not impact the results of the study. Furthermore, all speaker acoustical recordings were collected under flexible laryngoscopy. It is possible that the laryngoscopy procedure may interfere with typical speaking patterns and induce stress and tension during recordings. The certified SLP who verified normal perceptual vocal quality also made judgments during the typical speaking condition under laryngoscopy as well as other recordings made without laryngoscopy. The SLP determined that all typical recordings were within normal limits and had no concerns that the laryngoscopy procedure changed vocal quality. Thus, although possible, we think it unlikely that the laryngoscopy procedure affected the vocal recordings in this study.

Researchers have determined that the length and type of stimuli can affect perceptual ratings (Barsties and Maryn, 2017; Bele, 2005; de Krom, 1994). The acoustical stimuli analyzed in this study were repetitions of the non-word utterance /ifi/. Importantly, the inter- and intra-rater reliability reported here were markedly similar to previously reported reliability values of inexperienced listeners rating vocal effort from full sentences (Eadie *et al.*, 2010). Still, it may be possible that vocal effort is more difficult to judge in non-word contexts, or possibly, easier to judge in this case as the listeners knew what stimuli to expect. Further work is needed to determine what kind of stimuli result in consistent inter- and intra-rater reliability during perceptual ratings of vocal effort.

The speaker and listener instructions were developed based on three objectives: (i) provide a definition of effort that could be used to describe vocal effort to both speakers and listeners, (ii) ensure proper understanding of vocal effort in both groups to decrease variance in perceptual measures, and (iii) ensure that vocal effort would be specific to the structures of the larynx (instead of other factors that may impact effort). With that said, these instructions may have acted to limit the free interpretation of vocal effort and reduce the ability to compare the present findings to other studies that did not use the same instructions. Furthermore, the instructions were similar to descriptions of vocal strain, which can be described as a hyperadduction of the larynx and/or excessive subglottal pressure (Netsell *et al.*, 1984); however, our acoustical results were not consistent with measures of strain (i.e., weak relationship with the measure of CPP), indicating that the speakers and listeners were in fact, perceiving and rating effort. We recommend future work focus on how the instructions provided to speakers and listeners may impact the perception of vocal effort.

Finally, it must be noted that approximately 12% of the RFF values were missing from the statistical analysis. Few studies have reported on how the number of missing RFF values impacts the accuracy of estimation and the clinical applicability of this measure. A study by Roy *et al.* (2016) assessed a large database of female speakers with muscle tension dysphonia, finding that RFF could only be determined 1.87 times out of three opportunities, or 62% of the time. Moreover, a study focused on the algorithmic calculation of RFF (the method used in the present study) found that the environment of the acquisition (sound-treated room vs quiet room) may have an impact on the number of calculable RFF values (Lien *et al.*, 2017). Therefore, it seems that missing RFF data is a common occurrence, but which factors contribute to the missing data and how many utterances are needed for averaging requires more study.

V. CONCLUSION

Vocal effort manifests as a series of changes to the speech signal, including those that can be quantified by amplitude-, time-, and spectral-based measures. There were strong relationships between inexperienced listener-perceptual ratings and speaker self-perceptual ratings of vocal effort, with an average correlation of $r = 0.86$. Likewise, there were similar acoustical predictors of self- and listener-perceptual ratings, which included mean SPL, L/H ratio, and HNR. However, listeners also used time-based acoustical cues when rating vocal effort (mean f_o and RFF offset cycle 10). The reason for the discrepancy between acoustical predictors in self- and listener-perception warrants further investigation and should be examined in speakers with voice disorders.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health Grant Nos. R01DC015570 (CES) and T32DC013017 (CAM), from the National Institute of Deafness and Other Communication Disorders. It was also supported by a Sargent College Dudley Allen Research Grant (VSM) from Boston University. We would like to thank Zachary Morgan and Ashling Lupiani for their assistance with acoustical data processing.

APPENDIX

TABLE IV. Averaged within-speaker correlations (r) and SD between acoustical measures. Note: SPL = sound pressure level; RFF = relative fundamental frequency; ST = semitone; CPP = cepstral peak prominence; L/H = low-to-high; HNR = harmonics-to-noise-ratio; f_o = fundamental frequency.

	Averaged Within-Speaker Correlations (r)								
	Mean SPL (dB SPL)	RFF Offset 10 (ST)	RFF Onset 1 (ST)	CPP (dB)	CPP SD (dB)	L/H Ratio (dB)	L/H SD (dB)	HNR (dB)	Mean f_o (ST)
Mean SPL (dB SPL)	1.00	−0.47(0.49)	−0.10(0.53)	0.39(0.49)	0.57(0.38)	−0.51(0.54)	0.27(0.54)	0.14(0.54)	0.66(0.38)
RFF Offset 10 (ST)	–	1.00	−0.01(0.49)	−0.13(0.57)	−0.31(0.48)	0.28(0.43)	−0.12(0.51)	0.01(0.46)	−0.43(0.49)
RFF Onset 1 (ST)	–	–	1.00	0.11(0.47)	−0.01(0.47)	0.13(0.47)	−0.21(0.37)	0.01(0.41)	−0.25(0.47)
CPP (dB)	–	–	–	1.00	0.70(0.29)	−0.12(0.47)	0.13(0.40)	0.17(0.42)	0.19(0.49)
CPP SD (dB)	–	–	–	–	10.00	−0.29(0.46)	0.28(0.37)	0.09(0.45)	0.29(0.40)
L/H Ratio (dB)	–	–	–	–	–	1.00	−0.40(0.48)	0.08(0.54)	−0.33(0.55)
L/H SD (dB)	–	–	–	–	–	–	1.00	0.10(0.43)	0.14(0.56)
HNR (dB)	–	–	–	–	–	–	–	1.00	0.36(0.54)
Mean f_o (ST)	–	–	–	–	–	–	–	–	1.00

Altman, K. W., Atkinson, C., and Lazarus, C. (**2005**). "Current and emerging concepts in muscle tension dysphonia: A 30-month review," J. Voice **19**(2), 261–267.

Anand, S., Kopf, L., Shrivastav, R., and Eddins, D. (**2018**). "Objective indices of perceived vocal strain," J. Voice (in press).

Awan, S. N. (**2011**). *Analysis of Dysphonia in Speech and Voice: An Application Guide* (KAYPENTAX, Montvale, NJ).

Awan, S. N., Giovinco, A., and Owens, J. (**2012**). "Effects of vocal intensity and vowel type on cepstral analysis of voice," J. Voice **26**(5), 670.e15–670.e20.

Awan, S. N., and Roy, N. (**2005**). "Acoustic prediction of voice type in women with functional dysphonia," J. Voice **19**(2), 268–282.

Awan, S. N., Roy, N., and Cohen, S. M. (**2014b**). "Exploring the relationship between spectral and cepstral measures of voice and the voice handicap index (VHI)," J. Voice **28**(4), 430–443.

Awan, S. N., Roy, N., Jette, M. E., Meltzner, G. S., and Hillman, R. E. (**2010**). "Quantifying dysphonia severity using a spectral/cepstral-based acoustic index: Comparisons with auditory-perceptual judgements from the CAPE-V," Clin. Ling. Phon. **24**(9), 742–758.

Bach, K. K., Belafsky, P. C., Wasylik, K., Postma, G. N., and Koufman, J. A. (**2005**). "Validity and reliability of the glottal function index," Arch. Otolaryngol. Head Neck Surg. **131**(11), 961–964.

Baldner, E. F., Doll, E., and van Mersbergen, M. R. (**2015**). "A review of measures of vocal effort with a preliminary study on the establishment of a vocal effort measure," J. Voice **29**(5), 530–541.

Barsties, B., and Maryn, Y. (**2017**). "The influence of voice sample length in the auditory-perceptual judgment of overall voice quality," J. Voice **31**(2), 202–210.

Bastian, R. W., Keidar, A., and Verdolini-Marston, K. (**1990**). "Simple vocal tasks for detecting vocal fold swelling," J. Voice **4**(2), 172–183.

Bauer, J. J., Mittal, J., Larson, C. R., and Hain, T. C. (**2006**). "Vocal responses to unanticipated perturbations in voice loudness feedback: An automatic mechanism for stabilizing voice amplitude," J. Acoust. Soc. Am. **119**(4), 2363–2371.

Behroozmand, R., Korzyukov, O., Sattler, L., and Larson, C. R. (**2012**). "Opposing and following vocal responses to pitch-shifted auditory feedback: Evidence for different mechanisms of voice pitch control," J. Acoust. Soc. Am. **132**(4), 2468–2477.

Bele, I. V. (**2005**). "Reliability in perceptual analysis of voice quality," J. Voice **19**(4), 555–573.

Boersma, P. (**1993**). "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," Institute of Phonetic Sciences, University of Amsterdam, pp. 97–110.

Boersma, P. (**2001**). "Praat, a system for doing phonetics by computer," Glot Int. **5**(9/10), 341–345.

Bogert, B. P., Healy, M. J. R., and Tukey, J. W. (**1963**). "The quefrency analysis of time series for echoes: Cepstrum, pseudo autocovariance, cross-cepstrum and saphe cracking," in *Processings of the Sumposium of Time Series Analysis*, edited by M. Rosenblatt (Wiley, New York), pp. 209–243.

Boone, D. R., McFarlane, S. C., Von Berg, S. L., and Zraick, R. I. (**2014**). *The Voice and Voice Therapy*, 9th ed. (Pearson, Boston, MA).

Borg, G. A. (**1982**). "Psychophysical bases of perceived exertion," Med. Sci. Sports Exer. **14**(5), 377–381.

Bottalico, P., Graetzer, S., and Hunter, E. J. (**2016**). "Effects of speech style, room acoustics, and vocal fatigue on vocal effort," J. Acoust. Soc. Am. **139**(5), 2870–2879.

Brinca, L., Nogueira, P., Tavares, A. I., Batista, A. P., Goncalves, I. C., and Moreno, M. L. (**2015**). "The prevalence of laryngeal pathologies in an academic population," J. Voice **29**(1), 130.e131–130.e139.

Burk, M. H., and Wiley, T. L. (**2004**). "Continuous versus pulsed tones in audiometry," Am. J. Audiol. **13**(1), 54–61.

Burnett, T. A., Freedland, M. B., Larson, C. R., and Hain, T. C. (**1998**). "Voice F0 responses to manipulations in pitch feedback," J. Acoust. Soc. Am. **103**(6), 3153–3161.

Cannito, M. P., Doiuchi, M., Murry, T., and Woodson, G. E. (**2012**). "Perceptual structure of adductor spasmodic dysphonia and its acoustic correlates," J. Voice **26**(6), 818.e5–818.e13.

Chang, A., and Karnell, M. P. (**2004**). "Perceived phonatory effort and phonation threshold pressure across a prolonged voice loading task: A study of vocal fatigue," J. Voice **18**(4), 454–466.

de Alvear, R. M. B., Baron, F. J., and Martinez-Arquero, A. G. (**2011**). "School teachers' vocal use, risk factors, and voice disorder prevalence: Guidelines to detect teachers with current voice problems," Folia Phoniatr. Logopaed. **63**(4), 209–215.

de Krom, G. (**1994**). "Consistency and reliability of voice quality ratings for different types of speech fragments," J. Speech Hear. Res. **37**(5), 985–1000.

Dworkin, J. P., Meleca, R. J., Simpson, M. L., and Garfield, I. (**2000**). "Use of topical lidocaine in the treatment of muscle tension dysphonia," J. Voice **14**(4), 567–574.

Eadie, T. L., Day, A. M. B., Sawin, D. E., Lamvik, K., and Doyle, P. C. (**2013**). "Auditory-perceptual speech outcomes and quality of life after total laryngectomy," Otolaryngol. Head Neck Surg. **148**(1), 82–88.

Eadie, T. L., Kapsner, M., Rosenzweig, J., Waugh, P., Hillel, A., and Merati, A. (**2010**). "The role of experience on judgments of dysphonia," J. Voice **24**(5), 564–573.

Eadie, T. L., Nicolici, C., Baylor, C., Almand, K., Waugh, P., and Maronian, N. (**2007**). "Effect of experience on judgments of adductor spasmodic dysphonia," Ann. Otol. Rhinol. Laryngol. **116**(9), 695–701.

Eadie, T. L., and Stepp, C. E. (**2013**). "Acoustic correlate of vocal effort in spasmodic dysphonia," Ann. Otol. Rhinol. Laryngol. **122**(3), 169–176.

Espinoza, V. M., Zanartu, M., Van Stan, J. H., Mehta, D. D., and Hillman, R. E. (**2017**). "Glottal aerodynamic measures in women with phonotraumatic and nonphonotraumatic vocal hyperfunction," J. Speech Lang. Hear. Res. **60**(8), 2159–2169.

Friedman, A. D., Hillman, R. E., Landau-Zemer, T., Burns, J. A., and Zeitels, S. M. (**2013**). "Voice outcomes for photoangiolytic KTP laser treatment of early glottic cancer," Ann. Otol. Rhinol. Laryngol. **122**(3), 151–158.

Fujiki, R. B., Chapleau, A., Sundarrajan, A., McKenna, V., and Sivasankar, M. P. (**2017**). "The interaction of surface hydration and vocal loading on voice measures," J. Voice **31**(2), 211–217.
Gerratt, B. R., Kreiman, J., Antonanzas-Barroso, N., and Berke, G. S. (**1993**). "Comparing internal and external standards in voice quality judgments," J. Speech Hear. Res. **36**(1), 14–20.
Ghassemi, M., Van Stan, J. H., Mehta, D. D., Zanartu, M., Cheyne, H. A., Hillman, R. E., and Guttag, J. V. (**2014**). "Learning to detect vocal hyperfunction from ambulatory neck-surface acceleration features: Initial results for vocal fold nodules," IEEE Trans. Biomed. Eng. **61**(6), 1668–1675.
Granqvist, S. (**2003**). "The visual sort and rate method for perceptual evaluation in listening tests," Logoped. Phoniatr. Vocol. **28**(3), 109–116.
Hair, J. F., Anderson, R. E., Tatham, R. L., and Black, W. C. (**1995**). *Multivariate Data Analysis*, 3rd ed. (Macmillan, New York).
Heller Murray, E. S., Hands, G. L., Calabrese, C. R., and Stepp, C. E. (**2016**). "Effects of adventitious acute vocal trauma: Relative fundamental frequency and listener perception," J. Voice **30**(2), 177–185.
Heller Murray, E. S., Lien, Y. S., Van Stan, J. H., Mehta, D. D., Hillman, R. E., Pieter Noordzij, J., and Stepp, C. E. (**2017**). "Relative fundamental frequency distinguishes between phonotraumatic and non-phonotraumatic vocal hyperfunction," J. Speech Lang. Hear. Res. **60**(6), 1507–1515.
Heman-Ackah, Y. D., Sataloff, R. T., Laureyns, G., Lurie, D., Michael, D. D., Heuer, R., Rubin, A., Eller, R., Chandran, S., Abaza, M., Lyons, K., Divi, V., Lott, J., Johnson, J., and Hillenbrand, J. (**2014**). "Quantifying the cepstral peak prominence, a measure of dysphonia," J. Voice **28**(6), 783–788.
Hillenbrand, J., Cleveland, R., and Erickson, R. (**1994**). "Acoustic correlates of breathy vocal quality," J. Speech Lang. Hear. Res. **37**, 769–778.
Hillenbrand, J., and Houde, R. A. (**1996**). "Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech," J. Speech Hear. Res. **39**(2), 311–321.
Hillman, R. E., Holmberg, E. B., Perkell, J. S., Walsh, M., and Vaughan, C. (**1989**). "Objective assessment of vocal hyperfunction: An experimental framework and initial results," J. Speech Lang. Hear. Res. **32**(2), 373–392.
Hirano, M. (**1981**). *Clinical Examination of Voice* (Springer-Verlag, New York).
Hogikyan, N. D., and Sethuraman, G. (**1999**). "Validation of an instrument to measure voice-related quality of life (V-RQOL)," J. Voice **13**(4), 557–569.
Holmberg, E. B., Doyle, P., Perkell, J. S., Hammarberg, B., and Hillman, R. E. (**2003**). "Aerodynamic and acoustic voice measurements of patients with vocal nodules: Variation in baseline and changes across voice therapy," J. Voice **17**(3), 269–282.
Hunter, E. J., and Titze, I. R. (**2009**). "Quantifying vocal fatigue recovery: Dynamic vocal recovery trajectories after a vocal loading exercise," Ann. Otol. Rhinol. Laryngol. **118**(6), 449–460.
Isetti, D., Xuereb, L., and Eadie, T. L. (**2014**). "Inferring speaker attributes in adductor spasmodic dysphonia: Ratings from unfamiliar listeners," Am. J. Speech Lang. Pathol. **23**(2), 134–145.
ISO (**2002**). ISO 9921:2002(E), *Ergonomics-Assessment of Speech Communication* (ISO, Geneva, Switzerland).
Jacobson, B. H., Johnson, A., Grywalski, C., Silbergleit, A., Jocobson, G., Benninger, M. S., and Newman, C. W. (**1997**). "The voice handicap index (VHI): Development and validation," Am. J. Speech Lang. Pathol. **6**(3), 66–70.
Johnson, J. (**2012**). "A comparison between self-rated and listener-rated outcomes in tracheoesophageal speech," M.S. thesis, University of Washington, Seattle, WA.
Kempster, G. B., Gerratt, B. R., Abbott, K. V., Barkmeier-Kraemer, J., and Hillman, R. E. (**2009**). "Consensus auditory-perceptual evaluation of voice: Development of a standardized clinical protocol," Am. J. Speech Lang. Pathol. **18**(2), 124–132.
Klatt, D. H., and Klatt, L. C. (**1990**). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," J. Acoust. Soc. Am. **87**(2), 820–857.
Kleber, B., Zeitouni, A. G., Friberg, A., and Zatorre, R. J. (**2013**). "Experience-dependent modulation of feedback integration during singing: Role of the right anterior insula," J. Neurosci. **33**(14), 6070–6080.
Koo, T. K., and Li, M. Y. (**2016**). "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," J. Chiropr. Med. **15**(2), 155–163.
Kwan, L. C., and Whitehill, T. L. (**2011**). "Perception of speech by individuals with Parkinson's disease: A review," Parkinson's Disease **2011**, 389767.
Lametti, D. R., Nasir, S. M., and Ostry, D. J. (**2012**). "Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback," J. Neurosci. **32**(27), 9351–9358.
Lane, H. L., Catania, A. C., and Stevens, S. S. (**1961**). "Voice level: Autophonic scale, perceived loudness, and effects of sidetone," J. Acoust. Soc. Am. **33**(2), 160–167.
Larson, C. R., Sun, J., and Hain, T. C. (**2007**). "Effects of simultaneous perturbations of voice pitch and loudness feedback on voice F0 and amplitude control," J. Acoust. Soc. Am. **121**(5), 2862–2872.
Laukkanen, A. M., Ilomaki, I., Leppanen, K., and Vilkman, E. (**2008**). "Acoustic measures and self-reports of vocal fatigue by female teachers," J. Voice **22**(3), 283–289.
Lee, M., Drinnan, M., and Carding, P. (**2005**). "The reliability and validity of patient self-rating of their own voice quality," Clin. Otolaryngol. **30**(4), 357–361.
Lien, Y. A., Michener, C. M., Eadie, T. L., and Stepp, C. E. (**2015**). "Individual monitoring of vocal effort with relative fundamental frequency: Relationships with aerodynamics and listener perception," J. Speech Lang. Hear. Res. **58**(3), 566–575.
Lien, Y. A., and Stepp, C. E. (**2014**). "Comparison of voice relative fundamental frequency estimates derived from an accelerometer signal and low-pass filtered and unprocessed microphone signals," J. Acoust. Soc. Am. **135**(5), 2977–2985.
Lien, Y. S., Heller Murray, E. S., Calabrese, C. R., Michener, C. M., Van Stan, J. H., Mehta, D. D., Hillman, R. E., Noordzil, J. P., and Stepp, C. E. (**2017**). "Validation of an algorithm for semi-automated estimation of voice relative fundamental frequency," Ann. Otol. Rhinol. Laryngol. **126**(10), 712–716.
Lofqvist, A., Baer, T., McGarr, N. S., and Story, R. S. (**1989**). "The cricothyroid muscle in voicing control," J. Acoust. Soc. Am. **85**(3), 1314–1321.
Loucks, T. M. J., Poletto, C. J., Saxon, K. G., and Ludlow, C. L. (**2005**). "Laryngeal muscle responses to mechanical displacement of the thyroid cartilage in humans," J. Appl. Physiol. **99**(3), 922–930.
Lowell, S. Y., Colton, R. H., Kelley, R. T., and Mizia, S. A. (**2013**). "Predictive value and discriminant capacity of cepstral- and spectral-based measures during continuous speech," J. Voice **27**(4), 393–400.
Lowell, S. Y., Kelley, R. T., Awan, S. N., Colton, R. H., and Chan, N. H. (**2012**). "Spectral- and cepstral-based acoustic features of dysphonic, strained voice quality," Ann. Otol. Rhinol. Laryngol. **121**(8), 539–548.
McCabe, D. J., and Titze, I. R. (**2002**). "Chant therapy for treating vocal fatigue among public school teachers: A preliminary study," Am. J. Speech Lang. Pathol. **11**(4), 356–369.
McKenna, V. S., Murray, E. S. H., Lien, Y. A. S., and Stepp, C. E. (**2016**). "The relationship between relative fundamental frequency and a kinematic estimate of laryngeal stiffness in healthy adults," J. Speech Lang. Hear. Res. **59**(6), 1283–1294.
Merrill, R. M., Roy, N., and Lowe, J. (**2013**). "Voice-related symptoms and their effects on quality of life," Ann. Otol. Rhinol. Laryngol. **122**(6), 404–411.
Murphy, P. J., McGuigan, K. G., Walsh, M., and Colreavy, M. (**2008**). "Investigation of a glottal related harmonics-to-noise ratio and spectral tilt as indicators of glottal noise in synthesized and human voice signals," J. Acoust. Soc. Am. **123**(3), 1642–1652.
Neely, G., Ljunggren, G., Sylven, C., and Borg, G. (**1992**). "Comparison between the Visual Analogue Scale (VAS) and the Category Ratio Scale (CR-10) for the evaluation of leg exertion," Int. J. Sports Med. **13**(2), 133–136.
Netsell, R., Lotz, W., and Shaughnessy, A. L. (**1984**). "Laryngeal aerodynamics associated with selected voice disorders," Am. J. Otolaryngol. **5**(6), 397–403.
Nikjeh, D., Lister, J., and Frisch, S. (**2009**). "The relationship between pitch discrimination and vocal production: Comparison of vocal and instrumental musicians," J. Acoust. Soc. Am. **124**(1), 328–338.
Noll, A. M. (**1964**). "Short-term spectrum and 'cepstrum' techniques for vocal pitch detection," J. Acoust. Soc. Am. **36**, 296–302.
Noll, A. M. (**1967**). "Cepstrum pitch determination," J. Acoust. Soc. Am. **41**, 293–309.
Oates, J. (**2009**). "Auditory-perceptual evaluation of disordered voice quality pros, cons and future directions," Folia Phon. Logopaed. **61**(1), 49–56.
Patel, R. R., Awan, A. N., Barkmeier-Kraemer, J., Courey, M., Deliyski, D., Eadie, T., Paul, D., Svec, J. G., and Hillman, R. (**2018**). "Recommended protocols for instrumental assessment of voice: American speech-language-hearing association expert panel to develop a protocol for instrumental assessment of vocal function," Am. J. Speech-Lang. Pathol. **27**, 887–905.

Qi, Y. Y., and Hillman, R. E. (**1997**). "Temporal and spectral estimations of harmonics-to-noise ratio in human voice signals," J. Acoust. Soc. Am. **102**(1), 537–543.
Rantala, L., Lindholm, P., and Vilkman, E. (**1998**). "F0 change due to voice loading under laboratory and field conditions. A pilot study," Logoped. Phon. Vocol. **23**(4), 164–168.
Rosenthal, A. L., Lowell, S. Y., and Colton, R. H. (**2014**). "Aerodynamic and acoustic features of vocal effort," J. Voice **28**(2), 144–153.
Roy, N., Fetrow, R. A., Merrill, R. M., and Dromey, C. (**2016**). "Exploring the clinical utility of relative fundamental frequency as an objective measure of vocal hyperfunction," J. Speech Lang. Hear. Res. **59**(5), 1002–1017.
Roy, N., Merrill, R. M., Gray, S. D., and Smith, E. M. (**2005**). "Voice disorders in the general population: Prevalence, risk factors, and occupational impact," Laryngoscope **115**(11), 1988–1995.
Sapir, S., Baker, K. K., Larson, C. R., and 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," J. Speech Lang. Hear. Res. **43**(1), 268–276.
Schindler, A., Mozzanica, F., Vedrody, M., Maruzzi, P., and Ottaviani, F. (**2009**). "Correlation between the Voice Handicap Index and voice measurements in four groups of patients with dysphonia," Otolaryngol. Head Neck Surg. **141**(6), 762–769.
Schlow, R. L. (**1991**). "Considerations in selecting and validating an adults/elderly hearing screening protocol," Ear Hear. **12**(5), 337–348.
Schmidt, C. P., Gelfer, M. P., and Andrews, M. L. (**1990**). "Intensity range as a function of task and training," J. Voice **4**(1), 30–36.
Selby, J. C., Gilbert, H. R., and Lerman, J. W. (**2003**). "Perceptual and acoustic evaluation of individuals with laryngopharyngeal reflux pre- and post-treatment," J. Voice **17**(4), 557–570.
Severin, F., Bozkurt, B., and Dutoit, T. (**2005**). "HNR extraction in voiced speech, oriented towards voice quality analysis," in *Proceedings of the 13th European Signal Processing Conference*, September 4–8, Antalya, Turkey.
Shipp, T. (**1975**). "Vertical laryngeal position during continuous and discrete vocal frequency change," J. Speech Hear. Res. **18**(4), 707–718.
Smith, E., Gray, S. D., Dove, H., Kirchner, L., and Heras, H. (**1997**). "Frequency and effects of teachers' voice problems," J. Voice **11**(1), 81–87.
Smith, E., Taylor, M., Mendoza, M., Barkmeier, J., Lemke, J., and Hoffman, H. (**1998**). "Spasmodic dysphonia and vocal fold paralysis: Outcomes of voice problems on work-related functioning," J. Voice **12**(2), 223–232.
Somodi, L. B., Robin, D. A., and Luschei, E. S. (**1995**). "A model of 'sense of effort' during maximal and submaximal contractions of the tongue," Brain Lang. **51**(3), 371–382.
Stemple, J. C., Stanley, J., and Lee, L. (**1995**). "Objective measures of voice production in normal subjects following prolonged voice use," J. Voice **9**(2), 127–133.
Stepp, C. E., Merchant, G. R., Heaton, J. T., and Hillman, R. E. (**2011**). "Effects of voice therapy on relative fundamental frequency during voicing offset and onset in patients with vocal hyperfunction," J. Speech Lang. Hear. Res. **54**(5), 1260–1266.
Stepp, C. E., Sawin, D. E., and Eadie, T. L. (**2012**). "The relationship between perception of vocal effort and relative fundamental frequency during voicing offset and onset," J. Speech Lang. Hear. Res. **55**(6), 1887–1896.
Sundarrajan, A., Huber, J. E., and Sivasankar, M. P. (**2017**). "Respiratory and laryngeal changes with vocal loading in younger and older individuals," J. Speech Lang. Hear. Res. **60**(9), 2551–2556.
Sundberg, J., Iwarsson, J., and Billstrom, A. H. (**1995**). "Significance of mechanoreceptors in the subglottal mucosa for subglottal pressure control in singers," J. Voice **9**(1), 20–26.
Titze, I. R. (**1989**). "On the relation between subglottal pressure and fundamental frequency in phonation," J. Acoust. Soc. Am. **85**(2), 901–906.
Verdolini, K., Titze, I. R., and Fennell, A. (**1994**). "Dependence of phonatory effort on hydration level," J. Speech Hear. Res. **37**(5), 1001–1007.
Vilkman, E., Lauri, E. R., Alku, P., Sala, E., and Sihvo, M. (**1999**). "Effects of prolonged oral reading on F0, SPL, subglottal pressure and amplitude characteristics of glottal flow waveforms," J. Voice **13**(2), 303–312.
Vogel, A. P., Maruff, P., Snyder, P. J., and Mundt, J. C. (**2009**). "Standardization of pitch-range settings in voice acoustic analysis," Behav. Res. Methods **41**(2), 318–324.
Witte, R., and Witte, J. (**2010**). *Statistics* (Wiley, Hoben, NJ).
Xue, C., Kang, J., and Jiang, J. (**2018**). "Dynamically monitoring vocal fatigue and recovery using aerodynamic, acoustic, and subjective self-rating measurements," J. Voice (in press).
Zanartu, M., Galindo, G. E., Erath, B. D., Peterson, S. D., Wodicka, G. R., and Hillman, R. E. (**2014**). "Modeling the effects of a posterior glottal opening on vocal fold dynamics with implications for vocal hyperfunction," J. Acoust. Soc. Am. **136**(6), 3262–3271.