JSLHR

Research Article

Differences in Daily Voice Use Measures Between Female Patients With Nonphonotraumatic Vocal Hyperfunction and Matched Controls

Jarrad H. Van Stan,[a,b,c] Andrew J. Ortiz,[a] Juan P. Cortes,[a,b] Katherine L. Marks,[a,c] Laura E. Toles,[a,c] Daryush D. Mehta,[a,b,c] James A. Burns,[a,b] Tiffiny Hron,[a,b] Tara Stadelman-Cohen,[a,c] Carol Krusemark,[a,c] Jason Muise,[a,c] Annie B. Fox-Galalis,[c] Charles Nudelman,[a,c] Steven Zeitels,[a,b] and Robert E. Hillman[a,b,c]

Purpose: The purpose of this study was to obtain a more comprehensive understanding of the pathophysiology and impact on daily voice use of nonphonotraumatic vocal hyperfunction (NPVH).
Method: An ambulatory voice monitor collected 1 week of data from 36 patients with NPVH and 36 vocally healthy matched controls. A subset of 11 patients with NPVH were monitored after voice therapy. Daily voice use measures included neck-skin acceleration magnitude, fundamental frequency (f_o), cepstral peak prominence (CPP), and the difference between the first and second harmonic magnitudes (H1–H2). Additional comparisons included 118 patients with phonotraumatic vocal hyperfunction (PVH) and 89 additional vocally healthy controls.
Results: The NPVH group, compared to the matched control group, exhibited increased f_o (Cohen's d = 0.6), reduced CPP (d = –0.9), and less positive H1–H2 skewness (d = –1.1). Classifiers used CPP mean and H1–H2 mode to maximally differentiate the NPVH and matched control groups (area under the receiver operating characteristic curve of 0.78). Classifiers performed well on unseen data: the logit decreased in patients with NPVH after therapy; ≥ 85% of the control and PVH groups were identified as "normal" or "not NPVH," respectively.
Conclusions: The NPVH group's daily voice use is less periodic (CPP), is higher pitched (f_o), and has less abrupt vocal fold closure (H1–H2 skew) compared to the matched control group. The combination of CPP mean and H1–H2 mode appears to reflect a pathophysiological continuum in NPVH patients of inefficient phonation with minimal potential for phonotrauma. Further validation of the classification model is needed to better understand potential clinical uses.
Supplemental Material: https://doi.org/10.23641/asha.14390771

Nonphonotraumatic vocal hyperfunction (NPVH; Hillman et al., 2020)—also referred to as primary muscle tension dysphonia or functional dysphonia (Verdolini et al., 2006)—is one of the most common and highly variable voice disorders treated by laryngologists and speech-language pathologists (Coyle et al., 2001; Herrington-Hall et al., 1988; Kridgen et al., 2020). NPVH is characterized by myriad habitual, chronic voice-related symptoms in daily life—for example, dysphonia (Altman et al., 2005; Dworkin et al., 2000; Van Houtte et al., 2011), increased vocal fatigue/effort (Koufman & Blalock, 1988; Solomon, 2008), and anterior neck soreness/excess muscle activation (Mathieson et al., 2009; Roy et al., 1997)—in the absence of any signs of phonotrauma or other phonation-disrupting structural or neurological impairments. There are also multiple causative factors (Kridgen et al., 2020; besides or in addition to voice use) that have been associated with the onset of NPVH such as stress reactivity (Dietrich et al., 2008; Helou et al., 2013), psychological predispositions (Misono et al., 2016; Roy et al., 2000a, 2000b; Van Mersbergen et al.,

[a]Massachusetts General Hospital, Boston
[b]Harvard Medical School, Boston, MA
[c]MGH Institute of Health Professions, Boston, MA
Correspondence to Jarrad H. Van Stan: jvanstan@mgh.harvard.edu

Editor-in-Chief: Bharath Chandrasekaran
Editor: Jack J. Jiang
Received September 11, 2020
Revision received December 3, 2020
Accepted January 8, 2021
https://doi.org/10.1044/2021_JSLHR-20-00538

Disclosure: *Robert Hillman and Steven Zeitels have a financial interest in InnoVoyce LLC, a company focused on developing and commercializing technologies for the prevention, diagnosis, and treatment of voice-related disorders. Hillman's and Zeitels' interests were reviewed and are managed by Massachusetts General Hospital and Partners HealthCare in accordance with their conflict of interest policies. All other authors have declared that no other competing interests existed at the time of publication.*

2008), and external irritants to the upper airway like a respiratory infection or laryngo-pharyngeal reflux (Koufman et al., 2000; Ross et al., 1998; Roy et al., 2005). Furthermore, the vocal presentation of patients with NPVH is highly variable in regard to overall severity, combination of auditory-perceptual characteristics (e.g., pitch breaks, overall dysphonia, strained voice quality, pulse registration), and the consistency of an individual patient's voice production over time (e.g., constant level of severity, quickly varying within a sentence, and possibly slowly varying based on vocal demands across an entire day or week; Fernández et al., 2020; Gillespie et al., 2013; Solomon, 2008; Spencer, 2015).

Despite the large amount of variability in NPVH, studies using short-duration, in-laboratory recordings have found that groups of these patients produce abnormal acoustic measures such as sound pressure level (SPL; e.g., reduced mean and range, increased shimmer), fundamental frequency (f_o; e.g., increased mode or mean, decreased standard deviation, increased jitter), and periodicity (e.g., increased noise-to-harmonic ratio, decreased cepstral peak prominence [CPP]); for example of studies, see Awan and Roy (2005); Carding et al. (1999); Cooper (1974); Koufman and Blalock (1988); Mathieson et al. (2009); Nguyen and Kenny (2009); Rattenbury et al. (2004); Roy and Hendarto (2005); Roy and Leeper (1993); and Van Lierde et al. (2010, 2004). Lab-based aerodynamic studies have also shown that patients with NPVH phonate with increased subglottal pressure and open quotient to produce a desired vocal intensity compared to normal control subjects (Espinoza et al., 2020, 2017; Hillman et al., 1989). Taken together, the results from these lab-based assessments support the view that the variable presentations of patients with NPVH are associated with an underlying pathophysiology that causes inefficient daily voice use but does not result in true vocal fold tissue trauma (Hillman et al., 2020). The only curative option for these patients is voice therapy, which focuses on behaviorally improving the patient's voice use (e.g., increasing vocal efficiency; Altman et al., 2005; Dworkin et al., 2000; Van Houtte et al., 2011) with hopes that such improvements will carry over into the patient's daily life (Ziegler et al., 2014).

Although information about the impact of NPVH on daily voice use is thought to be critical in the diagnosis and treatment of NPVH, to date, only one study has used long-duration ambulatory voice monitoring to investigate differences between a group of 20 patients with NPVH and matched healthy controls (Mehta et al., 2015). Somewhat surprisingly, this study found no differences in weeklong voice use between the patient and control groups for traditional acoustic measures previously shown to be different based on brief laboratory assessments (i.e., SPL, f_o, and CPP). The failure of average univariate ambulatory voice measures to differentiate NPVH from normal daily voice use may be due to the large variability in voice production (e.g., type and severity of dysphonia) that these patients display, and because such measures may not adequately reflect salient pathophysiological mechanisms. If any singular quantitative measure could be broadly associated across most patients with NPVH, it would need to be strongly and independently connected to inefficient voicing with minimal risk of phonotrauma. Theoretically, one could voice inefficiently with minimal phonotrauma at any intensity, f_o, or degree of periodicity, weakening the potential connection between any of these individual metrics and the NPVH pathophysiology. This outcome is similar to multiple investigations of phonotraumatic vocal hyperfunction (PVH) that also failed to find differences between patients and controls based on average univariate ambulatory voice measures (Mehta et al., 2015; Szabo Portela et al., 2018; Van Stan et al., 2015). Subsequent ambulatory studies of PVH did find salient differences between patients and controls by using distributional characteristics of ambulatory voice use (i.e., skew and kurtosis), measures indicative of glottal closure (i.e., statistical measures based on the difference between the first and second harmonic magnitudes [H1–H2]), and the development of a multidimensional data-driven classifier (Cortés et al., 2018; Van Stan, Mehta, Ortiz, Burns, Marks, et al., 2020; Van Stan, Mehta, Ortiz, Burns, Toles, et al., 2020). It is reasonable to assume that similar approaches could help to better characterize (differentiate from normal) the phonatory pathophysiology of NPVH based on ambulatory voice data.

Vocal fold closure dynamics are thought to be a critical determinant of the collision and shearing forces on vocal fold tissue during phonation, and phonotrauma is unlikely to occur if vocal fold closure is less complete or abrupt than normal (Berry et al., 2001; Gunter et al., 2005; Hillman et al., 1989). Estimates of H1–H2 that reflect increased abruptness of vocal fold closure have already shown strong, independent relationships with the pathophysiology of PVH (Cortés et al., 2018; Van Stan, Mehta, Ortiz, Burns, Marks, et al., 2020; Van Stan, Mehta, Ortiz, Burns, Toles, et al., 2020). Thus, H1–H2 also has great potential to independently discriminate between groups of patients with NPVH and their matched controls based on evidence that the absence of phonotrauma in NPVH patients is associated with less abrupt and/or incomplete vocal fold closure (Espinoza et al., 2020, 2017; Hillman et al., 1989). Specifically, H1–H2 is the difference between the first and second harmonic magnitudes during voicing, and higher/lower H1–H2 values (including those extracted from neck accelerometer recordings) represent less/more abrupt vocal fold closure, respectively (Klatt & Klatt, 1990; Mehta et al., 2019; Stevens, 1998). The hypothesis would be that patients with NPVH produce higher H1–H2 (less abrupt closure) values more often than vocally healthy controls, even when there may be no differences in vocal intensity, f_o, CPP, and vocal doses. In other words, the patients' nonphonotraumatic inefficiency would be using less vocal fold closure to achieve the same (or nearly the same) vocal output.

As patients with NPVH are thought to voice inefficiently with minimal potential for phonotrauma, it is expected that they would produce lower-than-typical levels of overall laryngeal force during phonation (Hillman et al., 2020). Traditionally, the amount of laryngeal force during ambulatory monitoring is estimated by transforming the neck-surface acceleration magnitude (NSAM) into an estimate

of SPL (Švec et al., 2005). However, this article keeps the NSAM in physical units of vibration (cm/s^2) because the SPL calibration procedure has a known estimation error as high as ± 5–6 dB, which could introduce more uncertainty (noise) into the model (Švec et al., 2005). The NSAM could be viewed as generally representing the magnitude of laryngeal forces associated with phonation. This interpretation is based on evidence that the NSAM is correlated with low-bandwidth (subglottal pressure; Fryd et al., 2016) and high-bandwidth (peak-to-peak glottal airflow, maximum flow declination rate; Zañartu et al., 2013, 2012) aero-acoustic parameters, as well as mechanical forces generated by the tissue-to-tissue contact associated with vocal fold vibration (Coleman, 1988; Wokurek & Pützer, 2009, 2011, 2013). Thus, the NSAM probably represents a combination of these aero-acoustic and mechanical forces. However, it is currently not possible to determine the relative contribution of each force to the NSAM as these are hypothetically expected to vary according to how the subject is voicing, for example, at different f_o, intensities, level of periodicity, and amount of vocal fold contact (Jiang & Titze, 1994).

The main purpose of this study was to obtain a more comprehensive understanding of the pathophysiology and impact on daily voice use of NPVH through pursuing two aims: (a) determine if there are any significant one-dimensional differences in daily voice use by reexamining previously used ambulatory voice measures (NSAM, f_o, CPP) and a novel measure associated with vocal fold closure patterns (H1–H2) in a larger group of patients with NPVH and matched controls than previously studied with ambulatory monitoring and (b) develop and test a data-driven model capable of reflecting the multidimensional continuum of inefficient vocal behaviors that are not phonotraumatic. Weeklong ambulatory phonation data were acquired using a smartphone-based ambulatory voice monitor (Mehta et al., 2012) in groups of patients with NPVH and age-, sex-, and occupation-matched controls. All data were collected as part of a larger, ongoing project aimed at attaining a better understanding of the etiology and pathophysiology of hyperfunctional voice disorders. The governing institutional review board approved all experimental aspects related to the use of human subjects for this study.

Method

Participants

A total of 279 female subjects were consented for participation in this study. Only female participants were selected to be in this study to provide a homogenous sample of a group that has a significantly higher incidence of vocal hyperfunction (Goldman et al., 1996; Herrington-Hall et al., 1988). The patient groups were recruited through sequential convenience sampling. The NPVH patient group consisted of 36 females with a diagnosis of NPVH and was used (in combination with their matched-control group) to train classification models. One hundred eighteen females with PVH were used to test the resulting classification models, and the PVH group demographics can be found in a previous paper (Van Stan, Mehta, Ortiz, Burns, Toles, et al., 2020). Testing the classification models on a group of patients with PVH is especially meaningful, as it is believed there are two primary types of vocal hyperfunction: one with no signs of phonotrauma (NPVH) and one with signs of phonotrauma (PVH). If the classification models have found a multidimensional space that has any chance of representing vocal hyperfunction with minimal risk of phonotrauma (i.e., "NPVH"), the PVH patients should rarely be found in those areas associated with NPVH (i.e., classified as "not NPVH"). A subset of the NPVH patients (n = 11) was monitored after completing voice therapy and used to test the classification models. Of note, all patients with NPVH were offered voice therapy, but only these 11 completed a full course of treatment. Diagnoses were based on a comprehensive team evaluation (laryngologist and speech-language pathologist) at the Center for Laryngeal Surgery and Voice Rehabilitation at Massachusetts General Hospital (MGH Voice Center) that included the collection of a complete case history and videostroboscopic imaging of the larynx. Each patient's vocal function was further characterized by completion of the Voice-Related Quality of Life (V-RQOL) questionnaire (Hogikyan & Sethuraman, 1999), an auditory-perceptual evaluation using the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V; Kempster et al., 2009), and aerodynamic and acoustic assessments of vocal function (Patel et al., 2018). All participants were engaged in occupations considered to be at a higher-than-normal risk for developing a voice disorder (Ramig & Verdolini, 1998). Specifically, the NPVH group consisted of 11 health care professionals, seven office workers, five teachers, four full-time parents, four administrators, three college students, one entertainer, and one public relation representative. Of note, singers were prospectively excluded from recruitment in the NPVH population as it can be unclear to what degree singing is responsible for the patient's voice-related symptoms. The average (standard deviation) age of the NPVH group was 46 (12) years.

Thirty-six subjects with healthy voices were recruited to serve as matched controls to the NPVH group and were used to train classification models. Eighty-nine subjects with healthy voices (not matched to the patients with NPVH) were used as a test set for the trained classification models; that is, can the classification models correctly label most of these subjects as normal? These 89 subjects with healthy voices were recruited as controls for patients with PVH, and their demographics have been reported in a previous paper (Van Stan, Mehta, Ortiz, Burns, Toles, et al., 2020). All control subjects were recruited through snowball sampling. The snowball sampling approach asked patients enrolled in the study to recommend a colleague with no history of voice disorders, approximately the same age (± 5 years), the same sex, and in a high voice-use occupation. The normal vocal status of all participants in the control groups was verified via interview and a laryngeal stroboscopic examination. During the interview, the control candidates were specifically asked if they had any voice difficulties that affected their daily life, and a speech-language pathologist evaluated

the auditory-perceptual quality of their voices. If the control candidate indicated voice difficulties or demonstrated a nonnormal voice quality, they were excluded from study enrollment and did not undergo a laryngeal stroboscopic examination. Due to the matching paradigm, the NPVH control group's occupations and ages were the same as the NPVH patient group. The total number of patients with NPVH and their matched controls (n = 36 pairs) resulted from a convenience sample that attained enough power (beta ≥ 0.8) to find medium-to-large differences—that is, Cohen's $d \geq 0.5$ (Cohen, 1988).

Table 1 reports subscale scores for the self-reported V-RQOL and clinician-judged CAPE-V ratings for the participants in the NPVH group before and after voice therapy. There are missing data points underlying the mean ± standard deviation reported in the table since these measures were extracted from a clinical database during the course of standard care (not all measures were taken before/after voice therapy on all patients). Specifically, the estimates at each time point are composed of the following number of patients: before treatment (V-RQOL based on 34 patients; CAPE-V based on all 36 patients) and after therapy (V-RQOL based on eight patients; CAPE-V based on all 11 patients). These subjective scales are reported only for the purpose of generally describing the severity level of the patient group, not for statistical analysis or results reporting. Therefore, reliability was not addressed. V-RQOL scores are normalized ordinal ratings that lie between 0 and 100, with higher scores indicating a higher voice-related quality of life. CAPE-V scores are visual analog scale ratings that range from 0 to 100, with 0 indicating normality and 100 indicating the most extreme example of deviance for a particular voice quality characteristic. The CAPE-V measurement for each patient came from one rater—a speech-language pathologist's single rating during a routine clinical evaluation using the CAPE-V standard reading and sustained vowel samples. Both subjective scales qualitatively indicate that, compared to before treatment, the voice quality and voice-related quality of life for the NPVH group improved after therapy.

Table 1. Patients' self-reported quality of life impact due to their voice disorder using the Voice-Related Quality of Life (V-RQOL) subscales, and the perceived qualities of their voice as judged by a speech-language pathologist using the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) form.

Subjective measures	Before therapy	After therapy
V-RQOL		
Social–Emotional	70.1 ± 28.3	93.0 ± 9.1
Physical Functioning	57.5 ± 25.5	87.6 ± 12.7
Total Score	62.8 ± 24.7	89.8 ± 9.1
CAPE-V		
Overall Severity	42.5 ± 28.7	10.1 ± 7.0
Roughness	27.2 ± 21.3	9.2 ± 8.0
Breathiness	19.3 ± 26.8	4.8 ± 4.6
Strain	33.8 ± 26.9	6.5 ± 6.0

Note. Mean ± standard deviation reported before therapy and after therapy.

Data Collection

The Voice Health Monitor (VHM; Mehta et al., 2012) was used to collect ambulatory voice data on all subjects in the study. The VHM employs a miniature accelerometer (Model BU-27135, Knowles Electronics) attached via doubled-sided medical grade tape to the anterior neck (below the larynx and above the sternal notch) to sense phonation. The sensor is connected to a custom smartphone application as the data acquisition platform, and the system records the unprocessed acceleration signal at 11025-Hz sampling rate, 16-bit quantization, and 80-dB dynamic range to obtain frequency content of neck-surface vibrations up to 5 kHz.

Participants in the NPVH and PVH groups were monitored for 1 week (7 days) before any treatment. A subset of the NPVH group (n = 11) was monitored for 1 week after completing voice therapy. Each control participant was monitored for 1 full week. Each morning, the VHM application led the participants through a daily process to calibrate the accelerometer (ACC) signal level to acoustic SPL recorded by a handheld microphone (H1 Handy Recorder, Zoom Corporation) positioned 15 cm from the lips (Švec et al., 2005). For the acoustic SPL calibration, the participant is asked to glide from soft to loud on an /a/ and is trained to perform the loudness glide during their initial study appointment by study staff. To also improve the quality of the loudness glide, three glides are elicited from the subjects every morning and the best glide (largest intensity range and most linear mapping between the neck skin and acoustic signal) is used. The most detailed description of the acoustic SPL calibration is included in a previous publication (Mehta et al., 2012). During the calibration procedure, participants take a picture of their neck to document the day-to-day placement of ACC. Participants were also taught to contact study staff if the sensor fell off their neck or loosened throughout the day. If the ACC sensor were misplaced (as evidenced by the daily photos) or the participants reported issues with the sensor coming off, those days of data were not included in the analysis. Of note, these type of sensor issues occurred very rarely, that is, less than 5% of days.

Data Analysis

The weeklong neck-skin acceleration recordings were divided into nonoverlapping frames of 50 ms in duration. As was done in previous studies (Mehta et al., 2015; Van Stan, Mehta, Ortiz, Burns, Marks, et al., 2020; Van Stan, Mehta, Ortiz, Burns, Toles, et al., 2020; Van Stan et al., 2015), each frame was considered voiced if it passed the following thresholds: (a) SPL was greater than 45 dB SPL at 15 cm, (b) the first non–zero-lag peak in the normalized autocorrelation exceeded a threshold of 0.6, (c) f_o (reciprocal of the time lag of the first non-zero autocorrelation peak) was between 70 and 1000 Hz, and (d) the ratio of low- to high-frequency energy exceeded 22 dB. These criteria were needed to eliminate several types of nonphonatory activity such as tapping or rubbing on the sensor, extremely high levels of environmental noise (e.g., rock concert), and electrical interference/artifacts.

To calculate NSAM, the root-mean-square of each 50-ms frame was transformed into physical units of cm/s^2 according to the linear mapping obtained from a calibration procedure for the specific miniature accelerometer used during recording. The calibration procedure was completed once for each accelerometer, where the accelerometer was calibrated to a reference accelerometer (4533-B, Brüel & Kjær, Denmark) by applying a known chirp vibration signal covering the 10- to 5000-Hz spectrum using an electrodynamic vibration exciter (Mini-Shaker Type 4810, Brüel & Kjær) on a vibration isolation table (BT-2024, Newport Corp.). This calibration procedure was developed and validated in previous work (Cheyne, 2002; Cheyne et al., 2003).

CPP and H1–H2 were two additional features calculated on each analysis frame. To calculate CPP, each 50-ms frame underwent two discrete Fourier transforms that were computed in succession with a logarithmic transformation between them. A regression line was then computed over quefrencies greater than 2 ms (corresponding to a quefrency range minimally affected by subglottal resonances). Finally, the CPP for each frame was defined as the difference, in dB, between the magnitude of the highest peak and the baseline regression level in the power cepstrum. The peak search was limited to quefrencies between 2.5 and 12 ms, corresponding to frequencies of 417 and 83 Hz, respectively. To calculate H1–H2, each 50-ms frame underwent one discrete Fourier transform. The H1–H2 for each frame was defined as the difference, in dB, between the magnitudes of the first and second harmonics in the frequency spectrum.

Statistical Analysis

Statistics computed were mean (NSAM and CPP), mode (f_o and H1–H2), standard deviation (*SD*), minimum (5th percentile), maximum (95th percentile), range (middle 90%), skewness, and kurtosis. In the data presented here, NSAM and CPP distributions tended to be normal (similar mean, median, mode), whereas f_o and H1–H2 distributions were often skewed toward lower values with a long, thin tail toward higher values. The f_o and H1–H2 modes were computed from histograms containing 30 equally spaced bins. Unidimensional differences between the two groups (NPVH and matched controls) were evaluated using summary statistics computed from daylong distributions—instead of weeklong estimates of voice use, as was done in previous studies (Mehta et al., 2015; Van Stan, Mehta, Ortiz, Burns, Toles, et al., 2020; Van Stan et al., 2015). Daily summary statistics (unlike weekly summary statistics) preserved individual subject variability at a daily level, and linear mixed-effects regression models provided adjusted group means that take this daily variability into account. The analyses use daylong distributions instead of time periods shorter than 1 day because (a) the purpose of this work is to identify habitual, long-term voice use differences between patients with NPVH and controls; (b) a day is the longest continuous duration of an individual recording; and (c) shorter durations than the daylong recordings would require some theory-driven splitting of the daily recording or data-driven analysis at multiple time scales. Linear mixed-effects regression models were used to analyze the results across two paired observations (0 = NPVH control group, 1 = pretreatment NPVH patient group). To take full advantage of the matched patient–control study design, each control subject was used as the "0" observation for their matched patient with NPVH. The models assessed differences among the regression values of the NPVH patient group and the normative group. Instead of relying on *p* values or a Bonferroni correction, statistical significance was based on effect sizes representing clinically meaningful differences instead of the traditional alpha value of 0.05. As in previous investigations (Van Stan, Mehta, Ortiz, Burns, Marks, et al., 2020; Van Stan, Mehta, Ortiz, Burns, Toles, et al., 2020; Van Stan et al., 2015), clinically meaningful differences were considered to be medium-to-large effect sizes (Cohen's $|d| \geq 0.5$; Cohen, 1988). All statistics were completed using R 3.5.0 (R Core Team, 2018). The model-derived means and effect sizes were calculated using the *emmeans* packages (Version 1.4.4; Lenth et al., 2020).

A linear and a nonlinear data-driven classification models—logistic regression and quadratic discriminant analysis (QDA), respectively—were trained using all features in a stepwise, forward, conditional approach to minimize the total number of features and feature redundancy (i.e., minimal correlation between final variables). QDA was used instead of other traditional nonlinear classification techniques (e.g., Source Vector Machine) because it permits the output of probabilities and logits, which could be clinically interpreted. This data-driven approach was adopted to develop a model that could find the most discriminative multidimensional relationship with the fewest variables possible. If the approach were to only use features that were individually significant, this could prevent identifying a stronger relationship that combined one strong and one weak predictor. Specifically, the logistic regression and QDA will attempt to use the daily statistical estimates of NSAM, f_o, CPP, and H1–H2 to classify a subject's day on a probability scale of 0–1. On this scale, data are considered to be from a patient with probabilities ≥ 0.5 and from a control with probabilities < 0.5. After training the models on the 36 patient–control pairs, they were tested using three unseen ambulatory data sets with associated a priori hypotheses: (a) a subgroup of 11 patients with posttherapy data will produce lower logits on average than their pretherapy data, (b) a group of 89 female subjects without history of voice disorders and endoscopically verified normal laryngeal anatomy will be mostly classified as "normal," and (c) a group of 118 female patients with PVH will be mostly classified as "not NPVH" (i.e., "normal").

Results

Most subjects wore the monitoring system for more than 80 hr during the 7 days. Table 2 displays all summary statistics for voiced features (NSAM, f_o, CPP, H1–H2) that were compared between the NPVH and matched control groups. Nine measures were significantly different between

Table 2. Group-based adjusted means (standard errors) from the multilevel models across daily summary statistics of ambulatory estimates of neck-skin acceleration magnitude (NSAM), fundamental frequency (f_o), cepstral peak prominence (CPP), and H1–H2 measures collected from the patient and matched control groups (n = 36 pairs).

Voice use summary statistic	Patient group	Control group	Cohen's *d*
Monitored duration (hh:mm)	77:49 (14:04)	85:05 (11:18)	
Voicing time (%)	6.3 (0.4)	7.1 (0.4)	
NSAM (dB re cm/s^2)			
M	46.0 (0.7)	46.9 (0.7)	
SD	6.02 (0.1)	5.6 (0.1)	
5th percentile	36.0 (0.7)	37.2 (0.7)	
95th percentile	55.8 (0.8)	55.8 (0.8)	
Range	19.8 (0.3)	18.6 (0.3)	
Skewness	–0.03 (0.04)	–0.18 (0.04)	
Kurtosis	3.22 (0.06)	3.11 (0.06)	
f_o (Hz)			
Mode	202.4 (3.5)	182.8 (3.5)	0.64
SD	68.1 (2.5)	68.6 (2.5)	
5th percentile	163.6 (2.7)	153.2 (2.7)	0.61
95th percentile	367.2 (9.2)	358.4 (9.1)	
Range	203.6 (8.4)	205.1 (8.4)	
Skewness	2.08 (0.09)	2.20 (0.09)	
Kurtosis	12.38 (0.88)	10.97 (0.88)	
CPP (dB)			
M	20.6 (0.2)	22.1 (0.2)	–1.08
SD	3.9 (0.1)	4.3 (0.1)	–0.71
5th percentile	14.1 (0.1)	14.7 (0.1)	–0.95
95th percentile	26.74 (0.3)	28.5 (0.3)	–1.00
Range	12.7 (0.2)	13.8 (0.2)	–0.81
Skewness	–0.03 (0.04)	–0.24 (0.04)	–0.80
Kurtosis	2.62 (0.04)	2.45 (0.04)	
H1–H2 (dB)			
M	2.6 (0.5)	2.5 (0.5)	
SD	6.7 (0.1)	6.5 (0.1)	
5th percentile	–4.1 (0.4)	–3.3 (0.4)	
95th percentile	17.7 (0.4)	17.8 (0.4)	
Range	21.7 (0.4)	21.1 (0.4)	
Skewness	0.52 (0.05)	0.79 (0.05)	–1.08
Kurtosis	3.82 (0.1)	3.83 (0.1)	

Note. Comparisons reaching a medium-to-large difference (Cohen's $|d| \geq 0.5$) have *d* effect sizes listed. Directionality of effect sizes are derived from the pairwise comparison of each summary statistic for patient values minus their matched control values.

the two groups ($|d| \geq 0.5$): f_o (mode, 5th percentile), CPP (mean, standard deviation, 5th percentile, 95th percentile, range, skewness), and H1–H2 (skewness). Specifically, patients with NPVH exhibited significantly higher f_o mode and fifth percentile (d = 0.64 and 0.61, respectively), lower CPP values and overall variability (d = −0.71 to −1.08), more positively skewed CPP (d = −0.80), and more negatively skewed H1–H2 (d = −1.08) compared to their matched controls.

Both the logistic regression and QDA found that only two features were significant contributors to classification based on the training data of 36 patient–control pairs: CPP mean and H1–H2 mode. The logistic regression used CPP mean (standardized beta = −0.78, $p < .001$, odds ratio = 0.46) and H1–H2 mode (standardized beta = −0.23, $p = .009$, odds ratio = 0.79) to achieve an overall classification of 69.4 % accuracy; true positives and true negatives (n = 25), false positives and false negatives (n = 11), an area under the receiver operating characteristic curve (AUC) = 0.78, and positive and negative likelihood ratios = 2.23 and 0.45, respectively. The probability (p) of a subject's data being classified as coming from a patient or not results from a logistic transformation of the patient's daily CPP mean (C) and H1–H2 mode (H), represented in Equation 1:

$$p = \frac{1}{1 + e^{-(-0.781C - 0.230H + 17.184)}}. \quad (1)$$

Assuming that the feature space of NPVH subjects and controls each are drawn from a multivariate Gaussian distribution $p\ (C = c_1 | \boldsymbol{x}, \theta)$ with their own means and covariances, the output of the QDA for each subject can be transformed into a probability (p) estimate, where the probability of belonging to the "patient" class (c_1) out of both classes (C) given the QDA model (θ) is represented in Equation 2:

$$p(C = c_1|x,\theta) = \frac{\pi_{c_1}|2\pi\sum_{c_1}|^{-\frac{1}{2}} e^{\left(-\frac{1}{2}(x-\mu_{c_1})\sum_{c_1}^{-1}(x-\mu_{c_1})^T\right)}}{\sum_{c_i}^{C}\pi_{c_i}|2\pi\sum_{c_i}|^{-\frac{1}{2}} e^{\left(-\frac{1}{2}(x-\mu_{c_i})\sum_{c_i}^{-1}(x-\mu_{c_i})^T\right)}}. \quad (2)$$

In Equation 2 above, $\pi_{c_i} = 0.5$ is the prior probability for class i if there are two classes and the number of instances in each class is equal. Also, $|*|$ represent the determinant of a matrix. Finally, the mean vector and covariance matrix for class i are represented by μ_{c_i} and $\sum_{c_i}$, respectively. Specifically, the parameters for the control group were $\mu_{c_0} = [22.01 \;\; 2.886]$ and $\sum_{c_0} = \begin{bmatrix} 2.27 & -0.85 \\ -0.85 & 6.75 \end{bmatrix}$; and for the NPVH group were $\mu_{c_1} = [20.48 \quad 2.359]$ and $\sum_{c_1} = \begin{bmatrix} 3.91 & -6.37 \\ -6.37 & 22.39 \end{bmatrix}$.

The boundary decision of QDA can be derived by using the log-likelihood function, as illustrated in Equation 3:

$$\log\Lambda(x) = \log\frac{p(C = c_1|x,\theta)}{p(C = c_0|x,\theta)}. \quad (3)$$

By setting $\log(\Lambda(x)) = 1$ and calculating θ with a Bayesian optimization procedure (Snoek et al., 2012), we obtained the log-likelihood function at the boundary decision, which can be expressed as a quadratic function illustrated in Equation 4:

$$K + xL + xQx^T = y(x), \quad (4)$$

where K, L, and Q are the bias, linear, and quadratic terms, respectively, x is a feature vector composed of CPP mean and H1–H2 mode per subject, and T represents the transpose operation of a matrix; for derivation of Equation (4), see Othman et al. (2019).

The resulting parameters were $K = -9.25$, $L = [0.286 \;\; -1.24]^T$, $Q = \begin{bmatrix} 0.00659 & 0.0386 \\ 0.0386 & -0.0362 \end{bmatrix}$, and the QDA resulted in an overall classification of 70.8% accuracy: true positives ($n = 25$), true negatives ($n = 26$), false positives ($n = 10$), false negatives ($n = 11$), an AUC = 0.78, and positive and negative likelihood ratios = 2.46 and 0.43, respectively.

To represent the output of both classification models, the logit was used instead of the probability. As shown in Equation 5, the logit (L) is an inverse transformation (i.e., link function) of the nonlinear probability (p) estimate:

$$L = \log\frac{p}{(1-p)}. \quad (5)$$

The logit was used because changes in L (ΔL) can be interpreted equally throughout the scale. In contrast, this is not true for the probability, where, for example, a reduction in p from 0.99 to 0.98 ($\Delta L = -0.31$) represents a much larger improvement than a reduction in p from 0.50 to 0.49 ($\Delta L = -0.02$). Note that negative values of L represent voice use in the normative range.

Figure 1 plots the average weekly CPP mean and H1–H2 mode to illustrate the performance of the linear and nonlinear two-variable models on classification of each of the 36 patients with NPVH and their matched control.

Figure 2 plots the pretherapy training data and posttherapy testing data in reference to the two classification models for all patients who completed voice therapy. The logistic regression logit decreased (toward normal) after therapy compared to before therapy in nine of 11 patients and significantly decreased on average across the entire group; mean $\Delta L = -0.28$, $d = -1.15$. The QDA logit decreased after therapy compared to before therapy in eight of 11 patients and did not significantly decrease on average across the entire group; mean $\Delta L = -1.01$, $d = -0.37$.

Figure 3 plots the test sets of 89 vocally healthy control subjects and 118 patients with PVH in relation to the two classification models. The logistic regression model correctly classified 76 (85%) control subjects as controls—mean (standard deviation) $L = -0.50$ (0.42)—and 105 (89%) patients with PVH as "not NPVH"; $L = -0.60$ (0.46). The QDA model correctly classified 76 (85%) control subjects—$L = -0.50$ (0.47)—and 111 (94%) patients with PVH as "not NPVH"; $L = -0.70$ (0.48).

Discussion

One purpose of this study was to investigate if there were one-dimensional differences in average ambulatory measures of NSAM, f_o, CPP, and H1–H2 between patients with NPVH and matched controls. There were no differences in average signal magnitude (NSAM), replicating

Figure 1. Scatter plots of H1–H2 mode on the *y*-axis and cepstral peak prominence (CPP) on the *x*-axis (patients with vocal hyperfunction: black; matched controls: gray). Each dot represents a single subject's daily mean across 7 days of monitoring. The logistic regression cutoff is represented as a black, solid, diagonal line. The quadratic discriminant analysis cutoff is represented as a black, dashed, curved line. The black Xs and gray X denote the patient and control examples in Figure 4, respectively.

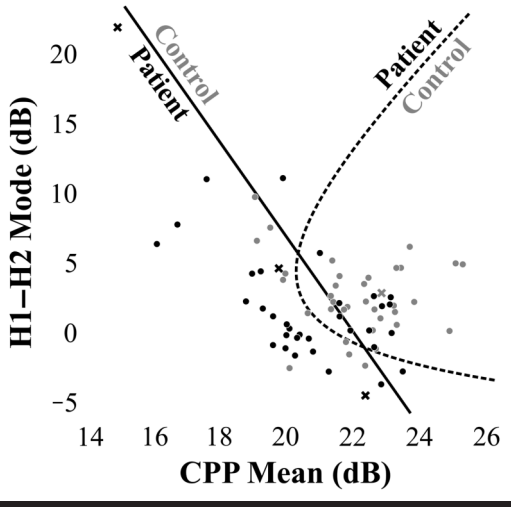
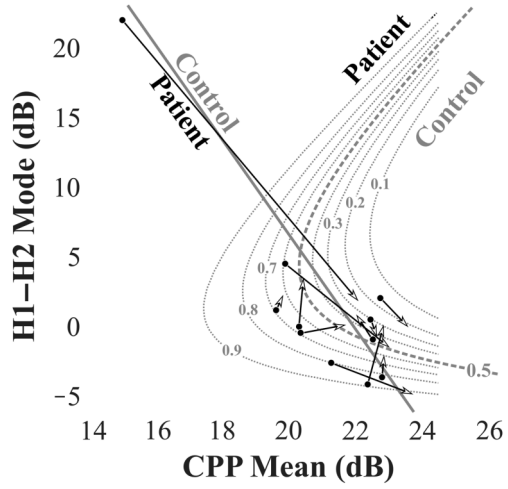
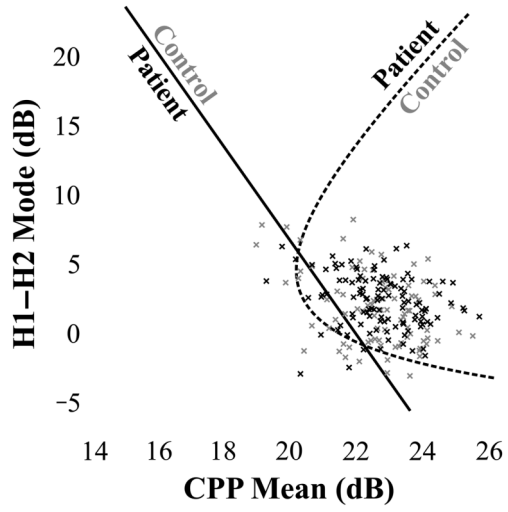


Figure 2. Scatter plots of H1–H2 mode on the *y*-axis and cepstral peak prominence (CPP) on the x-axis. Each black dot and white arrow represent a single patient's daily mean across 7 days of monitoring before and after voice therapy, respectively. The logistic regression cutoff is represented as a gray, solid, diagonal line. The quadratic discriminant analysis (QDA) cutoff is represented as a gray, dashed, curved line. Since the curved classification space with the QDA is not obvious, gray, dotted, thin, curved lines represent specific predicted probabilities in 0.1 increments.



findings by the only other known ambulatory monitoring study that included patients with NPVH and matched controls (Mehta et al., 2015). Group based differences could be reasonably expected because NSAM can strongly correlate

Figure 3. Scatter plots of H1–H2 mode on the *y*-axis and cepstral peak prominence (CPP) on the *x*-axis (patients with phonotraumatic vocal hyperfunction: black; vocally healthy subjects: gray). Each marker represents a single subject's daily mean across 7 days of monitoring. The logistic regression cutoff is represented as a black, solid, diagonal line. The quadratic discriminant analysis cutoff is represented as a black, dashed, curved line.



to measures that have been shown in previous work to be dissimilar between patients and controls, such as aerodynamic (Fryd et al., 2016; Zañartu et al., 2012) and tissue-to-tissue forces (Coleman, 1988; Wokurek & Madsack, 2011; Wokurek & Pützer, 2009, 2011, 2013). For example, studies using in-clinic recordings have shown that patients with NPVH phonate with increased subglottal pressure and open quotient to produce a desired vocal intensity compared to normal phonation (Espinoza et al., 2020, 2017; Hillman et al., 1989). However, this negative NSAM result probably occurred because patients with NPVH aberrantly voice in many different manners (e.g., degrees of aphonia, aperiodicity, strain, breathiness, vocal fry) and each voicing manner is associated with its own mapping between laryngeal forces and vocal intensity (Marks et al., 2020, 2019).

In contrast to the previous ambulatory monitoring study (Mehta et al., 2015), multiple features (f_o mode and 5th percentile, nearly all CPP statistics) were independently associated with medium-to-large differences between the two groups. This may be because the 21 new patients in the current study produced more severe overall dysphonia compared to the 15 patients in both this study and the previous study (mean CAPE-V overall dysphonia was 46.1 vs. 31.6, respectively). Despite this difference in overall dysphonia, the new patients reported a similar level of impaired voice-related quality of life compared to the patients in both studies (mean V-RQOL total was 63.2 and 64.2, respectively). Higher levels of overall dysphonia could have primarily contributed to lower ambulatory CPP values in the NPVH group compared to the matched controls ($d = -0.71$ to -1.08). Also, more severe cases have been associated with increases in f_o (Roy & Hendarto, 2005), and the NPVH group did voice at obviously higher f_o values than their matched controls ($d = 0.61$–0.64).

A novel finding was that H1–H2 skewness was less positive in the NPVH group than the matched-control group with a large effect size ($d = -1.1$). As skewness is a distributional feature, this finding is interpreted to indicate a daily tendency to produce inefficient phonation. More specifically, this feature indicates that the NPVH group, compared to the control group, phonated more often above their median H1–H2 than below (i.e., with less abrupt vocal fold closure more often than more abrupt vocal fold closure). This aligns with the hypothesized pathophysiology of NPVH, where patients inefficiently voice with minimal risk of phonotrauma due to decreased abruptness of vocal fold closure (Hillman et al., 2020).

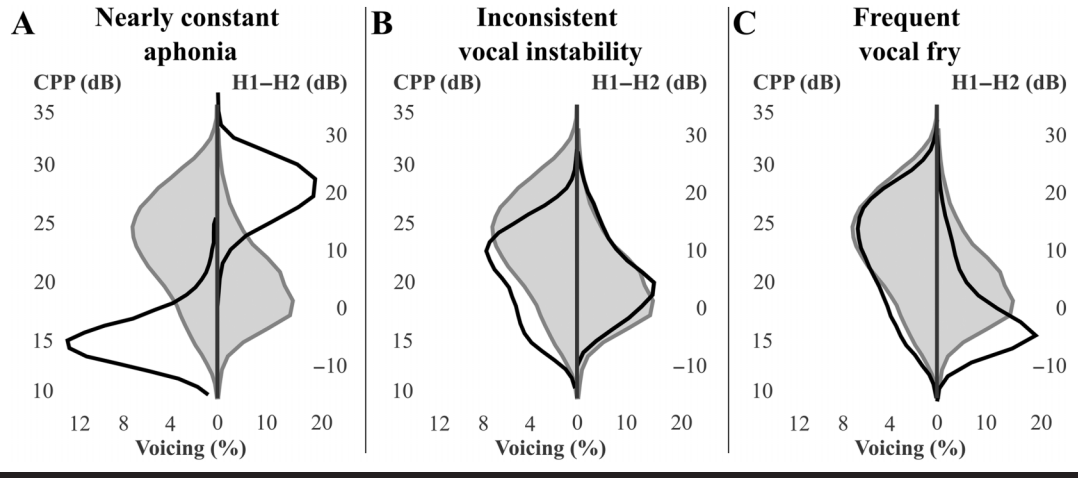
Linear and nonlinear data-driven classifiers significantly differentiated patients with NPVH from their matched controls (AUC = 0.78 for both models) using two lower order distributional statistics (CPP mean and H1–H2 mode). As seen in Figure 1, the patients with NPVH appear to vary along a continuum of lower CPP and higher H1–H2 than normal (upper left corner) to normal CPP and lower H1–H2 than normal (bottom right corner). The patients located at the top left extreme of the continuum represent the more typical characterization of NPVH with severe dysphonia and varying amounts of aphonia (low CPP) as well as minimal

vocal fold closure evidenced by a high degree of breathiness (high H1–H2). The patients located at the other extreme of the continuum in the bottom right appear to be producing periodic phonation (near normal CPP) with increased vocal fold closure (low H1–H2). In auditory-perceptual terms, the voices of these patients contained nearly constant vocal fry (i.e., pulse) registration. The pulse register has subharmonics that would not significantly reduce CPP (because it is "periodic"), would be associated with lower H1–H2 values (because this pulse registration is associated with complete glottal closure), and produce negative H1–H2 values (increased potential to have a subharmonic register as H1 and the "true" first harmonic to register as H2). However, we hypothesize that there is still diminished potential for phonotrauma because pulse register phonation is produced with a slack vocal fold cover (i.e., reduced stiffness of the vocal fold layers that participate in vibration; Hollien, 1974), which should reduce/dampen the forces (e.g., collision) that could contribute to tissue damage. The voices of the patients between these two extremes had inconsistent pulse registration and/or overall dysphonia with pitch breaks, phonation breaks, and so forth. Figure 4 shows distributional examples of individual patients (black Xs in Figure 1) in reference to a vocally healthy control (gray X in Figure 1) to help further illustrate how the NPVH group varies in this two-dimensional space. Audio samples of the three patients in Figure 4 reading the Rainbow Passage are provided in Supplemental Material S1. In summary, this CPP and H1–H2 space could represent a pathophysiological continuum of inefficient vocal behaviors that have minimal likelihood to produce phonotrauma across the spectrum of vocal fold closure dynamics ranging from dysphonic voicing with reduced abruptness of vocal fold closure (upper right corner of Figure 1) to producing vocal fry that has very abrupt vocal fold closure in the presence of slack vocal fold tissue.

The validity of this two-dimensional space in representing the pathophysiology of NPVH was strengthened by both models performing well on unseen data. Specifically, as shown in Figure 2, the logits for 11 patients with NPVH decreased after therapy with a large effect size (linear classifier $d = -1.15$). Also, as shown in Figure 3, individual subject logits were below the "NPVH" classification threshold (i.e., $L = 0$) for nearly all of the held-out 89 vocally healthy controls (both the linear and nonlinear classifiers were 85% accurate) and the 118 patients with PVH (linear and nonlinear classifier were 89% and 94%, respectively). Correctly classifying an overwhelming majority of patients with PVH as "not NPVH" is a critical result for two reasons. First, it further validates the model's ability to specifically characterize the pathophysiology of NPVH, as patients with phonotraumatic lesions (i.e., nodules, polyps) are obviously not displaying phonation that combines inefficiency with minimal likelihood of phonotrauma. Second, it lends credence to the overarching belief that vocal hyperfunction has two primary underlying unique pathophysiologies: phonotraumatic and nonphonotraumatic (Hillman et al., 2020). In other words, patients whose daily phonatory function falls along the NPVH pathophysiological continuum described are unlikely to develop phonotraumatic lesions in the future.

It is commonly believed that patients with NPVH may vocally deteriorate and/or improve throughout their daily lives due to many external factors. Therefore, it is possible that a patient who varies in and out of abnormal vocal states throughout a day could be classified as "normal" on an average daily basis. To potentially identify a subgroup of NPVH patients who were more variable (i.e., fluctuating in severity) than controls, all daily data for the patient and

Figure 4. Data representing an average normal subject (gray lines and fill) and example patient (black lines with no fill) weeklong histograms of cepstral peak prominence (CPP; left side) and H1–H2 (right side). All example patients and the example control can be seen as black or gray Xs, respectively. The normal subject was chosen because her data produced a logit closest to the average "normal" logit and the same normal is represented in Panels A through C. Data from three different patients are illustrated who voiced with nearly constant aphonia (Panel A), inconsistent vocal instabilities (e.g., pitch breaks, breathiness, and vocal fry; Panel B), and frequent vocal fry (Panel C). Audio samples of the three patients can be heard in a WAV file contained in Supplemental Material S1.



controls groups were subdivided into 1-hr time durations. There was no evidence, at the hour-long time scale, that patients with NPVH were more variable than the matched control group in their logit. More patients than controls had ≥ 80% of hours classified as patients (19 patients vs. three controls) and more controls than patients had ≤ 60% of hours classified as patient (27 controls vs. 11 patients). This essentially supports the clinical notion that patients with NPVH tend to stay "stuck" in a pathophysiological mode of phonating throughout daily life (Roy, 2008) and previous work showing that ~70%–80% of patients with vocal hyperfunction self-reported no variation in vocal status throughout their daily life (Van Stan et al., 2017). However, the variability analysis in this paragraph should be considered preliminary because it only investigated a single time scale (1 hr with 0% overlap) and one feature (the logistic classifier). Future work to comprehensively investigate variability differences between patients and controls (as well as identifying patient subgroups) should include multiple clinically relevant time scales (shorter and longer durations than 1 hr, potentially with various degrees of overlap), measures (NSAM, f_o, CPP, H1–H2), and contexts (louder vs. softer voicing, or phonating in the presence of louder vs. softer background noise levels).

The linear and nonlinear classifiers developed here were statistically significant with overall accuracies around 70%, but this performance is lower than most classification models that have discriminated among vocal states at ~90% accuracy (Hegde et al., 2019). However, the purpose of most classification studies in the literature were different than the goal of this study. Specifically, the end goal of many classification attempts has been to achieve maximum discrimination among voice qualities or between healthy and (generally) disordered voicing. In contrast, this study used classification as a means toward a different end: improved and/or new insights into the pathophysiology of a specific voice disorder, NPVH. Due to these different ends, classification attempts in the literature (compared to this study) have used many more features (10 s of features vs. two features, respectively) and more complicated models (neural networks or support vector machines vs. linear regressions or discriminant analyses, respectively). While this increased complexity achieves higher classification accuracies, it seriously hinders any interpretation of the results (i.e., pathophysiological insights into the disorder) that would make the model applicable to voice assessment and treatment. It is also important to note that these newly developed linear and nonlinear models were not designed to diagnose or screen for patients with NPVH (or exclude an NPVH diagnosis), but to characterize clinically meaningful differences in ambulatory voice use between already-diagnosed patients with NPVH and matched controls. The diagnosis of NPVH critically relies upon a complete case history and thorough visualization of the larynx to rule out any structural or neurological contribution to the patient's vocal complaints.

Some NPVH patients received voice therapy with generally positive outcomes, but it is important to note that strict therapeutic protocols were not established or used, and this study is not intended to support treatment efficacy or effectiveness. While the completion rate of voice therapy may appear low (11 of the 36 patients with NPVH successfully completed voice therapy), there are multiple reasons for this, including the following: Some patients did not agree to a second round of monitoring after completing therapy, some patients were recently enrolled in the study and were still undergoing therapy or were on the waiting list to begin therapy, some patients never attempted to begin therapy, and some patients dropped out of therapy before finishing. Future work should continue to support or refute the validity of these classification models for use as a clinical outcome measure. If the classification algorithms represent a significant aspect of the underlying NPVH pathophysiology, they should successfully classify a new group of female patients with NPVH, new posttherapy data, and generalize (albeit with different sex-specific weightings) to males with NPVH. Future work should also investigate where the models fail or succeed most often. For example, patients who were rated at 40 or higher on the CAPE-V dysphonia scale were classified at 89% accuracy (15 out of 17 patients) and patients who were below 40 on this scale were classified at 53% accuracy (10 out of 19 patients). Therefore, it is possible that dysphonia severity contributes to the model's performance. However, dysphonia may not be the main driver of classification because (a) many patients with milder dysphonia were still classified correctly and (b) the correlation between the model's logit and overall dysphonia across subjects was nonsignificant. Although ambulatory voice monitoring can provide insights into voice use outside of the clinic, a potential confound is that the assessment procedures (e.g., diagnosis, education) and/or monitoring could have changed the subject's typical behavior (Hunter, 2012). However, this seems improbable as it often takes weeks of voice therapy sessions to modify a patient's habitual voice use (Ziegler et al., 2014). Additionally, subjects often reported forgetting that they were being monitored. Finally, there are many other potentially confounding contextual and personal factors that were not included in this study, for example, acoustic environments (Bottalico et al., 2015; Whittico et al., 2020), personality (Roy & Bless, 2000), and stress (Dietrich et al., 2008; Helou et al., 2013). Future work could investigate how these variables interact with the classification models.

Conclusions

As a group, patients with NPVH appear to phonate in daily life with less periodicity (reduced CPP), at higher pitches (increased f_o), and with less abrupt vocal fold closure (more positive H1–H2 skew) compared to matched vocally healthy controls. Multidimensional data-driven classification models used the combination of daily CPP mean and H1–H2 mode to maximally discriminate between the NPVH and control groups. These classifiers appear to have some validity, as they correctly classified > 80% of unseen data, for example, posttherapy NPVH data, a held out set of vocally healthy subjects, and patients with PVH. The combination of CPP mean and H1–H2 mode appear to reflect a

pathophysiological continuum of inefficient phonation with reduced potential for phonotrauma. Further investigations involving measures with closer theoretical ties to the underlying physiology of phonation—for example, subglottal pressure, open quotient (Espinoza et al., 2020, 2017; Hillman et al., 1989)—and other contributing factors—for example, stress (Dietrich et al., 2008; Helou et al., 2013), personality (Roy & Bless, 2000)—are needed to further improve the understanding of how this model relates to the underlying etiology and pathophysiology of NPVH.

Acknowledgments

This work was supported by the Voice Health Institute and the National Institute on Deafness and Other Communication Disorders under Grants R33 DC011588 and P50 DC015446 (Principal Investigator for both: Robert Hillman). 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 acknowledge the contributions of Robert Petit for aid in designing and programming the smartphone application and Brianna E. Williams for assisting with data processing.

References

Altman, K. W., Atkinson, C., & Lazarus, C. (2005). Current and emerging concepts in muscle tension dysphonia: A 30-month review. *Journal of Voice, 19*(2), 261–267. https://doi.org/10.1016/j.jvoice.2004.03.007

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

Berry, D. A., Verdolini, K., Montequin, D. W., Hess, M. M., Chan, R. W., & Titze, I. R. (2001). A quantitative output-cost ratio in voice production. *Journal of Speech, Language, and Hearing Research, 44*(1), 29–37. https://doi.org/10.1044/1092-4388(2001/003)

Bottalico, P., Graetzer, S., & Hunter, E. J. (2015). Vocal effort and the effect of room acoustics in noisy environments. In C. Burroughs (Ed.), *Proceedings of the 44th International Congress and Exposition on Noise Control Engineering (Internoise 2015): Implementing Noise Control Technology* (pp. 4688–4698). Institute of Noise Control Engineering - USA.

Carding, P. N., Horsley, I. A., & Docherty, G. J. (1999). A study of the effectiveness of voice therapy in the treatment of 45 patients with nonorganic dysphonia. *Journal of Voice, 13*(1), 72–104. https://doi.org/10.1016/S0892-1997(99)80063-0

Cheyne, H. A. (2002). *Estimating glottal voicing source characteristics by measuring and modeling the acceleration of the skin on the neck* [Doctoral dissertation, Massachusetts Institute of Technology]. MIT Libraries. http://hdl.handle.net/1721.1/8351

Cheyne, H. A., Hanson, H. M., Genereux, R. P., Stevens, K. N., & Hillman, R. E. (2003). Development and testing of a portable vocal accumulator. *Journal of Speech, Language, and Hearing Research, 46*(6), 1457–1467. https://doi.org/10.1044/1092-4388(2003/113)

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Coleman, R. F. (1988). Comparison of microphone and neck-mounted accelerometer monitoring of the performing voice. *Journal of Voice, 2*(3), 200–205. https://doi.org/10.1016/S0892-1997(88)80077-8

Cooper, M. (1974). Spectrographic analysis of fundamental frequency and hoarseness before and after vocal rehabilitation. *The Journal of Speech and Hearing Disorders, 39*(3), 286–297. https://doi.org/10.1044/jshd.3903.286

Cortés, J. P., Espinoza, V. M., Ghassemi, M., Mehta, D. D., Van Stan, J. H., Hillman, R. E., Guttag, J. V., & Zañartu, M. (2018). Ambulatory assessment of phonotraumatic vocal hyperfunction using glottal airflow measures estimated from neck-surface acceleration. *PLOS ONE, 13*(12), e0209017. https://doi.org/10.1371/journal.pone.0209017

Coyle, S. M., Weinrich, B. D., & Stemple, J. C. (2001). Shifts in relative prevalence of laryngeal pathology in a treatment-seeking population. *Journal of Voice, 15*(3), 424–440. https://doi.org/10.1016/S0892-1997(01)00043-1

Dietrich, M., Verdolini Abbott, K., Gartner-Schmidt, J., & Rosen, C. A. (2008). The frequency of perceived stress, anxiety, and depression in patients with common pathologies affecting voice. *Journal of Voice, 22*(4), 472–488. https://doi.org/10.1016/j.jvoice.2006.08.007

Dworkin, J. P., Meleca, R. J., & Abkarian, G. G. (2000). Muscle tension dysphonia. *Current Opinion in Otolaryngology & Head and Neck Surgery, 8*(3), 169–173. https://doi.org/10.1097/00020840-200006000-00007

Espinoza, V. M., Mehta, D. D., Van Stan, J. H., Hillman, R. E., & Zañartu, M. (2020). Glottal aerodynamics estimated from neck-surface vibration in women with phonotraumatic and nonphonotraumatic vocal hyperfunction. *Journal of Speech, Language, and Hearing Research, 63*(9), 2861–2869. https://doi.org/10.1044/2020_JSLHR-20-00189

Espinoza, V. M., Zañartu, 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

Fernández, S., Garaycochea, O., Martinez-Arellano, A., & Alcalde, J. (2020). Does more compression mean more pressure? A new classification for muscle tension dysphonia. *Journal of Speech, Language, and Hearing Research, 63*(7), 2177–2184. https://doi.org/10.1044/2020_JSLHR-20-00042

Fryd, A. S., Van Stan, J. H., Hillman, R. E., & Mehta, D. D. (2016). Estimating subglottal pressure from neck-surface acceleration during normal voice production. *Journal of Speech, Language, and Hearing Research, 59*(6), 1335–1345. https://doi.org/10.1044/2016_JSLHR-S-15-0430

Gillespie, A. I., Gartner-Schmidt, J., Rubinstein, E. N., & Verdolini Abbott, K. (2013). Aerodynamic profiles of women with muscle tension dysphonia/aphonia. *Journal of Speech, Language, and Hearing Research, 56*(2), 481–488. https://doi.org/10.1044/1092-4388(2012/11-0217)

Goldman, S. L., Hargrave, J., Hillman, R. E., Holmberg, E., & Gress, C. (1996). Stress, anxiety, somatic complaints, and voice use in women with vocal nodules: Preliminary findings. *American Journal of Speech-Language Pathology, 5*(1), 44–54. https://doi.org/10.1044/1058-0360.0501.44

Gunter, H. E., Howe, R. D., Zeitels, S. M., Kobler, J. B., & Hillman, R. E. (2005). Measurement of vocal fold collision forces during phonation: Methods and preliminary data. *Journal of Speech, Language, and Hearing Research, 48*(3), 567–576. https://doi.org/10.1044/1092-4388(2005/039)

Hegde, S., Shetty, S., Rai, S., & Dodderi, T. (2019). A survey on machine learning approaches for automatic detection of voice disorders. *Journal of Voice, 33*(6), 947.e911–947.e933. https://doi.org/10.1016/j.jvoice.2018.07.014

Helou, L. B., Wang, W., Ashmore, R. C., Rosen, C. A., & Abbott, K. V. (2013). Intrinsic laryngeal muscle activity in response to autonomic nervous system activation. *The Laryngoscope, 123*(11), 2756–2765. https://doi.org/10.1002/lary.24109

Herrington-Hall, B. L., Lee, L., Stemple, J. C., Niemi, K. R., & McHone, M. M. (1988). Description of laryngeal pathologies by age, sex, and occupation in a treatment-seeking sample. *The Journal of Speech and Hearing Disorders, 53*(1), 57–64. https://doi.org/10.1044/jshd.5301.57

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., Stepp, C. E., Van Stan, J. H., Zañartu, 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

Hogikyan, N. D., & Sethuraman, G. (1999). Validation of an instrument to measure voice-related quality of life (V-RQOL). *Journal of Voice, 13*(4), 557–569. https://doi.org/10.1016/S0892-1997(99)80010-1

Hollien, H. (1974). On vocal registers. *Journal of Phonetics, 2*(2), 125–143. https://doi.org/10.1016/S0095-4470(19)31188-X

Hunter, E. J. (2012). Teacher response to ambulatory monitoring of voice. *Logopedics Phoniatrics Vocology, 37*(3), 133–135. https://doi.org/10.3109/14015439.2012.664657

Jiang, J. J., & Titze, I. R. (1994). Measurement of vocal fold intraglottal pressure and impact stress. *Journal of Voice, 8*(2), 132–144. https://doi.org/10.1016/S0892-1997(05)80305-4

Kempster, G. B., Gerratt, B. R., Verdolini Abbott, K., 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)

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustical Society of America, 87*(2), 820–857. https://doi.org/10.1121/1.398894

Koufman, J. A., Amin, M. R., & Panetti, M. (2000). Prevalence of reflux in 113 consecutive patients with laryngeal and voice disorders. *Otolaryngology—Head & Neck Surgery, 123*(4), 385–388. https://doi.org/10.1067/mhn.2000.109935

Koufman, J. A., & Blalock, P. D. (1988). Vocal fatigue and dysphonia in the professional voice user: Bogart-Bacall syndrome. *The Laryngoscope, 98*(5), 493–498. https://doi.org/10.1288/00005537-198805000-00003

Kridgen, S., Hillman, R. E., Stadelman-Cohen, T., Zeitels, S., Burns, J. A., Hron, T., Krusemark, C., Muise, J., & Van Stan, J. H. (2020). Patient-reported factors associated with the onset of hyperfunctional voice disorders. *Annals of Otology, Rhinology & Laryngology*.

Lenth, R., Buerkner, P., Herve, M., Love, J., Riebl, H., & Singmann, H. (2020). Package "emmeans." The Comprehensive R Archive Network (CRAN). https://cran.r-project.org/web/packages/emmeans/emmeans.pdf

Marks, K. L., Lin, J. Z., Burns, J. A., Hron, T. A., Hillman, R. E., & Mehta, D. D. (2020). Estimation of subglottal pressure from neck surface vibration in patients with voice disorders. *Journal of Speech, Language, and Hearing Research, 63*(7), 2202–2218. https://doi.org/10.1044/2020_JSLHR-19-00409

Marks, K. L., Lin, J. Z., Fox, A. B., Toles, L. E., & Mehta, D. D. (2019). Impact of nonmodal phonation on estimates of subglottal pressure from neck-surface acceleration in healthy speakers. *Journal of Speech, Language, and Hearing Research, 62*(9), 3339–3358. https://doi.org/10.1044/2019_JSLHR-S-19-0067

Mathieson, L., Hirani, S. P., Epstein, R., Baken, R. J., Wood, G., & Rubin, J. S. (2009). Laryngeal manual therapy: A preliminary study to examine its treatment effects in the management of muscle tension dysphonia. *Journal of Voice, 23*(3), 353–366. https://doi.org/10.1016/j.jvoice.2007.10.002

Mehta, D. D., Espinoza, V. M., Van Stan, J. H., Zañartu, M., & Hillman, R. E. (2019). The difference between first and second harmonic amplitudes correlates between glottal airflow and neck-surface accelerometer signals during phonation. *The Journal of the Acoustical Society of America, 145*(5), EL386–EL392. https://doi.org/10.1121/1.5100909

Mehta, D. D., Van Stan, J. H., Zañartu, M., Ghassemi, M., Guttag, J. V., Espinoza, V. M., Cortés, J. P., Cheyne, H. A., II, & Hillman, R. E. (2015). Using ambulatory voice monitoring to investigate common voice disorders: Research update. *Frontiers in Bioengineering and Biotechnology, 3,* 155. https://doi.org/10.3389/fbioe.2015.00155

Mehta, D. D., Zañartu, M., Feng, S. W., Cheyne, H. A., II, & Hillman, R. E. (2012). Mobile voice health monitoring using a wearable accelerometer sensor and a smartphone platform. *IEEE Transactions on Bio-medical Engineering, 59*(11), 3090–3096. https://doi.org/10.1109/TBME.2012.2207896

Misono, S., Meredith, L., Peterson, C. B., & Frazier, P. A. (2016). New perspective on psychosocial distress in patients with dysphonia: The moderating role of perceived control. *Journal of Voice, 30*(2), 172–176. https://doi.org/10.1016/.jvoice.2015.02.002

Nguyen, D. D., & Kenny, D. T. (2009). Randomized controlled trial of vocal function exercises on muscle tension dysphonia in Vietnamese female teachers. *Journal of Otolaryngology–Head & Neck Surgery, 38*(2), 261–278. https://doi.org/10.1016/j.jvoice.2007.09.003

Othman, N. H., Lee, K. Y., Radzol, A. R. M., Mansor, W., & Hisham, N. I. A. (2019). PCA-QDA model selection for detecting NS1 related diseases from SERS spectra of salivary mixtures. In L. Lhotska, L. Sukupova, I. Lacković, & G. Ibbott (Eds.), *World congress on medical physics and biomedical engineering 2018* (pp. 623–627). Springer. https://doi.org/10.1007/978-981-10-9035-6_116

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

Ramig, L. O., & Verdolini, K. (1998). Treatment efficacy: Voice disorders. *Journal of Speech, Language, and Hearing Research, 41*(1), S101–S116. https://doi.org/10.1044/jslhr.4101.s101

Rattenbury, H. J., Carding, P. N., & Finn, P. (2004). Evaluating the effectiveness and efficiency of voice therapy using transnasal flexible laryngoscopy: A randomized controlled trial. *Journal of Voice, 18*(4), 522–533. https://doi.org/10.1016/j.jvoice.2004.01.003

R Core Team. (2018). R: A language and environment for statistical computing, Vienna, Austria. https://www.R-project.org/

Ross, J. A., Noordzji, J. P., & Woo, P. (1998). Voice disorders in patients with suspected laryngo-pharyngeal reflux disease. *Journal of Voice, 12*(1), 84–88. https://doi.org/10.1016/S0892-1997(98)80078-7

Roy, N. (2008). Assessment and treatment of musculoskeletal tension in hyperfunctional voice disorders. *International Journal of Speech-Language Pathology, 10*(4), 195–209. https://doi.org/10.1080/17549500701885577

Roy, N., & Bless, D. M. (2000). Personality traits and psychological factors in voice pathology: A foundation for future research. *Journal of Speech, Language, and Hearing Research, 43*(3), 737–748. https://doi.org/10.1044/jslhr.4303.737

Roy, N., Bless, D. M., & Heisey, D. (2000a). Personality and voice disorders: A multitrait-multidisorder analysis. *Journal of Voice, 14*(4), 521–548. https://doi.org/10.1016/S0892-1997(00)80009-0

Roy, N., Bless, D. M., & Heisey, D. (2000b). Personality and voice disorders: A superfactor trait analysis. *Journal of Speech, Language, and Hearing Research, 43*(3), 749–768. https://doi.org/10.1044/jslhr.4303.749

Roy, N., Bless, D. M., Heisey, D., & Ford, C. N. (1997). Manual circumlaryngeal therapy for functional dysphonia: An evaluation of short-and long-term treatment outcomes. *Journal of Voice, 11*(3), 321–331. https://doi.org/10.1016/S0892-1997(97)80011-2

Roy, N., & Hendarto, H. (2005). Revisiting the pitch controversy: Changes in speaking fundamental frequency (SFF) after management of functional dysphonia. *Journal of Voice, 19*(4), 582–591. https://doi.org/10.1016/j.jvoice.2004.08.005

Roy, N., & Leeper, H. A. (1993). Effects of the manual laryngeal musculoskeletal tension reduction technique as a treatment for functional voice disorders: Perceptual and acoustic measures. *Journal of Voice, 7*(3), 242–249. https://doi.org/10.1016/S0892-1997(05)80333-9

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

Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian optimization of machine learning algorithms. In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 25 (NIPS 2012)*. Neural Information Processing Systems.

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

Spencer, M. L. (2015). Muscle tension dysphonia: A rationale for symptomatic subtypes, expedited treatment, and increased therapy compliance. *SIG 3 Perspectives on Voice and Voice Disorders, 25*(1), 5–15. https://doi.org/10.1044/vvd25.1.5

Stevens, K. N. (1998). *Acoustic phonetics*. MIT Press.

Szabo Portela, A., Granqvist, S., Ternström, S., & Södersten, M. (2018). Vocal behavior in environmental noise: Comparisons between work and leisure conditions in women with work-related voice disorders and matched controls. *Journal of Voice, 32*(1), 126.e23–126.e38. https://doi.org/10.1016/j.jvoice.2017.04.010

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 Lierde, K. M., De Bodt, M., Dhaeseleer, E., Wuyts, F., & Claeys, S. (2010). The treatment of muscle tension dysphonia: A comparison of two treatment techniques by means of an objective multiparameter approach. *Journal of Voice, 24*(3), 294–301. https://doi.org/10.1016/j.jvoice.2008.09.003

Van Lierde, K. M., De Ley, S., Clement, G., De Bodt, M., & Van Cauwenberge, P. (2004). Outcome of laryngeal manual therapy in four Dutch adults with persistent moderate-to-severe vocal hyperfunction: A pilot study. *Journal of Voice, 18*(4), 467–474. https://doi.org/10.1016/j.jvoice.2004.02.003

van Mersbergen, M., Patrick, C., & Glaze, L. (2008). Functional dysphonia during mental imagery: Testing the trait theory of voice disorders. *Journal of Speech, Language, and Hearing Research, 51*(6), 1405–1423. https://doi.org/10.1044/1092-4388(2008/06-0216)

Van Stan, J. H., Maffei, M., Masson, M. V., Mehta, D. D., Burns, J. A., & Hillman, R. E. (2017). Self-ratings of vocal status in daily life: Reliability and validity for patients with vocal hyperfunction and a normative group. *American Journal of Speech-Language Pathology, 26*(4), 1167–1177. https://doi.org/10.1044/2017_AJSLP-17-0031

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., & Hillman, R. E. (2020a). Changes in a Daily Phonotrauma Index (DPI) 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. (2020b). 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., Mehta, D. D., Zeitels, S. M., Burns, J. A., Barbu, A. M., & Hillman, R. E. (2015). Average ambulatory measures of sound pressure level, fundamental frequency, and vocal dose do not differ between adult females with phonotraumatic lesions and matched control subjects. *Annals of Otology, Rhinology & Laryngology, 124*(11), 864–874. https://doi.org/10.1177/0003489415589363

Verdolini, K., Rosen, C. A., & Branski, R. C. (Eds.). (2006). *Classification manual for voice disorders-I*. Erlbaum.

Whittico, T. H., Ortiz, A. J., Marks, K. L., Toles, L. E., Van Stan, J. H., Hillman, R. E., & Mehta, D. D. (2020). Ambulatory monitoring of Lombard-related vocal characteristics in vocally healthy female speakers. *The Journal of the Acoustical Society of America, 147*(6), EL552–EL558. https://doi.org/10.1121/10.0001446

Wokurek, W., & Madsack, A. (2011, August). *Acceleration sensor based estimates of subglottal resonances: Short vs. long vowels*. INTERSPEECH 2011, 12th Annual Conference of the International Speech Communication Association, Florence, Italy.

Wokurek, W., & Pützer, M. (2009). Acceleration sensor measurements of subglottal sound pressure for modal and breathy phonation quality. In C. Manfredi (Ed.), *Proceedings of the 6th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications* (pp. 153–156). Firenze University Press.

Wokurek, W., & Pützer, M. (2011). Acceleration sensor measurements of vibrations of the larynx in patients with vocal fold adduction deficiencies. In C. Manfredi (Ed.), *Proceedings of the 7th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications* (pp. 191–194). Firenze University Press.

Wokurek, W., & Pützer, M. (2013). Correlation analysis between acoustic source, electroglottogram and neck vibrations signals. In C. Manfredi (Ed.), *Proceedings of the 9th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications* (pp. 89–92). Firenze University Press.

Zañartu, M., Ho, J. C., Mehta, D. D., Hillman, R. E., & Wodicka, G. R. (2013). Subglottal impedance-based inverse filtering of

voiced sounds using neck surface acceleration. *IEEE Transactions on Audio, Speech, and Language Processing, 21*(9), 1929–1939. https://doi.org/10.1109/TASL.2013.2263138

Zañartu, M., Wodicka, G. R., Ho, J. C., Hillman, R. E., & Mehta, D. D. (2012). *Estimation of glottal aerodynamics using an impedance-based inverse filtering of neck surface acceleration* (U.S. Patent No. Application Number 61444199). U.S. Patent and Trademark Office. https://patents.google.com/patent/US20170014082A1/en

Ziegler, A., Dastolfo, C., Hersan, R., Rosen, C. A., & Gartner-Schmidt, J. (2014). Perceptions of voice therapy from patients diagnosed with primary muscle tension dysphonia and benign mid-membranous vocal fold lesions. *Journal of Voice, 28*(6), 742–752. https://doi.org/10.1016/j.jvoice.2014.02.007

Švec, J. G., Titze, I. R., & Popolo, P. S. (2005). Estimation of sound pressure levels of voiced speech from skin vibration of the neck. *The Journal of the Acoustical Society of America, 117*(3 Pt. 1), 1386–1394. https://doi.org/10.1121/1.1850074