ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

# Clinical Cutoff Scores for Acoustic Indices of Vocal Hyperfunction That Combine Relative Fundamental Frequency and Cepstral Peak Prominence

Mara R. Kapsner-Smith,[a] Manuel E. Díaz-Cádiz,[b] Jennifer M. Vojtech,[b,c] Daniel P. Buckley,[b,d] Daryush D. Mehta,[e,f,g] Robert E. Hillman,[e,f,g] Lauren F. Tracy,[b,d] J. Pieter Noordzij,[b,d] Tanya L. Eadie,[a] and Cara E. Stepp[b,c,d]

[a] Department of Speech & Hearing Sciences, University of Washington, Seattle [b] Department of Speech, Language & Hearing Sciences, Boston University, MA [c] Department of Biomedical Engineering, Boston University, MA [d] Department of Otolaryngology—Head & Neck Surgery, Boston University School of Medicine, MA [e] MGH Institute of Health Professions, Boston, MA [f] Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Boston [g] Department of Surgery, Harvard Medical School, Cambridge, MA

ARTICLE INFO

Article History:
Received August 30, 2021
Revision received November 15, 2021
Accepted December 15, 2021

Editor-in-Chief: Bharath Chandrasekaran
Editor: Jack J. Jiang

https://doi.org/10.1044/2021_JSLHR-21-00466

ABSTRACT

**Purpose:** This study examined the discriminative ability of acoustic indices of vocal hyperfunction combining smoothed cepstral peak prominence (CPPS) and relative fundamental frequency (RFF).
**Method:** Demographic, CPPS, and RFF parameters were entered into logistic regression models trained on two 1:1 case–control groups: individuals with and without nonphonotraumatic vocal hyperfunction (NPVH; $n$ = 360) and phonotraumatic vocal hyperfunction (PVH; $n$ = 240). Equations from the final models were used to predict group membership in two independent test sets ($n$ = 100 each).
**Results:** Both CPPS and RFF parameters significantly improved model fits for NPVH and PVH after accounting for demographics. CPPS explained unique variance beyond RFF in both models. RFF explained unique variance beyond CPPS in the PVH model. Final models included CPPS and RFF offset parameters for both NPVH and PVH; RFF onset parameters were significant only in the PVH model. Area under the receiver operating characteristic curve analysis for the independent test sets revealed acceptable classification for NPVH (72%) and good classification for PVH (86%).
**Conclusions:** A combination of CPPS and RFF parameters showed better discriminative ability than either measure alone for PVH. Clinical cutoff scores for acoustic indices of vocal hyperfunction are proposed for assessment and screening purposes.

Voice disorders affect as many as 30% of individuals at some point in their lives (Cohen, 2010) and may significantly impact an individual's ability to participate in social and work settings (Ma & Yiu, 2001). Vocal hyperfunction (VH), a physiological response to increased vocal demands or challenging vocal situations, may cause transient disruption of vocal function in speakers with typical voices (Hunter & Titze, 2009; Whitling et al., 2015;

Correspondence to Mara R. Kapsner-Smith: mkapsner@uw.edu. ***Disclosure:*** *Jennifer M. Vojtech is currently employed by Delsys, Inc., a commercial company that manufactures and markets sensor and software technologies for human movement, and Altec, Inc., an R&D company that performs research to design new technologies for human movement. Robert E. Hillman and Daryush Mehta have financial interests in InnoVoyce LLC, a company focused on developing and commercializing technologies for the prevention, diagnosis, and treatment of voice-related disorders. Their interests were reviewed and are managed by Massachusetts General Hospital and Mass General Brigham in accordance with their conflict of interest policies. Cara Stepp has received consulting fees from Altec, Inc./Delsys, Inc., companies focused on developing and commercializing technologies related to human movement. C. Stepp's interests were reviewed and are managed by Boston University in accordance with their conflict of interest policies. The other authors have declared that no other competing financial or nonfinancial interests existed at the time of publication.*

Xue et al., 2019). For some speakers, VH may become chronic, causing increased and poorly regulated muscle tension and varying degrees of vocal limitations, and is considered an etiological factor in the most frequently occurring types of voice disorders (Hillman et al., 2020; Oates & Winkworth, 2008). This atypical laryngeal muscle tension may manifest with altered aerodynamic and vibratory function consistent with increased vocal fold collision forces leading to phonotrauma (Espinoza et al., 2017; Hillman et al., 1990; Van Stan, Mehta, Ortiz, Burns, Marks, et al., 2020). Alternatively, VH may produce altered vocal fold biomechanics consistent with decreased collision forces and often with incomplete glottal closure that do not cause phonotrauma but reduce vocal efficiency and lead to fatigue and discomfort from voice use (Espinoza et al., 2017; Van Stan, Ortiz, Cortes, et al., 2021). A recently updated theoretical framework of VH (Hillman et al., 2020) designates these two fundamentally different types of VH as phonotraumatic VH (PVH), considered to be an etiological factor in benign lesions of the lamina propria, and nonphonotraumatic VH (NPVH), considered to be an etiological factor in primary muscle tension dysphonia (pMTD; Morrison & Rammage, 1993; Naunheim & Carroll, 2017; Van Houtte et al., 2011; Verdolini et al., 2014).

Current gold standard clinical assessments of increased laryngeal muscle tension, thought to be a core feature of both NPVH and PVH, rely on visual–perceptual judgments via laryngostroboscopy, auditory-perceptual evaluation of strain, manual palpation of the perilaryngeal region, and patient self-report (Kempster et al., 2009; Khoddami et al., 2015; Poburka et al., 2017; Ruel & Thibeault, 2020). Although these tools provide some information about vocal function and health, they are subject to rater error and bias due to their subjective nature (Kreiman & Gerratt, 1998; Poburka et al., 2017; Stepp, Heaton, et al., 2011; Yiu et al., 2014). Current clinical assessments lack objective, standardized measures of VH, which could aid in differential diagnosis and in documenting when vocal function has normalized. Although recent studies of daily voice use in patients with VH using ambulatory monitoring demonstrate that objective measures can differentiate speakers with NPVH, speakers with PVH, and speakers with typical voices and are sensitive to changes after treatment (Van Stan, Mehta, Ortiz, Burns, Marks, et al., 2020; Van Stan, Mehta, Ortiz, Burns, Toles, et al., 2020; Van Stan, Ortiz, Cortes, et al., 2021; Van Stan, Ortiz, Marks, et al., 2021), these measures require long-term monitoring and use of equipment outside the clinic. Development of objective measures of VH that can be administered in the clinic is the goal of this study.

Acoustic measures may provide objective, reliable, and relatively low-cost tools to assess aspects of the voice signal that convey information about underlying vocal function. Relative fundamental frequency (RFF) is an acoustic measure that is hypothesized to be sensitive to the changes in vocal fold tension, abduction, and transglottal pressure that take place during voicing offsets and onsets and occur during the production of a voiceless consonant between two vowels (vowel–voiceless consonant–vowel [VCV]; Fukui & Hirose, 1983; Halle & Stevens, 1971; Ladefoged, 1967; Stevens, 1977). The fundamental frequencies of the last 10 cycles of phonation during voicing offset and the first 10 cycles during voicing onset are measured and normalized in semitones relative to steady-state phonation. During voicing offset, fundamental frequency ($f_o$) would be expected to decrease due to decreased transglottal pressure as the vocal folds abduct (Ladefoged, 1967; Titze, 1989). Furthermore, results of a recent computational modeling study suggest that the decrease in vocal fold collision forces during the abductory gesture lead to decreases in $f_o$ (Serry et al., 2021). However, in speakers with typical voices, an increase in longitudinal vocal fold tension is hypothesized to counteract the expected drop in $f_o$ during voicing offset; this tension is thought to carry over and increase $f_o$ during voicing onset when elevated transglottal pressure also increases $f_o$ (Halle & Stevens, 1971; Jaiswal, 2011; Ladefoged, 1967; Löfqvist et al., 1989; Stevens, 1977). Because of elevated baseline longitudinal vocal fold tension (Hillman et al., 1989), speakers with VH are hypothesized to be less able to increase tension to counteract the effects of abduction and decreased transglottal pressure during voicing offsets, resulting in lower offset RFF values relative to speakers with typical voices (Stepp et al., 2010). Increased transverse tension and associated higher collision forces during steady-state phonation in speakers with PVH may cause a larger drop in $f_o$, further lowering offset RFF (Heller Murray et al., 2017; Kunduk et al., 2006; Serry et al., 2021; Watson, 1998). Onset RFF values, which are thought to be positive in speakers with typical voices due to increased longitudinal tension and increased transglottal pressure (Halle & Stevens, 1971; Löfqvist et al., 1989; Stevens, 1977), are similarly expected to be lowered in speakers with VH due to the inability to further increase tension. Transverse tension may also decrease onset RFF values because increased closing velocities could limit the time over which increased transglottal pressure is active (see Heller Murray et al., 2017, for discussion). Thus, because of the biomechanical differences between NPVH and PVH described earlier, including increased transverse tension and associated collision forces in PVH, RFF may be differentially sensitive to these two types of VH.

Prior studies support the sensitivity of RFF to VH. RFF explained a large portion of the variance in kinematic measures of laryngeal stiffness in healthy adults, supporting the hypothetical relationship between RFF and laryngeal tension (McKenna et al., 2016). RFF offset and onset values have been shown to be significantly lower in speakers with hyperfunctional voice disorders relative to

control speakers (Stepp et al., 2010). Furthermore, in a study of speakers with hyperfunctional voice disorders ($n$ = 13 with pMTD, $n$ = 3 with vocal fold nodules) before and after successful behavioral voice therapy, RFF significantly increased after treatment, becoming more similar to values observed in speakers with typical voices (Stepp, Merchant, et al., 2011). However, in a study comparing RFF measures obtained at pre- and postsurgical visits from speakers undergoing phonosurgery ($n$ = 18 with vocal fold nodules or polyps) with no behavioral intervention during the measurement period, RFF was not significantly different after surgery (Stepp et al., 2010). The authors interpreted this finding as suggesting that decreased RFF reflects functional (i.e., VH) rather than structural (i.e., phonotrauma) impairment, which, in the case of the postsurgical patients, may represent habituated compensatory hyperfunction that persists even in the absence of the lesion(s). In addition, RFF was more sensitive than listener ratings of overall severity (OS) and of vocal effort (i.e., strain) to the presence of a hyperfunctional voice disorder in speakers with very mild or no voice quality impairment (Stepp et al., 2012), suggesting that RFF may provide complementary information to measures of dysphonia severity.

In addition to being sensitive to differences in laryngeal tension during voice production, RFF has been shown to have diagnostic potential in detecting the presence of hyperfunctional voice disorders. Heller Murray et al. (2017) examined the ability of RFF values to discriminate between speakers with NPVH, speakers with PVH, and control speakers. Consistent with theoretical predictions, they found RFF Offset Cycle 10 to be significantly lower in speakers with NPVH than in control speakers. Furthermore, speakers with PVH had significantly lower RFF Offset Cycle 8–10 values than either control speakers or speakers with NPVH. Onset Cycle 1 was significantly lower in speakers with PVH than in control speakers; there were no other significant differences between groups related to onset RFF. Differences between speakers with NPVH and speakers with PVH may be explained by the presence of increased transverse tension in speakers with PVH, hypothesized to lower both offset and onset RFF values. The discriminative ability of RFF Offset Cycle 10 and Onset Cycle 1 was examined using receiver operating characteristic (ROC) curves (Heller Murray et al., 2017). Although RFF Offset Cycle 10 discriminated speakers with PVH from controls with typical voices with good accuracy (area under the ROC curve [AUC] = .80), the discriminative ability of RFF for NPVH was more limited. One limitation of this study was its relatively small sample size, with 49 control speakers, 54 speakers with PVH, and 35 speakers with NPVH. In a larger group of speakers, Roy et al. (2016) found a significant main effect of NPVH status on RFF onset, such that speakers with pMTD ($n$ = 111) had shallower slopes than controls ($n$ = 20), whereas there were no significant differences for RFF offset. A significant limitation of this study was the large number of unanalyzable samples, substantially higher than in some recent studies of RFF in speakers with voice disorders (Heller Murray et al., 2017; Vojtech et al., 2019). This may be due in part to methodological differences, in particular, the phonetic context of the analyzed tokens, which included unstressed vowels (Lien et al., 2014), limiting comparisons between this study and others.

In addition to elevated laryngeal tension, some, but not all, speakers with VH present with complaints of altered voice quality or dysphonia (Ruel & Thibeault, 2020). Dysphonia may occur due to dysregulated subglottal pressure and vibratory function secondary to atypical vocal fold posturing and/or structural changes due to phonotrauma (Espinoza et al., 2017; Hillman et al., 1990, 2020). Acoustic measures have shown potential for sensitivity to OS of dysphonia (Maryn et al., 2009), as well as specific voice quality parameters such as strain (e.g., Anand et al., 2019), roughness, and breathiness (von Latoszek et al., 2018). In recent decades, smoothed cepstral peak prominence (CPPS; Hillenbrand et al., 1994; Hillenbrand & Houde, 1996) has been established by numerous studies as a valid and reliable acoustic correlate of dysphonia (Awan & Roy, 2009; Awan et al., 2009; Heman-Ackah et al., 2002; Maryn, Corthals, et al., 2010; Maryn, De Bodt, & Roy, 2010; Wolfe & Martin, 1997; Wolfe et al., 2000) and was recommended in 2018 by an American Speech-Language-Hearing Association expert panel as part of a standard instrumental protocol for clinical voice assessment (Patel et al., 2018). A cepstrum is a Fourier transform of the power spectrum of a voice signal and measures the cepstral magnitude of "quefrencies" and corresponding "rahmonics." The dominant rahmonic is called the "cepstral peak," and voices that are more periodic—and therefore have more well-defined harmonic structures—display higher cepstral peaks. CPPS has been found to correlate with auditory-perceptual judgments of specific voice quality parameters including breathiness (Hillenbrand & Houde, 1996), roughness (Awan & Roy, 2005), and strain (Lowell et al., 2012), although it appears to have the strongest relationship with global judgments of OS of voice quality (Maryn et al., 2009).

Although CPPS alone has been shown to have acceptable accuracy rates in detecting the presence of a voice disorder (Murton et al., 2020; Sauder et al., 2017), multiparameter acoustic indices have been developed to improve sensitivity and specificity (Awan & Roy, 2009; Awan et al., 2009, 2016; Maryn, Corthals, et al., 2010; Maryn, De Bodt, & Roy, 2010; Peterson et al., 2013). Similarly, a multiparameter approach to the acoustic measurement of VH may prove superior to the use of RFF

alone. To our knowledge, CPPS has not specifically been examined for discriminative ability in VH. However, it was shown to be significantly lower in speakers with vocal fold nodules (a subgroup of speakers with PVH; $n = 50$) than in control speakers ($n = 50$; Radish Kumar et al., 2010). Furthermore, it predicted 48% of the variance in the Vocal Fatigue Index (Nanjundeswaran et al., 2015) scores in speakers with ($n = 50$) and without ($n = 50$) VH (Mahalingam et al., 2020). Given the predominant clinical signs of excessive laryngeal muscle tension and dysphonia in patients with hyperfunctional voice disorders, the acoustic measures RFF and CPPS may provide complementary information and enhanced accuracy in detection of VH.

The purpose of this study was to assess the discriminative ability of a combination of acoustic measures of laryngeal tension and dysphonia to detect hyperfunctional voice disorders in a large sample of speakers with NPVH, speakers with PVH, and control speakers without voice disorders. Specifically, this study builds upon prior work examining the relationship of RFF to VH (Heller Murray et al., 2017; Roy et al., 2016; Stepp et al., 2010; Stepp, Merchant, et al., 2011) by using a multidimensional acoustic approach to detecting NPVH and PVH, combining multiple parameters of RFF and CPPS in multivariate models. These measures are thought to be sensitive to different core clinical features of hyperfunctional voice disorders, including increased laryngeal tension (RFF) and dysphonia (CPPS), and therefore should provide better diagnostic accuracy in these populations than either measure alone. The study questions were as follows: (a) In a large data set, do RFF parameters and CPPS parameters distinguish between individuals with NPVH and control speakers, and individuals with PVH and control speakers, and do these categories of acoustic measures provide complementary information about the likelihood of VH? (b) What is the most parsimonious combination of RFF and CPPS parameters that best distinguishes NPVH from typical voice and PVH from typical voice, and with what degree of sensitivity and specificity? (c) What clinical cutoff scores for composite RFF/CPPS indices would be most appropriate to detect the likely presence of NPVH and PVH? We hypothesized that a combination of RFF and CPPS parameters would predict voice disorder status for both NPVH and PVH, with each type of acoustic measure (RFF and CPPS) providing unique information. We also hypothesized that the combination of RFF parameters that best predicted the presence of NPVH versus PVH would differ due to differences in underlying laryngeal function, with onset RFF serving as a significant predictor in PVH due to elevated transverse tension but not in NPVH. The resulting multiparameter acoustic indices may provide clinically meaningful objective indicators of NPVH and PVH to assist in diagnosis and track progress in response to intervention.

# Method

## Participant Characteristics

Speech recordings were selected from a database of extant data from prior studies conducted in the Stepp Lab for Sensorimotor Rehabilitation Engineering at Boston University and the Center for Laryngeal Surgery and Voice Rehabilitation at Massachusetts General Hospital (MGH). Speakers were categorized into three groups: speakers with no history of a voice disorder (control speakers), speakers with NPVH, and speakers with PVH. All speakers were at least 18 years of age and had a minimum of six VCV stimuli available in the database. Eligibility criteria for control speakers included no self-reported history of a voice disorder and voice quality rated as within normal limits by an experienced voice clinician (see below for rating procedures). Eligibility criteria for speakers with VH included a diagnosis consistent with NPVH or PVH confirmed by a referring laryngologist. Speakers with neurogenic diagnoses (e.g., laryngeal dystonia, vocal fold paresis/paralysis) or structural pathology not consistent with PVH (e.g., papilloma, carcinoma, and trauma not related to phonation) were excluded from the study. Speakers with VH were categorized into NPVH or PVH by the authors (M.R.K.S., J.M.V., J.P.N.) based on the referring diagnosis. These criteria resulted in the identification of 312 potential control speakers (203 women, 109 men, $M_{age} = 37$ years, $SD = 22$, range: 18–100), 252 speakers with NPVH (172 women, 80 men, $M_{age} = 41$ years, $SD = 18$, range: 18–88), and 182 speakers with PVH (146 women, 36 men, $M_{age} = 33$ years, $SD = 15$, range: 18–77). Speakers in the NPVH group were diagnosed with pMTD. Speakers in the PVH group were diagnosed with benign vocal fold lesions or other diagnosis consistent with a phonotraumatic etiology, including vocal fold nodules, polyps, cysts, or scar. There was no inclusion criterion for lesion size. Information about voice diagnoses for the speakers in the PVH group is included in Table 1. Participants either completed written consent in compliance with either the Boston University Institutional Review Board or the Mass General Brigham Institutional Review Board, or approval for use of clinical data was obtained from the Mass General Brigham Institutional Review Board for retrospective analysis of acoustic data.

## Audio Recording Procedures and Speech Stimuli

All signals were acquired digitally and recorded for off-line analysis. Participants were recorded at one of two locations: (a) in a sound-treated room at Boston University using a head-mounted microphone (model WH20XLR; Shure) sampled at 44.1 kHz with 16-bit resolution or (b) in

**Table 1.** Diagnoses in the phonotraumatic vocal hyperfunction group ($N$ = 182).

| Diagnosis | No. of participants |
|---|---|
| Vocal fold nodules (bilateral) | 114 |
| Vocal fold polyp (unilateral or bilateral) | 60 |
| Vocal fold fibrovascular lesion (unilateral) | 3 |
| Vocal fold cyst (unilateral) and vocal fold fibrovascular lesion (unilateral) | 2 |
| Vocal fold polyp (unilateral) and vocal fold fibrovascular lesion (unilateral) | 1 |
| Vocal fold polyp (unilateral) and vocal fold sulcus (unilateral) | 1 |
| Vocal fold scar | 1 |

a sound-treated room at MGH using a Sony ECM-44B microphone sampled at 50 kHz with 16-bit resolution. During each recording, an examiner first modeled the target utterances before the participant repeated them. Utterances consisted of three sets of three VCV instances, with the voiceless consonant /f/ and the corner vowels /ɑ/, /i/, and /u/ (i.e., /ɑfɑ ɑfɑ ɑfɑ/, /ifi ifi ifi/, and /ufu ufu ufu/), for a total of nine VCVs. VCVs were modeled with equal stress on each syllable. These stimuli were chosen as they were shown to yield low intraspeaker variability in RFF measures in previous studies (Lien et al., 2014; Park & Stepp, 2019). Due to low-frequency noise present in some of the samples recorded at MGH, a high-pass filter was applied to the affected samples, with a cutoff frequency of 70 Hz.

## Auditory-Perceptual Ratings of Voice Quality

Binary voice quality ratings (a response of "yes" or "no" to the prompt, "Within normal limits?") were collected to confirm inclusion in the control speaker group. Stimuli from these speakers (/ifi ifi ifi/) were presented in random order using a custom MATLAB interface. Fifteen percent of samples were randomly repeated for intrarater reliability and agreement analyses. All samples were rated by a voice-specialized clinician with 4 years of clinical experience assessing and treating patients with voice disorders (D.B.P.). A rating of "yes" was used as an inclusion criterion for control speakers, resulting in a final set of 312 speakers out of 319 potential control speakers. Fifteen percent of samples were rated by a second voice-specialized clinician with 12 years of experience (M.R.K.S.) in order to assess interrater agreement. Agreement was assessed by calculating the percentage of agreement in repeated ratings of the same stimulus (probability of chance agreement = 50%), within (D.B.P.) and across (D.B.P. and M.R.K.S.) raters. The percentage of intrarater agreement was 98.6%. The percentage of interrater agreement was 95.2%.

Auditory-perceptual ratings of OS of voice quality were collected for speakers in the NPVH and PVH groups for descriptive purposes. The stimuli (/ifi ifi ifi/) were presented in random order using a custom MATLAB interface. Fifteen percent of samples were randomly repeated for intrarater reliability analysis. All samples were rated by D.B.P. and 15% were also rated by M.R.K.S. to assess interrater reliability. Reliability was assessed by calculating Pearson's correlation coefficients ($r$). Intrarater reliability for D.B.P. was $r$ = .92. Interrater reliability was $r$ = .68, within the range of interrater reliability values found in previous studies of highly trained listeners (Kreiman et al., 1993). The interrater reliability result may have been impacted by a relative lack of variance in the severity of our sample, which comprised many speakers with mild–moderate dysphonia. For speakers with NPVH, the mean rating of OS was 16.2 ($SD$ = 12.6, range: 0–77). For speakers with PVH, the mean rating of OS was 20.6 ($SD$ = 13.4, range: 0–94).

## Acoustic Analyses

RFF was calculated using a semi-automated MATLAB algorithm (MathWorks, 2016; Vojtech et al., 2019). After identification of voiceless consonant locations in each utterance, the algorithm determines the boundaries of voiceless consonants and vowels and calculates the $f_o$ of the 10 voicing cycles immediately preceding (voicing offset) and immediately following (voicing onset) the voiceless segment. The $f_o$ of each cycle is converted to semitones using the cycle farthest from the voicing boundary as reference, as these are closer to the center of the vowel and therefore most representative of steady-state phonation (i.e., offset cycles are converted to semitones relative to Offset Cycle 1, and onset cycles are converted to semitones relative to Onset Cycle 10). All available RFF instances for each speaker were averaged to obtain mean values for each offset and onset cycle. Across speakers, at least two RFF instances were available for 90.1% of speakers.

Average RFF values were used to derive six RFF parameters, after Buckley et al. (2020): *RFFoff10* and *RFFon1* (the mean RFF value for the boundary cycles at voicing offset and onset), *RFFoff10-9* and *RFFon1-2* (the difference between the two cycles closest to voicing offset and the two cycles closest to voicing onset to approximate the rate of change closest to voicing transitions), and *RFFoff10-5* and *RFFon1-6* (the difference between the midpoint of offset/onset and the respective boundary cycle). The boundary cycles *RFFoff10* and *RFFon1* provide information about the difference between steady-state phonation and the cycle closest to the voiceless period. However, the rate of change in RFF across cycles does not tend to be linear, and the other parameters provide

information about the rate of change where the slope tends to be steepest (e.g., Heller Murray et al., 2017; Stepp et al., 2010). Changes in slope across RFF cycles could relate to underlying biomechanical factors such as the loss of collision forces during voicing offset (Serry et al., 2021). In addition to these parameters, mean RFF values of all cycles were calculated for each group and visually inspected for patterns to identify additional potential parameters. For the NPVH group, group differences were apparent across all offset cycles and Onset Cycles 1 and 2. Similar patterns were observed for PVH, though with a larger magnitude of group differences, particularly for onset cycles. Therefore, two additional RFF parameters were calculated: the sum of all offset cycles excluding the reference cycle (i.e., 2 through 10, *RFFoffTotal*) and the sum of Onset Cycles 1 and 2 (*RFFon1+2*). RFF parameters were calculated for all speakers who had a minimum of two usable RFF instances, after McKenna and Stepp (2018). With a minimum of six recorded VCV samples per participant, conservatively at least two samples should be usable to calculate RFF for most speakers (Eadie & Stepp, 2013), and averaging across at least two RFF instances should provide a more robust estimate of RFF than using a single instance.

Two CPPS parameters were calculated from the concatenated vowel segments of the VCV samples. *CPPS mean* has been used in multiparameter acoustic indices of voice quality (Awan et al., 2010; Maryn, De Bodt, & Roy, 2010) and has been shown to discriminate between speakers with and without dysphonia (Sauder et al., 2017). In addition, *CPPS SD* significantly predicted listener ratings of dysphonia severity, with higher *CPPS SD* values associated with increased dysphonia severity in sustained phonation (Awan et al., 2010). Because the VCV samples used in this study were produced with equal stress, the concatenated vowel segments are more akin to sustained phonation than connected speech; thus, a similar relationship would be expected.

CPPS parameters were obtained by a combination of MATLAB (MathWorks, 2016) and Praat (Boersma & Weenink, 2013) scripts. First, a custom MATLAB script concatenated all audio samples per participant, selecting only vowel segments with a root-mean-square thresholding procedure. The threshold for voicing was 10% of the maximum signal value found in each recording. Then, these concatenated samples were split into smaller audio frames using an 87-ms window size with a window overlap of 79 ms. A Praat script was used to calculate single CPPS values for each audio frame. From these time series data, the parameters *CPPS mean* and *CPPS SD* were calculated. In order to reduce potential biases in the *CPPS mean* estimated by this procedure, window size and window overlap parameters were determined empirically in such a way that the error between the CPPS value obtained by the Praat script with the entire concatenated audio sample and the *CPPS mean* value obtained by the method described was minimized.

## Statistical Analysis

All statistical analyses were conducted using IBM SPSS Statistics software Version 26. Significance for all statistical tests was set a priori at $p < .05$.

Missing data were assessed prior to dividing speakers into training and test sets. Across all speakers, 9.9% of cases had missing data. All missing data were RFF values due to fewer than two usable onset or offset RFF instances. Little's Missing Completely at Random (MCAR) test was conducted to assess patterns of missing data. The result suggested data were not MCAR ($\chi^2$ = 132.378, $df$ = 15, $p < .001$). It is likely that RFF data were missing not at random (MNAR), that is, influenced by voice signal type, severity, or other systematic factors. Because missing data estimation methods under the MNAR assumption are complex and cannot with certainty account for potential bias, these methods were not used (Jakobsen et al., 2017). Instead, cases with missing data were deleted listwise (i.e., all data were removed for cases with missing values for any parameter) before preparing training and test sets. Implications for potential bias are included in the discussion.

Four sample sets were created with a 1:1 ratio of cases and controls: NPVH training, NPVH test, PVH training, and PVH test. Due to higher recruitment of university students and older speakers with typical voices from previous experiments in the Stepp Lab, there was an excess of control speakers in the age brackets of 18–28 and 58–68 years. Therefore, control speakers in these ranges were randomized, and a subset was removed such that the total number of remaining control speakers was equal to the number of remaining VH cases ($n$ = 230 for NPVH, $n$ = 170 for PVH). This resulted in a combined NPVH/control sample of $n$ = 460 ($n$ = 230 cases, $n$ = 230 controls) and a combined PVH/control sample of $n$ = 340 ($n$ = 170 cases, $n$ = 170 controls). A random subset of 100 speakers ($n$ = 50 cases, $n$ = 50 controls) was held out from each combined sample (NPVH/control, PVH/control) for use as independent test sets to assess the classification performance of the regression equations derived from the training sets in untrained sets of data. The final sample characteristics are presented in Table 2.

Multiple logistic regression was used to examine relationships between predictor parameters and voice disorder status in the NPVH and PVH training samples. The RFF parameters implemented in the models were the six parameters included in previous studies—namely, *RFFoff10*, *RFFon1*, *RFFoff10-9*, *RFFoff10-5*, *RFFon1-2*, and *RFFon1-6* (Buckley et al., 2020)—and two parameters derived from

**Table 2.** Final sample characteristics for training and independent test sets.

| Data set | *N* | Female | Male | Age *M* | Age range (years) | Age *SD* |
|---|---|---|---|---|---|---|
| NPVH training | | | | | | |
| Cases | 180 | 125 | 55 | 40 | 17–77 | 17 |
| Controls | 180 | 113 | 67 | 39 | 18–100 | 22 |
| PVH training | | | | | | |
| Cases | 120 | 95 | 25 | 32 | 18–77 | 14 |
| Controls | 120 | 79 | 41 | 38 | 18–90 | 22 |
| NPVH test | | | | | | |
| Cases | 50 | 36 | 14 | 42 | 19–83 | 19 |
| Controls | 50 | 36 | 14 | 37 | 18–83 | 22 |
| PVH test | | | | | | |
| Cases | 50 | 43 | 7 | 33 | 18–73 | 15 |
| Controls | 50 | 31 | 19 | 34 | 18–100 | 21 |

*Note.* NPVH = nonphonotraumatic vocal hyperfunction; PVH = phonotraumatic vocal hyperfunction.

inspection of group means in the NPVH and PVH training samples, *RFFoffTotal* and *RFFon1+2*. These RFF parameters encompass aspects of the RFF measure known to be sensitive to VH, including late offset cycles (NPVH and PVH) and early onset cycles (PVH; Heller Murray et al., 2017). Differences between Offset Cycles 9 and 10, and 5 and 10, and between Onset Cycles 1 and 2, and 1 and 6, were used to approximate the slope of the offset and onset during the portion of cycles that are changing the most (Lien et al., 2014) while reducing the number of parameters (e.g., individual RFF cycles) entered into the model. The CPPS parameters included in the models were *CPPS mean* and *CPPS SD*, which have been shown to correlate with listener ratings of dysphonia severity and discriminate between speakers with and without voice disorders (Awan et al., 2010; Sauder et al., 2017).

To contrast the relative contributions of RFF and CPPS parameters in predicting disorder status in NPVH versus PVH, separate models were constructed to predict voice disorder status for speakers with NPVH versus control speakers and for speakers with PVH versus control speakers. Sequential predictor entry was used to assess the relative contributions of control parameters (age and sex), RFF parameters, and CPPS parameters in separate blocks. Each model was fit to the training sets twice to assess the effect of RFF parameters above and beyond CPPS (Block 1, control parameters; Block 2, eight RFF parameters; Block 3, two CPPS parameters) and the effect of CPPS parameters above and beyond RFF (Block 1, control parameters; Block 2, two CPPS parameters; Block 3, eight RFF parameters), as shown in (Equations 1–3). Sex was effect-coded (1 = female, −1 = male). Disorder status was dummy-coded (1 = disorder present, 0 = disorder absent). The final model in each case included all control, RFF, and CPPS parameters. Sensitivity and specificity were determined for each model. Pearson correlation coefficients were calculated between all pairs of predictor and outcome parameters to provide descriptive information about relationships among parameters.

$$\text{Logit (Disorder Status)} = b_0 + b_1 \times \text{Age} + b_2 \times \text{Sex} \quad (1)$$

$$\begin{aligned}\text{Logit (Disorder Status)} = {} & b_0 + b_1 \times \text{Age} + b_2 \times \text{Sex} + b_3 \\ & \times \textit{RFFoff 10} + b_4 \times \textit{RFFon1} \\ & + b_5 \times \Delta \textit{RFFoff 10-9} + b_6 \\ & \times \Delta \textit{RFFoff 10-5} + b_7 \\ & \times \textit{RFFoffTotal} + \mathrm{b}_8 \\ & \times \Delta \textit{RFFon1-2} + \mathrm{b}_9 \\ & \times \Delta \textit{RFFon1-6} + \mathrm{b}_{10} \\ & \times \textit{RFFon1+2}\end{aligned} \quad (2a)$$

or

$$\begin{aligned}\text{Logit (Disorder Status)} = {} & b_0 + b_1 \times \text{Age} + b_2 \times \text{Sex} + b_3 \\ & \times \textit{CPPS mean} + b_4 \\ & \times \textit{CPPS SD}\end{aligned} \quad (2b)$$

$$\begin{aligned}\text{Logit (Disorder Status)} = {} & b_0 + b_1 \times \text{Age} + b_2 \times \text{Sex} \\ & + b_3 \times \textit{CPPS mean} + b_4 \\ & \times \textit{CPPS SD} + b_5 \times \textit{RFFoff 10} \\ & + b_6 \times \textit{RFFon1} + b_7 \\ & \times \Delta \textit{RFFoff 10-9} + b_8 \\ & \times \Delta \textit{RFFoff 10-5} + b_9 \\ & \times \textit{RFFoffTotal} + b_{10} \\ & \times \Delta \textit{RFFon1-2} + b_{11} \\ & \times \Delta \textit{RFFon1-6} + b_{12} \\ & \times \textit{RFFon1+2}\end{aligned} \quad (3)$$

In order to determine the simplest models with the best performance in predicting voice disorder status, stepwise logistic regression with forward predictor selection using the likelihood ratio method was subsequently applied to the two training sets (NPVH/control and PVH/control). Alpha for entrance was set to .05, and alpha for removal was .10. Because the case and control samples

were not age- and sex-matched, these demographic parameters were not included in the stepwise models. The two CPPS parameters and eight RFF parameters were entered in the stepwise regression models.

For each model (NPVH and PVH), the equation resulting from the stepwise regression performed on the training set was applied to the independent test set to predict voice disorder status. ROC analyses were used to assess the discriminative performance of the models in detecting the presence of a voice disorder in the test sets. The AUC was calculated for each test set. Sensitivity, specificity, and likelihood ratios (LR+ and LR−) were also calculated to describe the models' performance in classifying the test sets. Sensitivity refers to the percentage of individuals who have a condition who receive a positive result on a test (i.e., true positive rate). Specificity refers to the percentage of individuals who do not have a condition who receive a negative result on a test (i.e., true negative rate). LRs provide information about the diagnostic value of a test by estimating the odds that a person with a positive result will have the condition (LR+ = sensitivity/[1 − specificity]) and a person with a negative result will not (LR− = [1 − sensitivity]/specificity). For example, if LR+ = 2.0, a person who receives a positive test result is twice as likely to have the condition (e.g., VH) than someone with a negative test result. Two different cutoff criteria were used. A cutoff of 0.50 was the criterion used in the regression models and was used to calculate sensitivity, specificity, and LRs. A second cutoff criterion was also selected to maximize sensitivity while maintaining a specificity of at least 60.0% in order to assess the potential of the models as screening tools for VH (Awan et al., 2016). Such screening tools could be useful to detect potential VH in high-risk populations, such as students entering teacher training programs or studying performing voice, or people employed in vocally demanding professions.

## Results

Group mean data for RFF and CPPS parameters in the NPVH and PVH training sets are presented in Figures 1 and 2 for descriptive purposes. Means, standard deviations, and zero-order correlations among all parameters in the NPVH and PVH training sets are provided in Tables 3 and 4.

### Sequential Logistic Regression Models

Two sequential logistic regression models were fit for each of the training sets (NPVH and PVH). The order of entry of predictor parameter blocks was varied across the two fits as follows: (a) demographic parameters, RFF parameters, and CPPS parameters and (b) demographic parameters, CPPS parameters, and RFF parameters. Note that the final models are identical, regardless of the order of entry of predictor parameter blocks. The results of these models are given in Tables 5–8.

**Figure 1.** Relative fundamental frequency (RFF) means and 95% confidence intervals in semitones (ST) for (a) the nonphonotraumatic vocal hyperfunction (NPVH) training set (speakers with NPVH, $n$ = 180, blue triangles; controls, $n$ = 180, black squares), and (b) the phonotraumatic vocal hyperfunction (PVH) training set (speakers with PVH, $n$ = 120, red circles; controls, $n$ = 120, black squares).

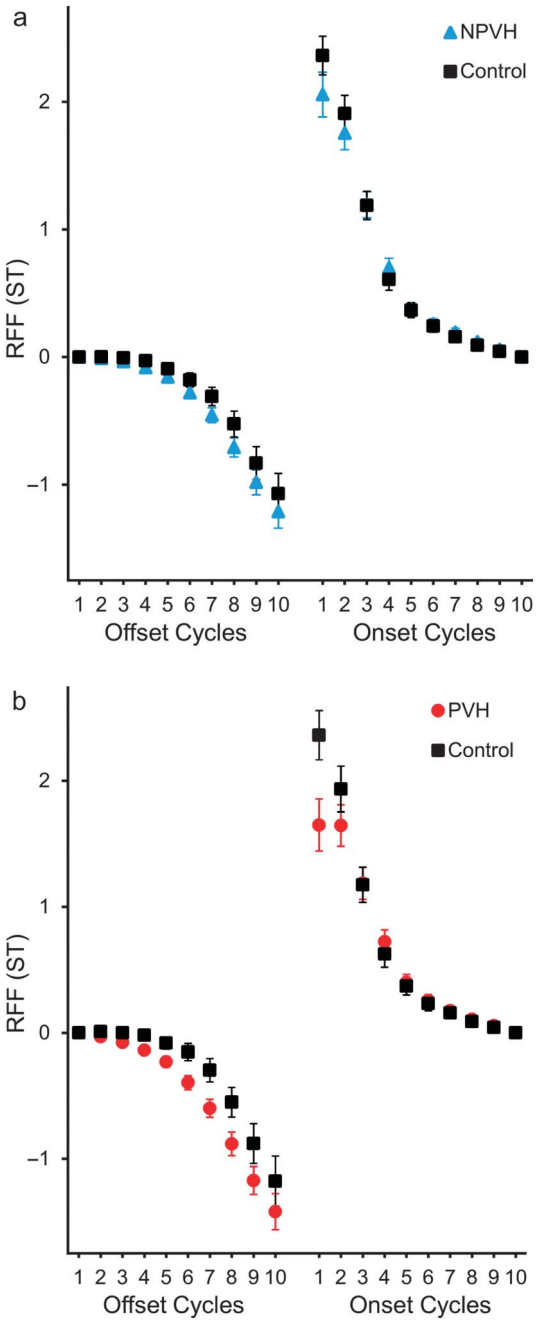
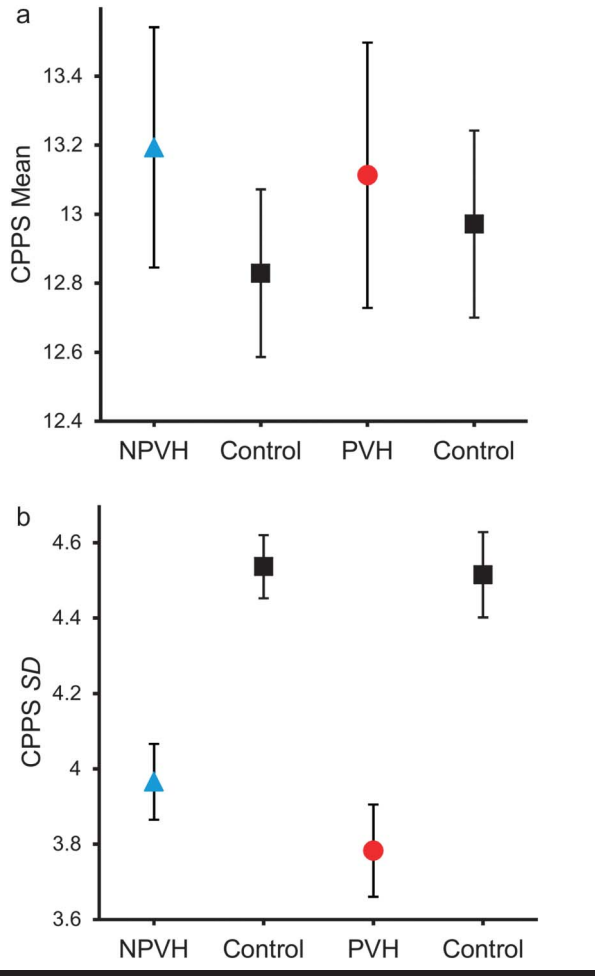


**Figure 2.** Means and 95% confidence intervals for (a) *CPPS mean* and (b) *CPPS SD*, for the nonphonotraumatic vocal hyperfunction (NPVH) training set (speakers with NPVH, *n* = 180, blue triangles; controls, *n* = 180, black squares) and the phonotraumatic vocal hyperfunction (PVH) training set (speakers with PVH, *n* = 120, red circles; controls, *n* = 120, black squares). CPPS = smoothed cepstral peak prominence.




## NPVH

As shown in Table 5, Block 1, which included the control parameters sex (effect coded, 1 = female, −1 = male) and age, did not significantly improve model fit for NPVH compared to the null model with no predictors, $\chi^2 = 2.10$, $p = .350$, Nagelkerke pseudo $R^2 = .01$ (correct classification hit rate of 53.3%). When RFF parameters were entered in Block 2, there was a significant improvement in model fit to the data, $\chi^2_{change} = 17.55$, $p = .014$, Nagelkerke pseudo $R^2 = .07$ (correct classification hit rate increased to 59.4%). This change was driven by *RFFoff10-9* and *RFFoffTotal*. When CPPS parameters were entered in Block 3, there was a significant improvement in model fit, $\chi^2_{change} = 67.87$, $p < .001$, Nagelkerke pseudo $R^2 = .29$ (correct classification hit rate increased to 71.4%). This change was driven by both *CPPS mean* and *CPPS SD*. This indicates that CPPS parameters uniquely explained variance in NPVH status above and beyond sex, age, and RFF parameters. Sensitivity (correct identifications) of the final model was 71.1%, and specificity (correct rejections) was 71.7%.

As shown in Table 6, when CPPS parameters were instead entered in Block 2, there was a significant improvement in model fit to the data, $\chi^2_{change} = 74.47$, $p < .001$, Nagelkerke pseudo $R^2 = .26$ (correct classification hit rate increased to 66.9%) over Block 1. This change was driven by both *CPPS mean* and *CPPS SD*. When RFF parameters were entered in Block 3, there was not a significant improvement in model fit, $\chi^2_{change} = 10.95$, $p = .141$, Nagelkerke pseudo $R^2 = .29$, though there was an increase in the correct classification hit rate (71.4%). This indicates that RFF parameters did not uniquely explain variance in NPVH status above and beyond sex, age, and CPPS parameters.

## PVH

As shown in Table 7, Block 1, which included the control parameters sex (effect coded, 1 = female, −1 = male) and age, significantly improved model fit for PVH compared to the null model with no predictors, $\chi^2 = 10.50$, $p = .005$, Nagelkerke pseudo $R^2 = .06$ (correct classification hit rate of 58.3%). This was driven by a small but significant coefficient for age. When RFF parameters were entered in Block 2, there was a significant improvement in model fit to the data, $\chi^2_{change} = 47.83$, $p < .001$, Nagelkerke pseudo $R^2 = .29$ (correct classification hit rate increased to 70.0%). This change was driven by *RFFoff10* and *RFFoffTotal*. When CPPS parameters were entered in Block 3, there was a significant improvement in model fit, $\chi^2_{change} = 42.01$, $p < .001$, Nagelkerke pseudo $R^2 = .46$ (correct classification hit rate increased to 77.5%). This change was driven by *CPPS SD*. This indicates that CPPS parameters uniquely explained variance in PVH status above and beyond sex, age, and RFF parameters. Sensitivity (correct identifications) of the final model was 78.3%, and specificity (correct rejections) was 76.7%.

As shown in Table 8, when CPPS parameters were instead entered in Block 2, there was a significant improvement in model fit to the data, $\chi^2_{change} = 62.73$, $p < .001$, Nagelkerke pseudo $R^2 = .35$ (correct classification hit rate increased to 71.3%). This change was driven by *CPPS SD*. When RFF parameters were entered in Block 3, there was a significant improvement in model fit, $\chi^2_{change} = 27.12$, $p < .001$, Nagelkerke pseudo $R^2 = .46$ (increase in the correct classification hit rate to 78%). This change was driven by *RFFoffTotal*. This indicates that RFF parameters uniquely explained variance in PVH status above and beyond sex, age, and CPPS parameters.

**Table 3.** Pairwise Pearson correlation coefficients between outcome and predictor parameters in the nonphonotraumatic vocal hyperfunction (NPVH) training set.

| Measure | *M* | *SD* | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. | 9. | 10. | 11. | 12. | 13. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Outcomes | | | | | | | | | | | | | | | |
| 1. NPVH | 0.50 | (0.50) | — | | | | | | | | | | | | |
| Demographic parameters | | | | | | | | | | | | | | | |
| 2. Sex (female) | 0.66 | (0.47) | .07 | — | | | | | | | | | | | |
| 3. Age | 39.21 | (19.65) | .03 | −.03 | — | | | | | | | | | | |
| CPPS parameters | | | | | | | | | | | | | | | |
| 4. *CPPS mean* | 13.01 | (2.06) | .09 | **−.12** | −.10 | — | | | | | | | | | |
| 5. *CPPS SD* | 4.25 | (0.69) | **−.41** | **−.19** | .07 | .10 | — | | | | | | | | |
| RFF parameters | | | | | | | | | | | | | | | |
| 6. *RFFoff10* | −1.14 | (0.99) | −.07 | .02 | **−.20** | .02 | .00 | — | | | | | | | |
| 7. *ΔRFFoff10-9* | −0.23 | (0.58) | .01 | .08 | −.07 | .01 | −.02 | **.61** | — | | | | | | |
| 8. *ΔRFFoff10-5* | −1.02 | (0.93) | −.04 | .05 | **−.20** | .03 | −.05 | **.97** | **.68** | — | | | | | |
| 9. *RFFoffTotal* | −3.49 | (3.26) | **−.14** | −.05 | **−.16** | .01 | .08 | **.82** | **.15** | **.68** | — | | | | |
| 10. *RFFon1* | 2.21 | (1.13) | **−.14** | **−.20** | .00 | .07 | **.17** | .05 | −.05 | .03 | **.11** | — | | | |
| 11. *ΔRFFon1-2* | 0.38 | (0.85) | −.09 | **−.36** | .05 | .09 | **.21** | **.11** | .02 | .09 | **.14** | **.58** | — | | |
| 12. *ΔRFFon1-6* | 1.96 | (1.08) | **−.15** | **−.24** | .01 | .05 | **.17** | .08 | −.05 | .05 | **.14** | **.96** | **.63** | — | |
| 13. *RFFon1+2* | 4.04 | (1.90) | **−.12** | −.08 | −.02 | .04 | **.11** | .02 | −.08 | .00 | .07 | **.93** | **.25** | **.87** | — |

*Note.* $N = 360$. Bolded values indicate statistically significant correlations ($p < .05$). CPPS = smoothed cepstral peak prominence; RFF = relative fundamental frequency.

## Stepwise Logistic Regression Models

To generate parsimonious regression equations for application to the independent test sets and potential future use, two stepwise logistic regression models were run to predict NPVH and PVH status using the training sets. As shown in Table 9, this resulted in a significant model fit for NPVH, $\chi^2(3) = 79.02$, $p < .001$, Nagelkerke pseudo $R^2 = .26$ (correct classification hit rate of 67.2%). Sensitivity (correct identifications) of this final model was 63.9%, and specificity (correct rejections) was 70.6%. The resulting regression equation is shown in Equation 4. Note that the correct classification hit rate and the variance explained by the NPVH stepwise model is equal to the NPVH CPPS-only model, as can be seen by comparing Tables 9 and 6. As shown in Table 10, stepwise regression also resulted in a significant model fit for PVH, $\chi^2(4) = 88.90$, $p < .001$, Nagelkerke pseudo $R^2 = .41$ (correct classification hit rate of 75.0%). Sensitivity (correct identifications) of this final model was 72.0%, and specificity (correct rejections) was 78.0%. The resulting regression equation is shown in Equation 5.

**Table 4.** Pairwise Pearson correlation coefficients between outcome and predictor parameters in the phonotraumatic vocal hyperfunction (PVH) training set.

| Measure | *M* | *SD* | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. | 9. | 10. | 11. | 12. | 13. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Outcomes | | | | | | | | | | | | | | | |
| 1. PVH | 0.50 | (0.50) | — | | | | | | | | | | | | |
| Demographic parameters | | | | | | | | | | | | | | | |
| 2. Sex (female) | 0.73 | (0.45) | **.15** | — | | | | | | | | | | | |
| 3. Age | 34.88 | (18.51) | **−.18** | **−.24** | — | | | | | | | | | | |
| CPPS parameters | | | | | | | | | | | | | | | |
| 4. *CPPS mean* | 13.04 | (1.86) | .04 | −.01 | −.10 | — | | | | | | | | | |
| 5. *CPPS SD* | 4.15 | (0.75) | **−.49** | **−.27** | .09 | **.17** | — | | | | | | | | |
| RFF parameters | | | | | | | | | | | | | | | |
| 6. *RFFoff10* | −1.30 | (0.98) | −.12 | −.01 | −.04 | .04 | .01 | — | | | | | | | |
| 7. *ΔRFFoff10-9* | −0.27 | (0.51) | .05 | **.21** | .02 | .07 | **−.17** | **.62** | — | | | | | | |
| 8. *ΔRFFoff10-5* | −1.14 | (0.89) | −.05 | .05 | −.05 | .04 | −.05 | **.97** | **.68** | — | | | | | |
| 9. *RFFoffTotal* | −4.05 | (3.37) | **−.27** | **−.18** | −.04 | .02 | **.19** | **.85** | **.23** | **.70** | — | | | | |
| 10. *RFFon1* | 2.01 | (1.18) | **−.30** | **−.28** | .05 | .08 | **.30** | .06 | −.03 | .00 | **.18** | — | | | |
| 11. *ΔRFFon1-2* | 0.22 | (0.91) | **−.23** | **−.30** | .07 | **.17** | **.27** | **.13** | −.06 | .07 | **.25** | **.59** | — | | |
| 12. *ΔRFFon1-6* | 1.76 | (1.14) | **−.32** | **−.29** | .07 | .09 | **.30** | .10 | −.01 | .03 | **.21** | **.97** | **.63** | — | |
| 13. *RFFon1+2* | 3.80 | (1.97) | **−.26** | **−.20** | .03 | .02 | **.23** | .02 | −.01 | −.03 | .10 | **.93** | **.24** | **.87** | — |

*Note.* $N = 240$. Bolded values indicate statistically significant correlations ($p < .05$). CPPS = smoothed cepstral peak prominence; RFF = relative fundamental frequency.

**Table 5.** Nonphonotraumatic vocal hyperfunction logistic regression results using sequential predictor entry, with relative fundamental frequency (RFF) entered before smoothed cepstral peak prominence (CPPS).

| | Block 1 | | | | Block 2 | | | | | Block 3 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Variable** | $\chi^2_{total}$ | **Nagel.** | **HR** | ***b*** | $\chi^2_{change}$ | $\chi^2_{total}$ | **Nagel.** | **HR** | ***b*** | $\chi^2_{change}$ | $\chi^2_{total}$ | **Nagel.** | **HR** | ***b*** |
| Model fit | 2.10 | .01 | 0.53 | | **17.55** | **19.65** | .07 | 0.59 | | **67.87** | **87.52** | .29 | 0.71 | |
| Coefficients | | | | | | | | | | | | | | |
| Intercept | | | | –0.17 | | | | | 0.26 | | | | | **4.38** |
| Female | | | | 0.53 | | | | | 0.08 | | | | | 0.02 |
| Age | | | | 0.00 | | | | | 0.00 | | | | | 0.01 |
| *RFFoff10* | | | | | | | | | 3.63 | | | | | 3.85 |
| *ΔRFFoff10-9* | | | | | | | | | **–1.08** | | | | | –0.81 |
| *ΔRFFoff10-5* | | | | | | | | | –2.05 | | | | | –2.54 |
| *RFFoffTotal* | | | | | | | | | **–0.56** | | | | | –0.52 |
| *RFFon1* | | | | | | | | | 0.20 | | | | | 0.28 |
| *ΔRFFon1-2* | | | | | | | | | 0.11 | | | | | 0.26 |
| *ΔRFFon1-6* | | | | | | | | | –0.50 | | | | | –0.58 |
| *RFFon1+2* | | | | | | | | | | | | | | |
| *CPPS mean* | | | | | | | | | | | | | | **0.16** |
| *CPPS SD* | | | | | | | | | | | | | | **–1.52** |

*Note.* $N = 360$. Bolded values indicate statistical significance ($p < .05$). Block 1 chi-square $df = 2$; Block 2 $df = 9$; Block 3 $df = 11$. *RFFon1+2* did not enter into the model due to redundancies. Nagel. = Nagelkerke pseudo $R_2$; HR = hit rate.

**Table 6.** Nonphonotraumatic vocal hyperfunction logistic regression results using sequential predictor entry, with smoothed cepstral peak prominence (CPSP) entered before relative fundamental frequency (RFF).

| | Block 1 | | | | Block 2 | | | | | Block 3 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Variable** | **$\chi^2_{total}$** | **Nagel.** | **HR** | ***b*** | **$\chi^2_{change}$** | **$\chi^2_{total}$** | **Nagel.** | **HR** | ***b*** | **$\chi^2_{change}$** | **$\chi^2_{total}$** | **Nagel.** | **HR** | ***b*** |
| Model fit | 2.10 | .01 | 0.53 | | **74.47** | **76.57** | .26 | 0.67 | | 10.95 | **87.52** | .29 | 0.71 | |
| Coefficients | | | | | | | | | | | | | | |
| Intercept | | | | –0.17 | | | | | **4.02** | | | | | **4.38** |
| Female | | | | 0.53 | | | | | 0.03 | | | | | 0.02 |
| Age | | | | 0.00 | | | | | 0.01 | | | | | 0.01 |
| *CPPS mean* | | | | | | | | | **0.17** | | | | | **0.16** |
| *CPPS SD* | | | | | | | | | **–1.54** | | | | | **–1.52** |
| *RFFoff10* | | | | | | | | | | | | | | 3.85 |
| *ΔRFFoff10-9* | | | | | | | | | | | | | | –0.81 |
| *ΔRFFoff10-5* | | | | | | | | | | | | | | –2.54 |
| *RFFoffTotal* | | | | | | | | | | | | | | –0.52 |
| *RFFon1* | | | | | | | | | | | | | | 0.28 |
| *ΔRFFon1-2* | | | | | | | | | | | | | | 0.26 |
| *ΔRFFon1-6* | | | | | | | | | | | | | | –0.58 |
| *RFFon1+2* | | | | | | | | | | | | | | |

*Note.* Bolded values indicate statistical significance ($p < .05$). $N = 360$. Block 1 chi-square $df = 2$; Block 2 $df = 4$; Block 3 $df = 11$. *RFFon1+2* did not enter into the model due to redundancies. Nagel. = Nagelkerke pseudo R2; HR = hit rate.

**Table 7.** Phonotraumatic vocal hyperfunction logistic regression results using sequential predictor entry, with relative fundamental frequency (RFF) entered before smoothed cepstral peak prominence (CPPS).

| | Block 1 | | | | Block 2 | | | | | Block 3 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Variable** | $\chi^2_{total}$ | **Nagel.** | **HR** | ***b*** | $\chi^2_{change}$ | $\chi^2_{total}$ | **Nagel.** | **HR** | ***b*** | $\chi^2_{change}$ | $\chi^2_{total}$ | **Nagel.** | **HR** | ***b*** |
| Model fit | **10.50** | .06 | 0.58 | | **47.83** | **58.34** | .29 | 0.70 | | **42.01** | **100.35** | .46 | 0.78 | |
| Coefficients | | | | | | | | | | | | | | |
| Intercept | | | | 0.46 | | | | | **1.50** | | | | | **5.85** |
| Female | | | | 0.26 | | | | | –0.15 | | | | | –0.32 |
| Age | | | | **–0.02** | | | | | **–0.03** | | | | | **–0.02** |
| *RFFoff10* | | | | | | | | | **6.69** | | | | | 4.87 |
| *ΔRFFoff10-9* | | | | | | | | | –1.37 | | | | | –1.33 |
| *ΔRFFoff10-5* | | | | | | | | | –3.84 | | | | | –2.66 |
| *RFFoffTotal* | | | | | | | | | **–1.06** | | | | | **–0.83** |
| *RFFon1* | | | | | | | | | 0.30 | | | | | 0.41 |
| *ΔRFFon1-2* | | | | | | | | | 0.01 | | | | | 0.11 |
| *ΔRFFon1-6* | | | | | | | | | –0.88 | | | | | –0.95 |
| *RFFon1+2* | | | | | | | | | | | | | | |
| *CPPS mean* | | | | | | | | | | | | | | 0.17 |
| *CPPS SD* | | | | | | | | | | | | | | **–1.62** |

*Note.* *N* = 240. Bolded values indicate statistical significance ($p < .05$). Block 1 chi-square *df* = 2; Block 2 *df* = 9; Block 3 *df* = 11. *RFFon1+2* did not enter into the model due to redundancies. Nagel. = Nagelkerke pseudo R2; HR = hit rate.

**Table 8.** Phonotraumatic vocal hyperfunction logistic regression results using sequential predictor entry, with smoothed cepstral peak prominence (CPPS) entered before relative fundamental frequency (RFF).

| | Block 1 | | | | Block 2 | | | | | Block 3 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Variable** | **$\chi^2_{total}$** | **Nagel.** | **HR** | ***b*** | **$\chi^2_{change}$** | **$\chi^2_{total}$** | **Nagel.** | **HR** | ***b*** | **$\chi^2_{change}$** | **$\chi^2_{total}$** | **Nagel.** | **HR** | ***b*** |
| Model fit | **10.50** | .06 | 0.58 | | **62.73** | **73.23** | .35 | 0.71 | | **27.12** | **100.35** | .46 | 0.78 | |
| Coefficients | | | | | | | | | | | | | | |
| Intercept | | | | 0.46 | | | | | **6.01** | | | | | **5.85** |
| Female | | | | 0.26 | | | | | –0.06 | | | | | –0.32 |
| Age | | | | **–0.02** | | | | | **–0.02** | | | | | **–0.02** |
| *CPPS mean* | | | | | | | | | 0.16 | | | | | 0.17 |
| *CPPS SD* | | | | | | | | | **–1.78** | | | | | **–1.62** |
| *RFFoff10* | | | | | | | | | | | | | | 4.87 |
| *ΔRFFoff10-9* | | | | | | | | | | | | | | –1.33 |
| *ΔRFFoff10-5* | | | | | | | | | | | | | | –2.66 |
| *RFFoffTotal* | | | | | | | | | | | | | | **–0.83** |
| *RFFon1* | | | | | | | | | | | | | | 0.41 |
| *ΔRFFon1-2* | | | | | | | | | | | | | | 0.11 |
| *ΔRFFon1-6* | | | | | | | | | | | | | | –0.95 |
| *RFFon1+2* | | | | | | | | | | | | | | |

*Note.* *N* = 240. Bolded values indicate statistical significance ($p < .05$). Block 1 chi-square *df* = 2; Block 2 *df* = 4; Block 3 *df* = 11. *RFFon1+2* did not enter into the model due to redundancies. Nagel. = Nagelkerke pseudo R2; HR = hit rate.

**Table 9.** Nonphonotraumatic vocal hyperfunction (NPVH) logistic regression results using stepwise predictor entry.

| Variable | $\chi^2(3)$ | *p* | *Pseudo* $R^2$ | Sens | Spec | HR | *b* | *(SE)* | *Wald* | *p* |
|---|---|---|---|---|---|---|---|---|---|---|
| NPVH model fit | 79.02 | < .001 | .26 | 0.64 | 0.71 | 0.67 | | | | |
| Coefficients | | | | | | | | | | |
| Intercept | | | | | | | 4.05 | (1.05) | 14.81 | < .001 |
| *CPPS mean* | | | | | | | 0.16 | (0.06) | 7.63 | .006 |
| *CPPS SD* | | | | | | | –1.51 | (0.21) | 52.31 | < .001 |
| *RFFoffTotal* | | | | | | | –0.08 | (0.04) | 4.40 | .036 |

*Note.* *N* = 360. Sens = sensitivity; Spec = specificity; HR = hit rate; *SE* = standard error; CPPS = smoothed cepstral peak prominence; RFF = relative fundamental frequency.

$$\text{Logit (NPVH Status)} = 4.05 + 0.16 \times CPPS\ mean - 1.51 \times CPPS\ SD - 0.08 \times RFFoffTotal \quad (4)$$

$$\text{Logit (PVH Status)} = 4.56 + 0.20 \times CPPS\ mean - 1.66 \times CPPS\ SD - 0.46 \times \Delta RFFon1\text{-}6 - 0.13 \times RFFoffTotal \quad (5)$$

## Classification Accuracy Using Test Sets

The regression equations resulting from the two stepwise regression models were used to calculate predicted probabilities (1 = disorder, 0 = control) for the speakers in the untrained NPVH and PVH test sets. ROC curves were generated using the predicted probabilities from each test set (see Figure 3). ROC analysis revealed an AUC for the NPVH test set of .72 (95% CI [.61, .82]), consistent with an acceptable classifier. For PVH, AUC was .86 (95% CI [.79, .93]), consistent with a good classifier (Hosmer et al., 2013).

Likelihood ratios, sensitivity, and specificity were calculated for two different cutoff criteria. A cutoff of 0.50 resulted in an LR+ of 1.94 and an LR− of 0.52 for NPVH (sensitivity = 66.0%, specificity = 66.0%) and an LR+ of 3.27 and an LR− of 0.36 for PVH (sensitivity = 72.0%, specificity = 78.0%). For NPVH, a screening cutoff criterion of 0.48 yielded an LR+ of 1.70 and an LR− of 0.53 (sensitivity = 68.0%, specificity = 60%). For PVH, a screening cutoff criterion of 0.31 yielded an LR+ of 2.25 and an LR− of 0.17 (sensitivity = 90.0%, specificity = 60%).

## Discussion

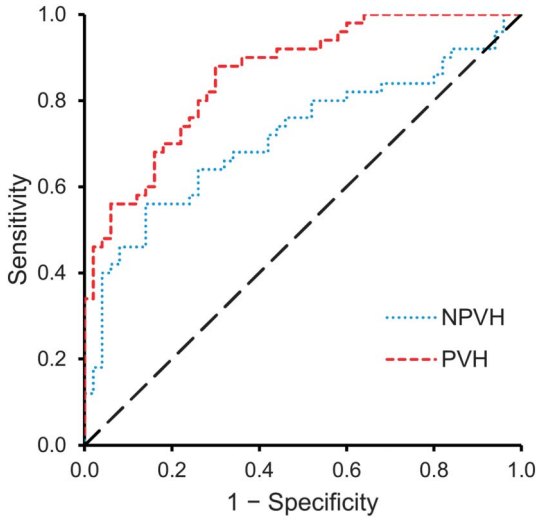
This study evaluated the discriminative ability of a combined RFF/CPPS acoustic index of VH in a large sample of speakers with NPVH, PVH, and typical voices. In addition to employing sample sizes that were substantially larger than those analyzed in prior studies of RFF (Heller Murray et al., 2017; Roy et al., 2016; Stepp et al., 2010) and CPPS (Radish Kumar et al., 2010; Mahalingam et al., 2020) in this population, we validated the performance of our models using large ($n$ = 100 each) independent test sets for NPVH and PVH. Validation of regression models using independent data (i.e., cases that were not included in the original analysis) is crucial to accurately assess the performance of a model, as overfitting of predictors to the original data set may inflate the model's performance and limit generalizability (Gareth et al., 2013). The results of this study show that a combination of RFF and CPPS acoustic parameters can differentiate between speakers with hyperfunctional voice disorders and speakers with typical voices with acceptable (NPVH) to good (PVH) discriminative performance based on AUC. When the results of logistic regression models derived from training sets were applied to untrained test stimuli, speakers with NPVH were classified with 66% sensitivity

**Table 10.** Phonotraumatic vocal hyperfunction (PVH) logistic regression results using stepwise predictor entry.

| Variable | $\chi^2(4)$ | *p* | *Pseudo* $R^2$ | Sens | Spec | HR | *b* | *(SE)* | *Wald* | *p* |
|---|---|---|---|---|---|---|---|---|---|---|
| PVH model fit | 88.90 | < .001 | .41 | 0.72 | 0.78 | 0.75 | | | | |
| Coefficients | | | | | | | | | | |
| Intercept | | | | | | | 4.56 | (1.42) | 10.26 | .001 |
| *CPPS mean* | | | | | | | 0.20 | (0.09) | 4.99 | .025 |
| *CPPS SD* | | | | | | | –1.66 | (0.28) | 36.31 | < .001 |
| *ΔRFFon1-6* | | | | | | | –0.46 | (0.15) | 9.09 | .003 |
| *RFFoffTotal* | | | | | | | –0.13 | (0.05) | 7.44 | .006 |

*Note.* *N* = 240. Sens = sensitivity; Spec = specificity; HR = hit rate; *SE* = standard error; CPPS = smoothed cepstral peak prominence; RFF = relative fundamental frequency.

**Figure 3.** Receiver operating characteristic curves for classification results of the nonphonotraumatic vocal hyperfunction (NPVH) test set ($n$ = 100, blue) and the phonotraumatic vocal hyperfunction (PVH) test set ($n$ = 100, red).




and 66% specificity, whereas a higher accuracy of 72% sensitivity and 78% specificity was obtained for PVH. Both RFF and CPPS parameters were significant predictors of NPVH and PVH status. CPPS parameters uniquely predicted variance in both NPVH and PVH status after accounting for variance associated with RFF parameters. However, whereas RFF parameters did not uniquely predict variance in NPVH status after accounting for variance associated with CPPS parameters, RFF did uniquely predict variance in PVH status above and beyond CPPS. Furthermore, the correct classification hit rate and variance explained were equal in a CPPS-only model (see Table 6) and a combined RFF/CPPS acoustic index (see Table 9) for NPVH. In the final regression models, *CPPS mean*, *CPPS SD*, and *RFFoffTotal* (the sum of RFF offset cycles) were significant predictors of both NPVH and PVH. Onset RFF parameters, specifically the slope between Onset Cycles 1 and 6 (*RFFon1-6*), were significantly predictive of PVH and not NPVH, consistent with a previous study of RFF in these populations (Heller Murray et al., 2017).

RFF offset parameters were significant predictors of both NPVH and PVH status. Specifically, in both the NPVH and PVH samples, a more negative *RFFoffTotal* was predictive of having a voice disorder, congruent with the larger downward trajectory in RFF offset cycles seen in group mean data for both NPVH and PVH compared to speakers with typical voices (see Figure 1). This is consistent with prior studies of RFF in speakers with VH (Heller Murray et al., 2017; Lien et al., 2015; Stepp et al., 2010, 2012). Heller Murray et al. (2017) proposed that elevated longitudinal vocal fold tension occurs in both speakers with NPVH and speakers with PVH and prevents these speakers from further increasing tension to counteract the effects of abduction and decreased transglottal pressure during voicing offsets, which cause $f_o$ to decrease. In addition to increased longitudinal vocal fold tension, speakers with PVH are hypothesized to have increased transverse vocal fold tension associated with higher collision forces, as a cause and/or compensation for phonotrauma (Hillman et al., 2020). The loss of collision forces during voicing offsets is associated with decreased $f_o$; thus, higher collision forces during steady-state phonation in PVH may lead to a larger decrease in $f_o$ during voicing offset (Heller Murray et al., 2017; Serry et al., 2021), as reflected by the more negative coefficient for *RFFoffTotal* in the PVH regression model than in the NPVH model.

RFF onset parameters significantly predicted PVH and not NPVH status. Specifically, a smaller value for *RFFon1-6*, indicative of a shallower RFF onset slope, was predictive of having PVH. This is consistent with prior work by Heller Murray et al. (2017) that showed lower RFF Onset Cycle 1 values in speakers with PVH compared to speakers with typical voices and provides further support for their hypothesized model of the effects of VH on RFF. Brief increases in longitudinal vocal fold tension that occur during devoicing are thought to cause an increase in $f_o$ during voicing onsets in speakers with typical voices (Halle & Stevens, 1971; Stepp et al., 2010). Elevated baseline longitudinal vocal fold tension is thought to reduce this effect in speakers with VH, leading to less positive onset RFF values (Stepp et al., 2010). Furthermore, during voicing onsets, increased transverse tension may result in lower RFF onset values, as faster closing velocities may reduce the impact of increased transglottal pressure that is thought to increase $f_o$ during onsets (Heller Murray et al., 2017; Ladefoged, 1967). The combined effects of increased longitudinal and transverse vocal fold tension in speakers with PVH may underlie our finding of lower onset RFF values in this population. In addition, the combination of both offset and onset RFF differences may be one reason why our regression models performed better at differentiating speakers with PVH from controls than speakers with NPVH.

It is noteworthy that CPPS parameters alone did not perform as well at classifying speakers with voice disorders in our study as in prior studies (e.g., Murton et al., 2020; Sauder et al., 2017). In addition, the average *CPPS mean* was similar across VH and control groups, as shown in Figure 2. This may be due to differences in the populations sampled in these studies. Prior studies of the discriminative ability of CPPS-derived acoustic measures have included speakers with a wide range of voice disorders (Awan et al., 2016; Sauder et al., 2017), whereas this study included only speakers with hyperfunctional voice disorders. In addition, the mean OS of voice quality in our sample was relatively low (NPVH mean = 16.2/100; PVH mean = 20.6/100). This may reflect differences between

people with VH versus the broader clinical population of people with voice disorders. Speakers with PVH and NPVH may or may not present with atypical voice quality and may instead report primary symptoms such as vocal fatigue and increased vocal effort (Hunter et al., 2020; Solomon, 2008). When atypical voice quality is present, it is possible that it is driven by factors other than those associated with low *CPPS mean* values (e.g., breathiness; Heman-Ackah et al., 2002), such as strain. Furthermore, in speakers with PVH, compensation for the presence of phonotrauma may explain the lack of difference in *CPPS mean*, which replicates findings from ambulatory monitoring studies (Van Stan, Mehta, Ortiz, Burns, Toles, et al., 2020). Our findings suggest that RFF, an acoustic measure thought to be sensitive to differences in laryngeal stiffness (McKenna et al., 2016), may provide complementary information to CPPS, an acoustic measure related to voice quality (Maryn, Corthals, et al., 2010). The combination of these two measures as an acoustic index of VH provided superior sensitivity and specificity, in particular for PVH.

In this study, we explored potential clinical cutoff scores for acoustic identification of hyperfunctional voice disorders. The choice of cutoff score may depend on the purpose for which it is used. Cutoff scores of 0.50, the criterion used in our regression models, yielded an LR+ of 1.94 and an LR− of 0.52 for NPVH (sensitivity = 66.0%, specificity = 66.0%) and an LR+ of 3.27 and an LR− of 0.36 for PVH (sensitivity = 72.0%, specificity = 78.0%) in our untrained test data. This criterion provided fair discriminative accuracy for NPVH and good discriminative accuracy for PVH. As part of a comprehensive voice assessment, this measure may provide objective information about the presence of VH. An objective measure of VH could improve clinical management of these disorders by informing treatment decisions as part of a comprehensive assessment (detecting persistent hyperfunction after surgery, assessing discharge readiness, etc.). In addition, this measure may serve as a screening tool to assist prevention in at-risk individuals. For screening purposes, maximizing sensitivity while allowing for somewhat decreased specificity may be appropriate. For PVH, a substantially lower cutoff score (0.31) yielded an excellent sensitivity of 90% while maintaining 60% specificity. This cutoff score may be appropriate for use in voice screenings or clinical contexts without access to laryngoscopy to identify patients at high risk of PVH and to assess the urgency of otolaryngology referrals.

Several limitations of this study should be noted. First, prior treatment history was not controlled in this sample. However, all speakers with voice disorders were recently seen by a laryngologist, had a current diagnosis consistent with NPVH or PVH, and were symptomatic at the time of recording. Second, as noted in the Method section, 9.9% of cases in our sample had missing RFF data, because an insufficient number of RFF instances (minimum of two) could be calculated from their voice recordings. It is likely that missingness was influenced by systematic factors, such as dysphonia severity (i.e., RFF data were MNAR). For this reason, data estimation methods were not used, and cases with missing data were deleted listwise. This approach to missing data may cause bias. However, if cases with missing data tended to be more severe, this would be likely to decrease the overall discriminative accuracy of the models, as discriminative accuracy is more difficult for cases that are mild to moderate (which includes the majority of our samples based on auditory-perceptual ratings of OS). Given that our models performed with acceptable-to-good accuracy despite this bias, this supports their clinical utility. However, further testing of the models with samples reflecting a wider range of severity would be necessary to confirm this. Finally, regression results in this study related to the demographic parameters age and sex were used as control variables only and should not be interpreted with regard to prevalence of VH. Epidemiological study of the prevalence and risk factors associated with voice disorders, including VH, has revealed a higher risk of occurrence in female individuals, and in certain age groups (Roy et al., 2005). In this study, sex was not a significant predictor of VH status in either model, and age was a modest predictor of PVH. However, because our control samples comprised speakers with typical voices from several different experiments that varied in demographic inclusion criteria, they do not constitute a truly random sample in terms of age and sex. Furthermore, although we did not match our samples for age or sex, we intentionally removed excess speakers from oversampled age ranges (18–28 and 58–68 years) in order to create similar age distributions between our patient and control samples. Thus, our findings should not be interpreted in terms of the demographic prevalence of VH.

Results of this study support the use of a multiparameter acoustic index of VH as part of a comprehensive voice assessment and for screening purposes to identify individuals at high risk of PVH. However, future studies of patients with VH before and after successful voice therapy are needed to evaluate sensitivity of the index to change and to define a minimally important difference. Longitudinal study of high-risk voice users would provide additional support for the screening value of the indices. Additional acoustic measures should also be evaluated for their potential to improve the discriminative accuracy of the NPVH model. For example, acoustic measures related to the perception of strain (Anand et al., 2019) and breathiness (von Latoszek et al., 2018) may be sensitive to changes in voice quality related to increased laryngeal tension in the context of a high open quotient of vibration (Espinoza et al., 2017). Specifically, a linear combination of vowel spectral moments and CPPS may explain more variance in perceptual ratings of strain than CPPS alone (Anand et al., 2019). Similarly, a linear combination of spectral and perturbation measures in addition to CPPS

may explain more variance in perceptual ratings of breathiness than CPPS alone (von Latoszek et al., 2017). These acoustic measures could be examined for their potential to improve discriminative accuracy in NPVH above and beyond CPPS. Furthermore, use of additional stimuli such as more natural connected speech for CPPS measures may also improve performance of the models.

## Conclusions

Multiparameter acoustic indices of NPVH and PVH comprising CPPS and RFF parameters were developed. Although CPPS and RFF parameters provided complementary information about the presence of PVH, RFF parameters did not add unique variance beyond that of CPPS measures for NPVH. Logistic regression models determined that RFF offset parameters were significant predictors of both NPVH and PVH status, whereas RFF onset significantly predicted PVH, but not NPVH, consistent with hypothesized laryngeal biomechanical differences in these two populations. The NPVH index showed acceptable performance in an untrained test set in discriminating speakers with voice disorders from speakers with typical voices (AUC = .72), and the PVH index showed good performance (AUC = .86). Proposed clinical cutoff scores may provide objective information about the likelihood of VH as part of a comprehensive voice evaluation and as a screening tool for PVH.

## Acknowledgments

This work was supported by the National Institutes of Health Grants R01 DC015570 (C.E.S.), P50 DC015446 (R.E.H.), and T32 DC005361 (D.J.P.) from the National Institute on Deafness and Other Communication Disorders. The article's contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health. The authors thank Anton Dolling for assistance with participant information retrieval and Jacob Noordzij, Jr., and Roxanne Segina for acoustic signal quality checks.

## References

**Anand, S., Kopf, L. M., Shrivastav, R., & Eddins, D. A.** (2019). Objective indices of perceived vocal strain. *Journal of Voice, 33*(6), 838–845. https://doi.org/10.1016/j.jvoice.2018.06.005

**Awan, S. N., & Roy, N.** (2005). Acoustic prediction of voice type in women with functional dysphonia. *Journal of Voice, 19*(2), 268–282. https://doi.org/10.1016/j.jvoice.2004.03.005

**Awan, S. N., & Roy, N.** (2009). Outcomes measurement in voice disorders: Application of an acoustic index of dysphonia severity. *Journal of Speech, Language, and Hearing Research, 52*(2), 482–499. https://doi.org/10.1044/1092-4388(2008/08-0034)

**Awan, S. N., Roy, N., & Dromey, C.** (2009). Estimating dysphonia severity in continuous speech: Application of a multi-parameter spectral/cepstral model. *Clinical Linguistics & Phonetics, 23*(11), 825–841. https://doi.org/10.3109/02699200903242988

**Awan, S. N., Roy, N., Jette, M. E., Meltzner, G. S., & Hillman, R. E.** (2010). Quantifying dysphonia severity using a spectral/cepstral-based acoustic index: Comparisons with auditory-perceptual judgements from the CAPE-V. *Clinical Linguistics & Phonetics, 24*(9), 742–758. https://doi.org/10.3109/02699206.2010.492446

**Awan, S. N., Roy, N., Zhang, D., & Cohen, S. M.** (2016). Validation of the cepstral spectral index of dysphonia (CSID) as a screening tool for voice disorders: Development of clinical cutoff scores. *Journal of Voice, 30*(2), 130–144. https://doi.org/10.1016/j.jvoice.2015.04.009

**Boersma, P., & Weenink, D**. (2013). *Praat: Doing phonetics by computer* (Version 5.3.51). http://www.praat.org

**Buckley, D. P., Diaz Cadiz, M., Eadie, T. L., & Stepp, C. E.** (2020). Acoustic model of perceived overall severity of dysphonia in adductor-type laryngeal dystonia. *Journal of Speech, Language, and Hearing Research, 63*(8), 2713–2722. https://doi.org/10.1044/2020_JSLHR-19-00354

**Cohen, S. M.** (2010). Self-reported impact of dysphonia in a primary care population: An epidemiological study. *The Laryngoscope, 120*(10), 2022–2032. https://doi.org/10.1002/lary.21058

**Eadie, T. L., & Stepp, C. E.** (2013). Acoustic correlate of vocal effort in spasmodic dysphonia. *Annals of Otology, Rhinology & Laryngology, 122*(3), 169–176. https://doi.org/10.1177/000348941312200305

**Espinoza, V. M., Zanartu, M., Van Stan, J. H., Mehta, D. D., & Hillman, R. E.** (2017). Glottal aerodynamic measures in women with phonotraumatic and nonphonotraumatic vocal hyperfunction. *Journal of Speech, Language, and Hearing Research, 60*(8), 2159–2169. https://doi.org/10.1044/2017_JSLHR-S-16-0337

**Fukui, N., & Hirose, H.** (1983). Laryngeal adjustments in Danish voiceless obstruent production. *Annual Bulletin Research Institute of Logopedics and Phoniatrics, 17,* 61–71.

**Gareth, J., Daniela, W., Trevor, H., & Robert, T.** (2013). *An introduction to statistical learning: With applications in R*. Springer.

**Halle, M., & Stevens, K. N.** (1971). A note on laryngeal features, IX. Speech communication. In *MIT R.L.E. Quarterly Progress Report, No. 101* (pp. 198–213). Massachusetts Institute of Technology.

**Heller Murray, E. S., Lien, Y. S., Van Stan, J. H., Mehta, D. D., Hillman, R. E., Pieter Noordzij, J., & Stepp, C. E.** (2017). Relative fundamental frequency distinguishes between phonotraumatic and non-phonotraumatic vocal hyperfunction. *Journal of Speech, Language, and Hearing Research, 60*(6), 1507–1515. https://doi.org/10.1044/2016_JSLHR-S-16-0262

**Heman-Ackah, Y. D., Michael, D. D., & Goding, G. S., Jr.** (2002). The relationship between cepstral peak prominence and selected parameters of dysphonia. *Journal of Voice, 16*(1), 20–27. https://doi.org/10.1016/S0892-1997(02)00067-X

**Hillenbrand, J., Cleveland, R. A., & Erickson, R. L.** (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research, 37*(4), 769–778. https://doi.org/10.1044/jshr.3704.769

**Hillenbrand, J., & Houde, R. A.** (1996). Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. *Journal of Speech and Hearing Research, 39*(2), 311–321. https://doi.org/10.1044/jshr.3902.311

**Hillman, R. E., Holmberg, E. B., Perkell, J. S., Walsh, M., & Vaughan, C.** (1989). Objective assessment of vocal

hyperfunction: An experimental framework and initial results. *Journal of Speech and Hearing Research, 32*(2), 373–392. https://doi.org/10.1044/jshr.3202.373

**Hillman, R. E., Holmberg, E. B., Perkell, J. S., Walsh, M., & Vaughan, C.** (1990). Phonatory function associated with hyperfunctionally related vocal fold lesions. *Journal of Voice, 4*(1), 52–63. https://doi.org/10.1016/S0892-1997(05)80082-7

**Hillman, R. E., Stepp, C. E., Van Stan, J. H., Zanartu, M., & Mehta, D. D.** (2020). An updated theoretical framework for vocal hyperfunction. *American Journal of Speech-Language Pathology, 29*(4), 2254–2260. https://doi.org/10.1044/2020_AJSLP-20-00104

**Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X.** (2013). *Applied logistic regression* (Vol. 398). Wiley. https://doi.org/10.1002/9781118548387

**Hunter, E. J., Cantor-Cutiva, L. C., van Leer, E., van Mersbergen, M. R., Nanjundeswaran, C. D., Bottalico, P., Sandage, M. J., & Whitling, S.** (2020). Toward a consensus description of vocal effort, vocal load, vocal loading, and vocal fatigue. *Journal of Speech, Language, and Hearing Research, 63*(2), 509–532. https://doi.org/10.1044/2019_JSLHR-19-00057

**Hunter, E. J., & Titze, I. R.** (2009). Quantifying vocal fatigue recovery: Dynamic vocal recovery trajectories after a vocal loading exercise. *Annals of Otology, Rhinology & Laryngology, 118*(6), 449–460. https://doi.org/10.1177/000348940911800608

**Jaiswal, S.** (2011). *Cricothyroid muscle activity at voicing transitions*. The University of Iowa.

**Jakobsen, J. C., Gluud, C., Wetterslev, J., & Winkel, P.** (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials—A practical guide with flowcharts. *BMC Medical Research Methodology, 17*(1), 1–10. https://doi.org/10.1186/s12874-017-0442-1

**Kempster, G. B., Gerratt, B. R., Abbott, K. V., Barkmeier-Kraemer, J., & Hillman, R. E.** (2009). Consensus auditory-perceptual evaluation of voice: Development of a standardized clinical protocol. *American Journal of Speech-Language Pathology, 18*(2), 124–132. https://doi.org/10.1044/1058-0360(2008/08-0017)

**Khoddami, S. M., Ansari, N. N., & Jalaie, S.** (2015). Review on laryngeal palpation methods in muscle tension dysphonia: Validity and reliability issues. *Journal of Voice, 29*(4), 459–468. https://doi.org/10.1016/j.jvoice.2014.09.023

**Kreiman, J., & Gerratt, B. R.** (1998). Validity of rating scale measures of voice quality. *The Journal of the Acoustical Society of America, 104*(3), 1598–1608. https://doi.org/10.1121/1.424372

**Kreiman, J., Gerratt, B. R., Kempster, G. B., Erman, A., & Berke, G. S.** (1993). Perceptual evaluation of voice quality: Review, tutorial, and a framework for future research. *Journal of Speech and Hearing Research, 36*(1), 21–40. https://doi.org/10.1044/jshr.3601.21

**Kunduk, M., Yan, Y., McWhorter, A. J., & Bless, D.** (2006). Investigation of voice initiation and voice offset characteristics with high-speed digital imaging. *Logopedics Phoniatrics Vocology, 31*(3), 139–144. https://doi.org/10.1080/14015430500364065

**Ladefoged, P.** (1967). *Three areas of experimental phonetics: Stress and respiratory activity, the nature of vowel quality, units in the perception and production of speech*. Oxford University Press.

**Lien, Y. S., Calabrese, C. R., Michener, C. M., Heller Murray, E., Van Stan, J. H., Mehta, D. D., Hillman, R. E., Noordzij, J. P., & Stepp, C. E.** (2015). Voice relative fundamental frequency via neck-skin acceleration in individuals with voice disorders. *Journal of Speech, Language, and Hearing Research, 58*(5), 1482–1487. https://doi.org/10.1044/2015_JSLHR-S-15-0126

**Lien, Y. S., Gattuccio, C. I., & Stepp, C. E.** (2014). Effects of phonetic context on relative fundamental frequency. *Journal of Speech, Language, and Hearing Research, 57*(4), 1259–1267. https://doi.org/10.1044/2014_JSLHR-S-13-0158

**Löfqvist, A., Baer, T., McGarr, N. S., & Story, R. S.** (1989). The cricothyroid muscle in voicing control. *The Journal of the Acoustical Society of America, 85*(3), 1314–1321. https://doi.org/10.1121/1.397462

**Lowell, S. Y., Kelley, R. T., Awan, S. N., Colton, R. H., & Chan, N. H.** (2012). Spectral and cepstral-based acoustic features of dysphonic, strained voice quality. *Annals of Otology, Rhinology & Laryngology, 121*(8), 539–548. https://doi.org/10.1177/000348941212100808

**Ma, E. P., & Yiu, E. M.** (2001). Voice activity and participation profile. *Journal of Speech, Language, and Hearing Research, 44*(3), 511–524. https://doi.org/10.1044/1092-4388(2001/040)

**Mahalingam, S., Boominathan, P., Arunachalam, R., Venkatesh, L., & Srinivas, S.** (2020). Cepstral measures to analyze vocal fatigue in individuals with hyperfunctional voice disorder. *Journal of Voice, 35*(6), 815–821. https://doi.org/10.1016/j.jvoice.2020.02.007

**Maryn, Y., Corthals, P., Van Cauwenberge, P., Roy, N., & De Bodt, M.** (2010). Toward improved ecological validity in the acoustic measurement of overall voice quality: Combining continuous speech and sustained vowels. *Journal of Voice, 24*(5), 540–555. https://doi.org/10.1016/j.jvoice.2008.12.014

**Maryn, Y., De Bodt, M., & Roy, N.** (2010). The acoustic voice quality index: Toward improved treatment outcomes assessment in voice disorders. *Journal of Communication Disorders, 43*(3), 161–174. https://doi.org/10.1016/j.jcomdis.2009.12.004

**Maryn, Y., Roy, N., De Bodt, M., Van Cauwenberge, P., & Corthals, P.** (2009). Acoustic measurement of overall voice quality: A meta-analysis. *The Journal of the Acoustical Society of America, 126*(5), 2619–2634. https://doi.org/10.1121/1.3224706

**MathWorks.** (2016). *MATLAB 2016* (Version b).

**McKenna, V. S., Heller Murray, E. S., Lien, Y. S., & Stepp, C. E.** (2016). The relationship between relative fundamental frequency and a kinematic estimate of laryngeal stiffness in healthy adults. *Journal of Speech, Language, and Hearing Research, 59*(6), 1283–1294. https://doi.org/10.1044/2016_JSLHR-S-15-0406

**McKenna, V. S., & Stepp, C. E.** (2018). The relationship between acoustical and perceptual measures of vocal effort. *The Journal of the Acoustical Society of America, 144*(3), 1643–1658. https://doi.org/10.1121/1.5055234

**Morrison, M. D., & Rammage, L. A.** (1993). Muscle misuse voice disorders: Description and classification. *Acta Oto-Laryngologica, 113*(3), 428–434. https://doi.org/10.3109/00016489309135839

**Murton, O., Hillman, R., & Mehta, D.** (2020). Cepstral peak prominence values for clinical voice evaluation. *American Journal of Speech-Language Pathology, 29*(3), 1596–1607. https://doi.org/10.1044/2020_AJSLP-20-00001

**Nanjundeswaran, C., Jacobson, B. H., Gartner-Schmidt, J., & Verdolini Abbott, K.** (2015). Vocal fatigue index (VFI): Development and validation. *Journal of Voice, 29*(4), 433–440. https://doi.org/10.1016/j.jvoice.2014.09.012

**Naunheim, M. R., & Carroll, T. L.** (2017). Benign vocal fold lesions: Update on nomenclature, cause, diagnosis, and treatment. *Current Opinion in Otolaryngology & Head and Neck Surgery, 25*(6), 453–458. https://doi.org/10.1097/MOO.0000000000000408

**Oates, J. M., & Winkworth, A.** (2008). Current knowledge, controversies and future directions in hyperfunctional voice

disorders. *International Journal of Speech-Language Pathology, 10*(4), 267–277. https://doi.org/10.1080/17549500802140153

**Park, Y., & Stepp, C. E.** (2019). The effects of stress type, vowel identity, baseline f0, and loudness on the relative fundamental frequency of individuals with healthy voices. *Journal of Voice, 33*(5), 603–610. https://doi.org/10.1016/j.jvoice.2018.04.004

**Patel, R. R., Awan, S. N., Barkmeier-Kraemer, J., Courey, M., Deliyski, D., Eadie, T., Paul, D., Švec, J. G., & 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. *American Journal of Speech-Language Pathology, 27*(3), 887–905. https://doi.org/10.1044/2018_AJSLP-17-0009

**Peterson, E. A., Roy, N., Awan, S. N., Merrill, R. M., Banks, R., & Tanner, K.** (2013). Toward validation of the cepstral spectral index of dysphonia (CSID) as an objective treatment outcomes measure. *Journal of Voice, 27*(4), 401–410. https://doi.org/10.1016/j.jvoice.2013.04.002

**Poburka, B. J., Patel, R. R., & Bless, D. M.** (2017). Voice-vibratory assessment with laryngeal imaging (VALI) form: Reliability of rating stroboscopy and high-speed videoendoscopy. *Journal of Voice, 31*(4), 513.e1. https://doi.org/10.1016/j.jvoice.2016.12.003

**Radish Kumar, B., Bhat, J. S., & Prasad, N.** (2010). Cepstral analysis of voice in persons with vocal nodules. *Journal of Voice, 24*(6), 651–653. https://doi.org/10.1016/j.jvoice.2009.07.008

**Roy, N., Fetrow, R. A., Merrill, R. M., & Dromey, C.** (2016). Exploring the clinical utility of relative fundamental frequency as an objective measure of vocal hyperfunction. *Journal of Speech, Language, and Hearing Research, 59*(5), 1002–1017. https://doi.org/10.1044/2016_JSLHR-S-15-0354

**Roy, N., Merrill, R. M., Gray, S. D., & Smith, E. M.** (2005). Voice disorders in the general population: Prevalence, risk factors, and occupational impact. *The Laryngoscope, 115*(11), 1988–1995. https://doi.org/10.1097/01.mlg.0000179174.32345.41

**Ruel, B., & Thibeault, S.** (2020). Is there a relationship between vocal effort and VHI. *Journal of Voice, 34*(5), 809-e11. https://doi.org/10.1016/j.jvoice.2019.03.005

**Sauder, C., Bretl, M., & Eadie, T.** (2017). Predicting voice disorder status from smoothed measures of cepstral peak prominence using Praat and Analysis of Dysphonia in Speech and Voice (ADSV). *Journal of Voice, 31*(5), 557–566. https://doi.org/10.1016/j.jvoice.2017.01.006

**Serry, M. A., Stepp, C. E., & Peterson, S. D.** (2021). Physics of phonation offset: Towards understanding relative fundamental frequency observations. *The Journal of the Acoustical Society of America, 149*(5), 3654–3664. https://doi.org/10.1121/10.0005006

**Solomon, N. P.** (2008). Vocal fatigue and its relation to vocal hyperfunction. *International Journal of Speech-Language Pathology, 10*(4), 254–266. https://doi.org/10.1080/14417040701730990

**Stepp, C. E., Heaton, J. T., Braden, M. N., Jette, M. E., Stadelman-Cohen, T. K., & Hillman, R. E.** (2011). Comparison of neck tension palpation rating systems with surface electromyographic and acoustic measures in vocal hyperfunction. *Journal of Voice, 25*(1), 67–75. https://doi.org/10.1016/j.jvoice.2009.08.001

**Stepp, C. E., Hillman, R. E., & Heaton, J. T.** (2010). The impact of vocal hyperfunction on relative fundamental frequency during voicing offset and onset. *Journal of Speech, Language, and Hearing Research, 53*(5), 1220–1226. https://doi.org/10.1044/1092-4388(2010/09-0234)

**Stepp, C. E., Merchant, G. R., Heaton, J. T., & Hillman, R. E.** (2011). Effects of voice therapy on relative fundamental frequency during voicing offset and onset in patients with vocal hyperfunction. *Journal of Speech, Language, and Hearing Research, 54*(5), 1260–1266. https://doi.org/10.1044/1092-4388(2011/10-0274)

**Stepp, C. E., Sawin, D. E., & Eadie, T. L.** (2012). The relationship between perception of vocal effort and relative fundamental frequency during voicing offset and onset. *Journal of Speech, Language, and Hearing Research, 55*(6), 1887–1896. https://doi.org/10.1044/1092-4388(2012/11-0294)

**Stevens, K. N.** (1977). Physics of laryngeal behavior and larynx modes. *Phonetica, 34*(4), 264–279. https://doi.org/10.1159/000259885

**Titze, I. R.** (1989). On the relation between subglottal pressure and fundamental frequency in phonation. *The Journal of the Acoustical Society of America, 85*(2), 901–906. https://doi.org/10.1121/1.397562

**Van Houtte, E., Van Lierde, K., & Claeys, S.** (2011). Pathophysiology and treatment of muscle tension dysphonia: A review of the current knowledge. *Journal of Voice, 25*(2), 202–207. https://doi.org/10.1016/j.jvoice.2009.10.009

**Van Stan, J. H., Mehta, D. D., Ortiz, A. J., Burns, J. A., Marks, K. L., Toles, L. E., Stadelman-Cohen, T., Krusemark, C., Muise, J., Hron, T., Zeitels, S. M., Fox, A. B., & Hron, T.** (2020). Changes in a daily phonotrauma index after laryngeal surgery and voice therapy: Implications for the role of daily voice use in the etiology and pathophysiology of phonotraumatic vocal hyperfunction. *Journal of Speech, Language, and Hearing Research, 63*(12), 3934–3944. https://doi.org/10.1044/2020_JSLHR-20-00168

**Van Stan, J. H., Mehta, D. D., Ortiz, A. J., Burns, J. A., Toles, L. E., Marks, K. L., Vangel, M., Hron, T., Zeitels, S., & Hillman, R. E.** (2020). Differences in weeklong ambulatory vocal behavior between female patients with phonotraumatic lesions and matched controls. *Journal of Speech, Language, and Hearing Research, 63*(2), 372–384. https://doi.org/10.1044/2019_JSLHR-19-00065

**Van Stan, J. H., Ortiz, A. J., Cortes, J. P., Marks, K. L., Toles, L. E., Mehta, D. D., Burns, J. A., Hron, T., Stadelman-Cohen, T., Krusemark, C., Muise, J., Fox-Galalis, A. B., Nudelman, C., Zeitels, S., & Hillman, R. E.** (2021). Differences in daily voice use measures between female patients with nonphonotraumatic vocal hyperfunction and matched controls. *Journal of Speech, Language, and Hearing Research, 64*(5), 1457–1470. https://doi.org/10.1044/2021_JSLHR-20-00538

**Van Stan, J. H., Ortiz, A. J., Marks, K. L., Toles, L. E., Mehta, D., Burns, J. A., Hron, T., Stadelman-Cohen, T., Krusemark, C., Muise, J., Fox, A. B., Nudelman, C., Zeitels, S., & Hillman, R. E.** (2021). Changes in the Daily Phonotrauma Index (DPI) following the use of voice therapy as the sole treatment for phonotraumatic vocal hyperfunction in females. *Journal of Speech, Language, and Hearing Research, 64*(9), 3446–3455. https://doi.org/10.1044/2021_JSLHR-21-00082

**Verdolini, K., Rosen, C. A., & Branski, R. C.** (2014). *Classification manual for voice disorders-I*. Psychology Press. https://doi.org/10.4324/9781410617293

**Vojtech, J. M., Segina, R. K., Buckley, D. P., Kolin, K. R., Tardif, M. C., Noordzij, J. P., & Stepp, C. E.** (2019). Refining algorithmic estimation of relative fundamental frequency: Accounting for sample characteristics and fundamental frequency estimation method. *The Journal of the Acoustical Society of America, 146*(5), 3184–3202. https://doi.org/10.1121/1.5131025

**von Latoszek, B. B., Maryn, Y., Gerrits, E., & De Bodt, M.** (2017). The acoustic breathiness index (ABI): A multivariate

acoustic model for breathiness. *Journal of Voice, 31*(4), 511-e11. https://doi.org/10.1016/j.jvoice.2016.11.017

**von Latoszek, B. B., Maryn, Y., Gerrits, E., & De Bodt, M.** (2018). A meta-analysis: Acoustic measurement of roughness and breathiness. *Journal of Speech, Language, and Hearing Research, 61*(2), 298–323. https://doi.org/10.1044/2017_JSLHR-S-16-0188

**Watson, B. C.** (1998). Fundamental frequency during phonetically governed devoicing in normal young and aged speakers. *The Journal of the Acoustical Society of America, 103*(6), 3642–3647. https://doi.org/10.1121/1.423068

**Whitling, S., Rydell, R., & Åhlander, V. L.** (2015). Design of a clinical vocal loading test with long-time measurement of voice. *Journal of Voice, 29*(2), 261-e3. https://doi.org/10.1016/j.jvoice.2014.07.012

**Wolfe, V. I., & Martin, D.** (1997). Acoustic correlates of dysphonia: Type and severity. *Journal of Communication Disorders, 30*(5), 403–416. https://doi.org/10.1016/S0021-9924(96)00112-8

**Wolfe, V. I., Martin, D. P., & Palmer, C. I.** (2000). Perception of dysphonic voice quality by naive listeners. *Journal of Speech, Language, and Hearing Research, 43*(3), 697–705. https://doi.org/10.1044/jslhr.4303.697

**Xue, C., Kang, J., Hedberg, C., Zhang, Y., & Jiang, J. J.** (2019). Dynamically monitoring vocal fatigue and recovery using aerodynamic, acoustic, and subjective self-rating measurements. *Journal of Voice, 33*(5), 809-e11. https://doi.org/10.1016/j.jvoice.2018.03.014

**Yiu, E. M., Lau, V. C., Ma, E. P., Chan, K. M., & Barrett, E.** (2014). Reliability of laryngostroboscopic evaluation on lesion size and glottal configuration: A revisit. *The Laryngoscope, 124*(7), 1638–1644. https://doi.org/10.1002/lary.24521