



Published in final edited form as:

*Ann Otol Rhinol Laryngol.* 2015 November ; 124(11): 864–874. doi:10.1177/0003489415589363.

## 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

Jarrad H. Van Stan, MA, CCC-SLP, BCS-S<sup>1,2</sup>, Daryush D. Mehta, PhD<sup>1,2,3</sup>, Steven M. Zeitels, MD, FACS<sup>1,3</sup>, James A. Burns, MD, FACS<sup>1,3</sup>, Anca M. Barbu, MD, FACS<sup>1,3</sup>, and Robert E. Hillman, PhD, CCC-SLP<sup>1,2,3,\*</sup>

<sup>1</sup>Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital; 1 Bowdoin Square, 11<sup>th</sup> floor, Boston, MA 02114

<sup>2</sup>MGH Institute of Health Professions; 36 1<sup>st</sup> Avenue, Boston, MA 02129

<sup>3</sup>Department of Surgery, Harvard Medical School; 25 Shattuck Street, Boston, MA 02115

### Abstract

**Objectives**—Clinical management of phonotraumatic vocal fold lesions (nodules, polyps) is based largely on assumptions that abnormalities in habitual levels of sound pressure level (SPL), fundamental frequency (f<sub>0</sub>), and/or amount of voice use play a major role in lesion development and chronic persistence. This study used ambulatory voice monitoring to evaluate if significant differences in voice use exist between patients with phonotraumatic lesions and normal matched controls.

**Methods**—Subjects were 70 adult females: 35 with vocal fold nodules or polyps and 35 age-, sex-, and occupation-matched normal individuals. Weeklong summary statistics of voice use were computed from anterior neck surface acceleration recorded using a smartphone-based ambulatory voice monitor.

**Results**—Paired t-tests and Kolmogorov-Smirnov tests resulted in no statistically significant differences between patients and matched controls regarding average measures of SPL, f<sub>0</sub>, vocal dose measures, and voicing/voice rest periods. Paired t-tests comparing f<sub>0</sub> variability between the groups resulted in statistically significant differences with moderate effect sizes.

**Conclusions**—Individuals with phonotraumatic lesions did not exhibit differences in average ambulatory measures of vocal behavior when compared with matched controls. More refined characterizations of underlying phonatory mechanisms and other potentially contributing causes are warranted to better understand risk factors associated with phonotraumatic lesions.

\*Corresponding Author & Requests for Reprints: Hillman.Robert@MGH.HARVARD.EDU; 617-643-2466, office; 617-643-0681, fax.

## Keywords

ambulatory voice monitoring; vocal loading; vocal fold nodules; vocal fold polyps; voice disorders; occupational voice use

---

## INTRODUCTION

The role of daily voice use in the development of phonotraumatic vocal fold lesions (nodules and polyps) is unclear, and there is a paucity of quantitative information about the impact of these lesions on daily vocal function. The literature contains inconsistent empirical findings, divergent theories regarding causes/effects of these lesions, and a limited number of clinical studies tracking the daily vocal behavior of individuals who develop these disorders. Current clinical management of phonotrauma is based largely on assumptions that abnormalities in habitual levels of vocal intensity, fundamental frequency ( $f_0$ ), and/or amount of voice use play a major role in the cause and chronic persistence of these disorders.<sup>1-5</sup> However, relationships between actual measures of daily/habitual voice use and phonotrauma have not been clearly established and could therefore be less straightforward than assumed.

Clinical behavioral treatment for phonotraumatic lesions frequently involves a comprehensive voice assessment followed by therapeutic procedures specifically aimed at the modification of habitual vocal intensity,  $f_0$ , and/or excessive voice use through either direct, indirect, or compensatory rehabilitative interventions.<sup>6</sup> Examples of direct interventions include the reduction of conversational vocal intensity, confidential voice therapy, finding one's "optimal pitch," and voice rest.<sup>7-9</sup> Examples of indirect interventions and compensatory strategies include sound field or voice amplification<sup>10-13</sup> and vocal hygiene goals, such as avoiding environments requiring louder speech and "voice conservation."<sup>14-16</sup> These therapeutic approaches have historically deep underpinnings—especially as they relate to symptomatic voice therapy paradigms<sup>9</sup>—and have been strengthened by clinical experience and epidemiologic studies indicating that individuals in high-voice use occupations are more likely to develop voice disorders than in other occupations.<sup>17-19</sup> However, this prevailing view is not based on empirical evidence that directly links the daily vocal behavior of individuals in high-voice use occupations with the development of phonotraumatic disorders, nor does it provide a clear explanation for why only a certain minority of individuals in these occupations develop voice disorders while the majority of those in high-voice use occupations do not.

Increased average vocal intensity has been hypothesized to be related to the biological development of phonotraumatic lesions<sup>20,21</sup> and the compromise of vocal fold epithelial structures.<sup>22</sup> However, there is a lack of agreement in the literature about the association between average vocal intensity and the development and/or presence of vocal fold lesions. Some studies using in-clinic recordings have reported no difference in average vocal intensity between participants with vocal fold nodules and a control group,<sup>1,23</sup> while others found that participants with nodules or polyps spoke at a significantly higher average vocal intensity.<sup>24</sup> Furthermore, no differences in average vocal intensity have been noted when

comparing patient characteristics before and after voice therapy<sup>23,24</sup> or before and after surgical procedures.<sup>24,25</sup>

Hypothesized effects of inappropriate habitual  $f_0$  on the formation of phonotraumatic lesions—or as a reaction to these lesions—have also been addressed in the clinical literature.<sup>26</sup> For example, structural changes in the true vocal folds have been attributed to both inappropriately low<sup>27,28</sup> and inappropriately high<sup>29</sup> habitual  $f_0$ . Additionally, whereas some hypothesize that the presence of edema or structural lesions causes a decreased habitual  $f_0$  due to increases in vocal fold mass,<sup>30,31</sup> others hypothesize an increased habitual  $f_0$  due to increased vocal fold stiffness that results from reactive hyperfunction to maintain phonation in the presence of vocal fold pathology.<sup>32</sup>

In a similar vein, empirical studies regarding the association of average  $f_0$  with phonotraumatic lesions have provided inconsistent results using in-clinic, short-duration measurements. Some studies have demonstrated no significant differences in  $f_0$  when comparing pre- and post-treatment recordings,<sup>23,24,33–38</sup> nor between normal subjects and those with phonotraumatic lesions.<sup>35,36</sup> Conversely, others have reported either increased<sup>25,39,40</sup> or decreased<sup>24,41</sup>  $f_0$  values for patients with nodules or polyps relative to normal subjects and/or following surgery.

With the development of ambulatory voice monitors,<sup>42–46</sup> a patient's typical/habitual vocal behaviors related to intensity,  $f_0$ , and amount of voice use or *vocal dose* can potentially be better characterized than in clinic assessments since much longer-term data can be collected as the individuals engage in their activities of daily living. Vocal dose measures have been proposed as a way to indirectly estimate the exposure of vocal fold tissue to mechanical stress during phonatory vibration. Suggested dose measures include the calculation of accumulated phonation time (time dose), number of oscillatory cycles (cycle dose), and parameters that are designed to more comprehensively characterize the total vibration exposure to the vocal folds by combining intensity,  $f_0$ , and phonation time (e.g., distance dose).<sup>2,47</sup> The general concept of vocal dose measures is based on occupational safety standards for exposure of body systems/tissue to external vibratory sources, such as those developed for auditory noise levels and mechanical vibration of hand tools.

To date there have been a few studies that have employed long-term monitoring to examine vocal behaviors potentially related to phonotrauma. These include documenting voice use in high-risk occupations such as teaching<sup>48</sup> and call center operation<sup>49</sup> as well as conflicting reports of phonation times being either correlated<sup>50</sup> or not correlated<sup>49,51</sup> with self-reported vocal complaints or changes in voice quality. Only three ambulatory voice monitoring studies have included small groups of patients who were explicitly diagnosed with phonotraumatic vocal fold lesions.<sup>21,52,53</sup> Masuda and colleagues<sup>21</sup> registered vocal intensity and phonation time from five adult females with vocal fold nodules who exhibited increased vocal intensity and higher phonation time when compared with individuals in low-voice use occupations (office work), while exhibiting similar measures when compared with individuals in high-voice use occupations (school teachers). Nacci and colleagues<sup>52</sup> compared 5 teachers with vocal nodules to 5 teachers with normal voices and reported no differences in overall vocal dose measures but did note different daily trends for  $f_0$  and SPL

(i.e., downward for the nodules group and upward for the normal group). Most recently, Ghassemi and colleagues<sup>53</sup> conducted a pilot study using ambulatory voice data from a small group of nodules patients (12 females) and matched controls (12 females) that was primarily designed as an initial test of using machine learning techniques to analyze the large volumes of data produced by ambulatory voice monitoring. Advanced machine learning algorithms were ultimately able to correctly classify 22 out of 24 participants into nodule and control groups using vocal intensity and f0 features, with salient features arising from extreme regions of the associated data distributions (e.g., 5<sup>th</sup> and 95<sup>th</sup> percentile values).

The purpose of the present study was to determine if there are significant differences between patients with phonotraumatic vocal fold lesions and normal matched (age, sex, and occupation) controls in terms of average voice use. Weeklong ambulatory phonation data were acquired using a smartphone-based ambulatory voice monitor<sup>43</sup> in groups of patients and controls that were large enough to provide adequate power for robust statistical testing. All data were collected as part of a larger, ongoing project aimed at developing ambulatory monitoring of phonation into an effective clinical assessment and biofeedback tool. The governing institutional review board approved all experimental aspects related to the use of human subjects for this study.

## METHOD

### Participant Recruitment

Thirty-five female patients with vocal fold nodules or polyps were recruited through sequential convenience sampling. 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 phonotraumatic vocal fold lesions.<sup>24,54</sup> Diagnoses were based on a comprehensive team evaluation (laryngologist and speech-language pathologist) at the MGH Voice Center that included 1) the collection of a complete case history, 2) endoscopic imaging of the larynx, 3) application of the Voice-Related Quality of Life (V-RQOL) questionnaire,<sup>55</sup> 4) an auditory-perceptual evaluation using the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V),<sup>56</sup> and 5) aerodynamic and acoustic assessment of vocal function. A control subject with no history of voice disorders was matched to each patient according to approximate age ( $\pm$  5 years), sex, and occupation. The normal vocal status of all control participants was verified via interview and a laryngeal stroboscopic examination. During the interview, the matched-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 matched-control candidate indicated voice difficulties or demonstrated a non-normal voice quality, they were excluded from study enrollment and did not undergo a laryngeal stroboscopic examination.

Table 1 lists the occupations and diagnoses of the participants in the study. All participants were engaged in occupations considered to be at a higher-than-normal risk for developing a voice disorder.<sup>57</sup> The majority of patients (28) were professional, amateur, or student singers; every effort was made to match singers with control subjects in a similar musical genre (classical or non-classical) to account for any genre-specific vocal behaviors. Thirty-

one patient was diagnosed with bilateral vocal fold nodules, and four patients had unilateral vocal fold polyps. The average (standard deviation) age of participants within each group was approximately 23 (7) years.

Table 2 reports subscale scores for the self-reported V-RQOL and clinician-judged CAPE-V ratings for the participants in the patient group. V-RQOL scores are normalized ordinal ratings that lie between 0 and 100, with higher scores indicating a higher quality of life.<sup>55</sup> CAPE-V scores are visual analog scale ratings that range from 0 to 100, with zero indicating normality and 100 indicating extremely severe abnormality of a particular voice quality characteristic.<sup>56</sup> Scores on both perceptual scales indicated that most participants exhibited mild-to-moderate dysphonia, with only a few falling on the very severe end of the scales.

### Data Collection

Ambulatory voice monitoring data were collected using a miniature accelerometer (ACC; model BU-27135, Knowles Electronics, Itasca, IL) positioned on the subglottal neck surface as the phonation sensor and a custom smartphone application (Voice Health Monitor; VHM) as the data acquisition platform.<sup>43</sup> The system recorded the unprocessed ACC signal at an 11,025 Hz sampling rate, 16-bit quantization, and 80-dB dynamic range to obtain frequency content of neck surface vibrations up to 5000 Hz. The VHM application provided a user-friendly interface for starting/stopping recording, daily sensor calibration, and periodic alert capabilities. Detailed specifications of the system have been published.<sup>43</sup>

Figure 1 illustrates the coupling and placement of the ACC sensor connected to the smartphone through an interface circuit. Participants affixed the ACC assembly to their neck using hypoallergenic double-sided tape (Model 2181, 3M, Maplewood, MN). The tape's circular shape and small tab allowed for easy placement and removal of the ACC on the neck skin a few centimeters above the suprasternal notch. The tape was strong enough to hold the silicone pad in place during a full day for the typical user (in rare cases, excessive sweat warranted multiple tape applications).

Participants in the patient group were monitored for one week (seven days) before any surgical and/or therapeutic intervention. Each control participant was also monitored for one full week. Each morning, the VHM application led the participants through a daily calibration sequence to map ACC signal level to acoustic sound pressure level (SPL) recorded by a handheld microphone positioned 15 cm from the lips.<sup>43,58</sup> These daily calibrations allowed for ACC-based estimates of vocal dose measures and provided ongoing verification that the system was operating properly.

### Data Analysis

Traditional measures of the cumulative effects of voice use rely on running estimates of SPL and  $f_0$  from ACC signal segments containing voicing. For this study, the signal was divided into nonoverlapping frames of 50 ms in duration. Each frame was considered voiced if the following criteria were passed: (1) SPL was greater than 45 dB SPL at 15 cm, (2) the first (non zero-lag) peak in the normalized autocorrelation exceeded a threshold of 0.6, (3)  $f_0$  (reciprocal of the time lag of that peak) was between 70 Hz and 1000 Hz, and (4) the ratio of low- to high-frequency energy exceeded 20 dB. These criteria were warranted for robust

voice activity detection to eliminate several types of non-phonatory activity that included inadvertent sensor tapping, clothing rubbing against the sensor, high levels of environmental noise, and electrical interference.

In addition to computing  $f_0$  and SPL time series from each week of data, three cumulative vocal dose measures—phonation time, cycle dose, and distance dose<sup>2</sup>—quantified average voice use for each participant. Phonation time yielded the total duration of segments classified as voiced during the total monitoring time. Cycle dose estimated the number of vocal fold oscillations during the monitored period of time to take into account increased vocal fold vibration for higher  $f_0$  segments. Finally, distance dose estimated the total distance traveled by the vocal folds, combining cycle dose with estimates of vibratory amplitude based on SPL to yield a comprehensive vocal dose measure.

Additionally, attempts were made to characterize vocal load and recovery time by keeping track of the occurrences and durations of contiguous voiced and non-voiced segments.<sup>59</sup> Voiced and non-voiced segment durations were binned into logarithmically spaced ranges from 0.100–0.316 s to 3160–10,000 s, where successively longer-duration segments represented successively higher-level speech segmentals (phoneme level, syllable level, word level, etc., for voiced segments; voiceless consonants, pauses between phrases, etc., for non-voiced segments) up to the longest-duration sung passages and silence periods. These data yielded two types of histograms: 1) “occurrence” histograms of the normalized (per-hour) counts of all contiguous voiced and non-voiced segments within each duration bin and (2) “accumulation” histograms of the total duration (per hour) of all contiguous voiced and non-voiced segments within each duration bin.

### Statistical Analysis

Within-subject univariate summary statistics characterized the distributions of the weeklong SPL and  $f_0$  time series of lengths ranging from 200,000 to over 1,000,000 voiced frames depending on how much participants phonated during their respective weeks. Statistics were computed for mean (SPL only), mode ( $f_0$  only), standard deviation (SD), minimum (5<sup>th</sup> percentile), maximum (95<sup>th</sup> percentile), and range (middle 90 %). The use of trimmed estimators for the minimum, maximum, and range statistics was necessary to handle spurious SPL and  $f_0$  outliers. In the data presented here, SPL distributions tended to be normal (similar mean, median, and mode) and  $f_0$  distributions were often skewed toward lower  $f_0$  values with a long tail toward higher  $f_0$  values (thus, the mode was often more meaningful than the mean to quantify average voice use). The  $f_0$  mode was computed from histograms containing 30 equally spaced frequency bins.

Vocal dose measures were computed both as total accumulated values over the week of each individual, as well as normalized values to account for differences in the total time monitored by each study participant. From the occurrence and accumulation histograms, per-hour counts and durations of voiced and non-voiced segments within each duration bin were recorded for each participant. In addition, mean and SD statistics were computed for the Poisson-like distributions of voiced and non-voiced segment durations.



To take full advantage of the matched patient-control paradigm ( $n = 35$  pairs), paired t-tests were used to assess differences between the summary statistics of average voice use. Paired t-tests parametrically assess the differences between pairs, thereby preventing non-paired individual differences from averaging out paired differences. Kolmogorov-Smirnov (K-S) tests for non-parametric data were used to assess between-group differences in the overall distribution of the average voice use measures. The application of both parametric and non-parametric tests was necessary to handle, respectively, normal and non-normal data distributions across subjects. To mitigate the possibility of false positives, a Bonferroni correction was applied to the baseline alpha level (0.05) for multiple hypothesis tests; therefore, differences were considered statistically significant at p-values less than 0.003. When significance was found, an associated effect size was determined using Cohen's  $d$  (i.e., the difference between the two groups' means divided by their pooled standard deviation). The effect size provided a standardized method to interpret the size of differences between the two groups, with effect sizes less than 0.19 interpreted as small, 0.20 to 0.79 as medium, and 0.80 and greater as large.<sup>60</sup>

## Results

Figure 2 illustrates the voice use profile across one day for one of the patients with vocal fold nodules. The patient was monitored for over 14 hours with 14% of the time classified as containing voicing. Bursts of singing activity are observed during the day with phonation time approaching 45% around 11:15 am, 4:15 pm, and 8:15 pm. Figures 2D and 2E illustrate the typical shapes of SPL and  $f_0$  histograms, respectively. For the day shown, mean SPL was 85.8 dB SPL re 15 cm and  $f_0$  mode was 195.1 Hz. Total cycle dose approached 2 million cycles (130,000 cycles per hour), and total distance dose approached 8,000 meters (530 meters per hour).

Table 3 displays the within-group summary statistics derived from SPL,  $f_0$ , and vocal dose measures for the patient and matched-control groups. All features in Table 3 were tested with the Shapiro-Wilk test to evaluate distributional normality and every feature was not significantly different than a normal distribution ( $p > 0.05$ ). Therefore, all paired t-test comparisons were valid. Most subjects wore the monitoring system for more than 80 hours during the seven days, with four participants logging over 100 hours of data. Paired t-tests only revealed statistically significant differences for three measures of  $f_0$  variability:  $f_0$  SD,  $f_0$  maximum, and  $f_0$  range (medium effect sizes from 0.63 to 0.65). Although these measures of  $f_0$  variability were significantly lower for the patient group compared to the same measures for the matched-control group, it was notable that  $f_0$  mode—the measure typically indicative of average speaking  $f_0$  behavior<sup>26</sup>—was not significantly different between the two groups. All other paired t-test and K-S test statistics comparing average vocal intensity, fundamental frequency, and vocal dose measures (both accumulated and normalized) between the two groups resulted in non-significant differences at the Bonferroni-corrected significance level.

Figure 2 displays occurrence and accumulation histograms for the voiced and non-voiced segments; logarithmic scales on the vertical axes aid in visualizing the rare occurrences of long-duration segments. Although most contiguous voiced segments lasted less than one

second, a substantial percentage of the voiced segments were longer than one second with rarer occurrences of segments longer than 3.16 seconds. Voiced segments in the 10–31.6 s bin were verified (via listening to the ACC waveform) to capture singing by the participants during rehearsals or performances. One participant in the control group (singer) had an exceptional number of these long voiced segments (24 throughout her week).

Comparing patient and control group data within each bin of the histograms in Figure 3, none of the differences between normalized counts or the accumulated durations of the voiced and nonvoiced segments reached statistical significance. Differences were also not statistically significant between patient and control groups using mean and SD statistics of within-subject occurrence and accumulation distributions.

## DISCUSSION

The purpose of this study was to determine if there are significant differences between patients with phonotraumatic vocal fold lesions and normal matched (age, sex, and occupation) controls in terms of average daily voice use and vocal function. This data set provides the most convincing empirical evidence to date that overall average SPL, f0, vocal dose measures, and voiced/non-voiced duration characteristics do not differentiate an individual with a normal voice from one with vocal fold nodules or polyps. Only f0 variability measures (lower SD, maximum f0, and f0 range) yielded discriminatory power with a moderate effect size. The reductions in measures related to long-term f0 variability are likely related to, and therefore generally in agreement with, the finding in many previous reports that patients with nodules demonstrated a decreased ability to reach higher frequencies.<sup>25</sup>

The lack of other significant differences is considered surprising from two points of view: 1) the very presence of the lesions might be expected to more clearly impact vocal function (pitch and loudness) and 2) the results contradict the simple classic view that phonotrauma is primarily associated with increased levels of habitual voice use.

Although clinicians have undoubtedly encountered patients with phonotraumatic lesions who appear to be loud talkers (producing significantly elevated average SPL), laboratory studies have already shown that patients do not necessarily present according to this clinical stereotype.<sup>1</sup> Conversely, attempts to model the effects of nodules on phonation by increasing vocal fold mass and decreasing glottal closure predict a lowering of f0 and SPL compared to healthy vocal folds;<sup>61</sup> however, these modeled effects on f0 and SPL were not reflected in our empirical data. Thus the finding that there are no significant differences in average intensity and average f0 between the pathological and normal groups in this study is clearly at odds with some clinical and model-based hypotheses about phonotraumatic vocal fold lesions. This lack of a difference might be due to compensatory effort by the pathological group to maintain normal values for these parameters in the presence of vocal pathology. Previous laboratory studies have shown that patients with phonotraumatic lesions appear to employ phonatory adjustments that maintain vocal SPL but also increase potential for vocal fold trauma (e.g., elevated maximum flow declination rate or AC flow compared to controls).<sup>1,23</sup> These findings have been viewed as quantitative evidence for the classic



“vicious cycle” that is associated with these disorders; i.e., people with vocal fold nodules or polyps compensate for inefficient vocal physiology by producing progressively damaging levels of physical forces and dynamics. Thus the lack of between group differences in f<sub>0</sub> and SPL does not necessarily mean that the underlying mechanisms for achieving f<sub>0</sub> and SPL are equivalent in these two groups.

It was also surprising that vocal dose measures (phonation time, cycle dose, and distance dose) were statistically indistinguishable between patients and their matched controls according to both parametric and non-parametric paired statistical assessments. One possible explanation is related to the fact that existing vocal dose measures are based on estimates of vocal fold vibration (duration, frequency, and amplitude) and do not yet take into account vocal fold contact (glottal closure) patterns. As such, the averaged measures do not assess the underlying collision forces thought to be vital factors in the development of superficial lamina propria tissue changes/damage.<sup>7,62,63</sup>

The measures based on characterizing phonatory and non-phonatory segment durations (applied for the first time in a patient-versus-control study design) were also applied to this dataset, but with non-discriminatory results. The non-phonatory data can be viewed as approximating the time during which vocal fold tissue is given the opportunity to recover from vibration and collision-related trauma. Thus, in addition to patients not talking more often than their controls, the two groups displayed an approximately equivalent partitioning between phonatory and non-phonatory activity, which suggests that the two groups did not differ with respect to their daily vocal recovery time over the course of the week monitored. Future work calls for the further examination and characterization of voiced and non-voiced segments to more fully explore the role of vocal recovery in the etiology of phonotraumatic lesions. This could include taking into account their time ordering (quantifying the occurrence of voice rest durations given a burst of preceding voicing activity) and/or the use of temporal decay/growth weighting paradigms.<sup>64</sup>

The lack of significant differences in average voice use-related parameters between patients and controls also supports the view that other factors may contribute to the etiology of nodules and polyps.<sup>4</sup> Factors could include sources of tissue irritation that increase susceptibility of the vocal fold mucosa to mechanical trauma (e.g., laryngopharyngeal reflux, upper respiratory infection, etc.) and/or make phonation more difficult, thus triggering the initial patterns of vocal hyperfunction leading to phonotraumatic lesion development and the ensuing “vicious cycle” of increasing compensatory hyperfunction. Pilot work in the use of machine learning approaches to analyze long-term data<sup>53</sup> also points to the need to look beyond average behaviors in these types of patients and to examine the prevalence of more extreme phonatory episodes which could potentially create conditions (e.g., vocal fold edema, hemorrhage, etc.) that trigger chronic hyperfunction.

Unlike most previous approaches to ambulatory voice monitoring that discard the ACC signal after onboard extraction of basic measures, our new approach described here captures and stores the raw ACC waveform. This offers the potential to extract additional information from the ACC signal. For example, laboratory-based measures of relative fundamental frequency (around voice onset and offset for voiceless consonant sounds) have been shown

to be sensitive to the presence of vocal hyperfunction<sup>65</sup> and could potentially be extracted from the ACC signal.<sup>66</sup> Additionally, research groups have demonstrated the feasibility of inverse filtering the ACC signal using a vocal system model<sup>67,68</sup> to estimate aerodynamic measures that have been shown in the laboratory to provide insight into underlying aberrant phonatory mechanisms associated with vocal hyperfunction.<sup>1,69</sup>

It is possible that the mere diagnosis of vocal fold pathology or the fact that the patients were being monitored could have affected their typical behavior, thereby limiting the interpretation of results from any ambulatory monitoring study.<sup>70</sup> Singers may be particularly sensitive to their diagnosis and may reduce their overall voice use and/or show restraint during high-performance periods. However, the majority of subjects reported forgetting that they wore the device during their monitored period. In addition, considering that patients often need extensive vocal rehabilitation over the course of weeks or months, it is doubtful that the hypothesized habitual vocal behaviors associated with phonotraumatic lesions would be modified to any great extent in this ambulatory phonation data set. Future studies could experiment with various monitoring schedules to capture vocal behaviors during different seasons, during vocal performance and non-performance settings, etc.

In patients who already have vocal nodules or polyps, it is not possible to determine which aspects of vocal function/behavior preceded lesion formation (primary vocal hyperfunction) and which are a reaction to the presence of the lesions (reactive vocal hyperfunction)<sup>1,53</sup>, or if in fact there is any difference. In our ongoing work we are attempting to gain further insight into this issue by tracking the voice use of individuals with nodules or polyps before and after treatment—including surgical excision and/or voice therapy. The pre- versus post-surgery comparison (prior to voice therapy) is of particular interest because it provides the opportunity to observe the hyperfunctional behavior that caused the tissue damage without the potentially confounding influence of lesions on vocal function. Consequently, monitoring the behavior of the same subjects after successful behavioral voice therapy has the potential to reveal features related to the progressive improvement of vocal behavior patterns.

## CONCLUSION

Overall, this study indicates that the average voice use and vocal function of patients with phonotraumatic lesions (vocal fold nodules and polyps) does not differ significantly from matched-control subjects in terms of average vocal intensity, fundamental frequency, dose measures, or voiced/non-voiced segment frequency and duration. Only measures of fundamental frequency variability show initial discriminative value. More refined ambulatory measurements of hyperfunctional phonatory mechanisms, along with the examination of other potential contributing etiologic factors, are needed to improve the understanding of causative or associative risk factors for common phonotraumatic vocal fold lesions.

## Acknowledgments

Statement of Support: The authors acknowledge the contributions of R. Petit for aid in designing and programming the smartphone application. This work was supported by the Voice Health Institute and the National Institutes of

Health (NIH) National Institute on Deafness and Other Communication Disorders under Grant R33 DC011588. The paper's contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

## References

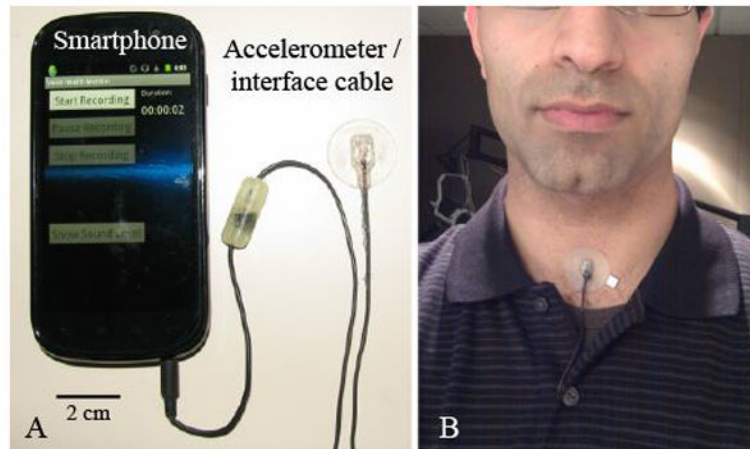
1. Hillman RE, Holmberg EB, Perkell JS, Walsh M, Vaughan C. Objective assessment of vocal hyperfunction: An experimental framework and initial results. *J Speech Hear Res.* 1989; 32(2):373–392. [PubMed: 2739390]
2. Titze IR, Švec JG, Popolo PS. Vocal dose measures: Accumulated vibration exposure in vocal fold tissues. *J Speech Lang Hear Res.* 2003; 46(4):919–932. [PubMed: 12959470]
3. Czerwonka L, Jiang JJ, Tao C. Vocal nodules and edema may be due to vibration-induced rises in capillary pressures. *Laryngoscope.* 2008; 118(4):748–752. [PubMed: 18300711]
4. Karkos PD, McCormick M. The etiology of vocal fold nodules in adults. *Curr Opin Otolaryngol Head Neck Surg.* 2009; 17(6):420–423. [PubMed: 19730264]
5. Kunduk M, McWhorter AJ. True vocal fold nodules: the role of differential diagnosis. *Curr Opin Otolaryngol Head Neck Surg.* 2009; 17(6):449–452. [PubMed: 19779347]
6. Behrman A, Rutledge J, Hembree A, Sheridan S. Vocal hygiene education, voice production therapy, and the role of patient adherence: a treatment effectiveness study in women with phonotrauma. *J Speech Lang Hear Res.* 2008; 51(2):350–366. [PubMed: 18367682]
7. Casper JK, Murry T. Voice therapy methods in dysphonia. *Otolaryngol Clin North Am.* 2000; 33(5): 983–1002. [PubMed: 10984765]
8. Roy N, Weinrich B, Gray SD, Tanner K, Stemple JC, Sapienza CM. Three treatments for teachers with voice disorders: A randomized clinical trial. *J Speech Lang Hear Res.* 2003; 46(3):670–688. [PubMed: 14696994]
9. Thomas LB, Stemple JC. Voice therapy: Does science support the art. *Commun Disord Rev.* 2007; 1(1):49–77.
10. Sapienza CM, Crandell CC, Curtis B. Effects of sound-field frequency modulation amplification on reducing teachers' sound pressure level in the classroom. *J Voice.* 1999; 13(3):375–381. [PubMed: 10498053]
11. Roy N, Weinrich B, Gray SD, et al. Voice amplification versus vocal hygiene instruction for teachers with voice disorders: A treatment outcomes study. *J Speech Lang Hear Res.* 2002; 45(4): 625–638. [PubMed: 12199394]
12. Morrow SL, Connor NP. Comparison of voice-use profiles between elementary classroom and music teachers. *J Voice.* 2011; 25(3):367–372. [PubMed: 20359861]
13. Jónsdóttir VI, Boyle BE, Martin PJ, Sigurdardóttir G. A comparison of the occurrence and nature of vocal symptoms in two groups of Icelandic teachers. *Logoped Phonatr Vocol.* 2002; 27:98–105.
14. Leonard R. Voice therapy and vocal nodules in adults. *Curr Opin Otolaryngol Head Neck Surg.* 2009; 17(6):453–457. [PubMed: 19741535]
15. Behlau M, Oliveira G. Vocal hygiene for the voice professional. *Curr Opin Otolaryngol Head Neck Surg.* 2009; 17(3):149–154. [PubMed: 19342952]
16. Behlau M, Zambon F, Madazio G. Managing dysphonia in occupational voice users. *Curr Opin Otolaryngol Head Neck Surg.* 2014; 22(3):188–94. [PubMed: 24670489]
17. Roy N, Merrill RM, Thibeault S, Parsa RA, Gray SD, Smith EM. Prevalence of voice disorders in teachers and the general population. *J Speech Lang Hear Res.* 2004; 47(2):281–293. [PubMed: 15157130]
18. Behlau M, Zambon F, Guerrieri AC, Roy N. Epidemiology of voice disorders in teachers and nonteachers in Brazil: Prevalence and adverse effects. *J Voice.* 2012; 26(5):e9–e18. [PubMed: 22516316]
19. Ohlsson A-C, Andersson EM, Södersten M, Simberg S, Barregård L. Prevalence of voice symptoms and risk factors in teacher students. *J Voice.* 2012; 26(5):629–634. [PubMed: 22578438]
20. Boone, R.; McFarlane, S. *The Voice and Voice Therapy.* 6. Needham Heights, MA: Allyn & Bacon; 2000.

21. Masuda T, Ikeda Y, Manako H, Komiyama S. Analysis of vocal abuse: fluctuations in phonation time and intensity in 4 groups of speakers. *Acta Otolaryngol.* 1993; 113(4):547–552. [PubMed: 8379311]
22. Levendoski EE, Leydon C, Thibeault SL. Vocal fold epithelial barrier in health and injury: A research review. *J Speech Lang Hear Res.* 2014; 57(5):1679–1691. [PubMed: 24686981]
23. Holmberg EB, Doyle P, Perkell JS, Hammarberg B, Hillman RE. Aerodynamic and acoustic voice measurements of patients with vocal nodules: variation in baseline and changes across voice therapy. *J Voice.* 2003; 17(3):269–282. [PubMed: 14513951]
24. Hirano M, Tanaka S, Fujita M, Terasawa R. Fundamental frequency and sound pressure level of phonation in pathological states. *J Voice.* 1991; 5(2):120–127.
25. Zeitels SM, Hillman RE, Desloge R, Mauri M, Doyle PB. Phonomicrosurgery in singers and performing artists: Treatment outcomes, management theories, and future directions. *Ann Otol Rhinol Laryngol.* 2002; 111(12):21–40.
26. Roy N, Hendarto H. Revisiting the pitch controversy: Changes in speaking fundamental frequency (SFF) after management of functional dysphonia. *J Voice.* 2005; 19(4):582–591. [PubMed: 16301104]
27. Van Riper, C.; Irwin, J. *Voice and Articulation.* 1. Englewood Cliffs, NJ: Prentice-Hall; 1958.
28. Wilson K. Voice re-education of adults with vocal nodules. *Arch Otol Rhinol Laryngol.* 1962; 76:68–73.
29. Cooper M. Spectrographic analyses of fundamental frequency of hoarseness before and after vocal rehabilitation. *J Speech Hear Disord.* 1974; 40(2):286–297. [PubMed: 4415856]
30. Morrison, M.; Rammage, L. *The Management of Voice Disorders.* 1. San Diego, CA: Singular Publishing Group; 1994.
31. Altman KW. Vocal fold masses. *Otolaryngol Clin North Am.* 2007; 40(5):1091–1108. [PubMed: 17765697]
32. Aronson, AE. *Clinical Voice Disorders.* 3. New York, NY: Thieme Stratton; 1990.
33. Languaite JK, Waldrop WF. Acoustic analysis of fundamental frequency of voices before and after voice therapy. *Folia Phoniatri Logop.* 1964; 16(3):183–192.
34. Hufnagle J, Hufnagle K. An investigation of the relationship between speaking fundamental frequency and vocal quality improvement. *J Commun Disord.* 1984; 17(2):95–100. [PubMed: 6725628]
35. Casado Morente JC, Adrian Torres JA, Conde Jimenez M, et al. Objective study of the voice in a normal population and in dysphonia caused by nodules and vocal polyps. *Acta Otorrinolaringol Esp.* 2001; 52(6):476–482. [PubMed: 11692962]
36. Chernobelsky SI. The treatment and results of voice therapy amongst professional classical singers with vocal fold nodules. *Logoped Phoniatri Vocol.* 2007; 32(4):178–184.
37. Halawa WE, Freire AFR, Muñoz IV, Pérez SS. Assessment of effectiveness of acoustic analysis of voice for monitoring the evolution of vocal nodules after vocal treatment. *Eur Arch Otorhinolaryngol.* 2014; 271(4):749–56. [PubMed: 24061570]
38. Štajner-Katušić S, Horga D, Zrinski KV. A longitudinal study of voice before and after phonosurgery for removal of a polyp. *Clin Linguist Phon.* 2008; 22(10–11):857–863. [PubMed: 18608247]
39. Schindler A, Mozzanica F, Maruzzi P, Atac M, De Cristofaro V, Ottaviani F. Multidimensional assessment of vocal changes in benign vocal fold lesions after voice therapy. *Auris Nasus Larynx.* 2013; 40(3):291–297. [PubMed: 23122629]
40. Valadez V, Ysunza A, Ocharan-Hernandez E, Garrido-Bustamante N, Sanchez-Valerio A, Pamplona MC. Voice parameters and videonasolaryngoscopy in children with vocal nodules: a longitudinal study, before and after voice therapy. *Int J Pediatr Otorhinolaryngol.* 2012; 76(9): 1361–1365. [PubMed: 22770591]
41. Toran KC, Lal BK. Objective voice analysis for vocal polyps following microlaryngeal phonosurgery. *Kathmandu Univ Med J.* 2010; 8(2):185–189.
42. Popolo PS, Švec JG, Titze IR. Adaptation of a Pocket PC for use as a wearable voice dosimeter. *J Speech Lang Hear Res.* 2005; 48(4):780–791. [PubMed: 16378473]

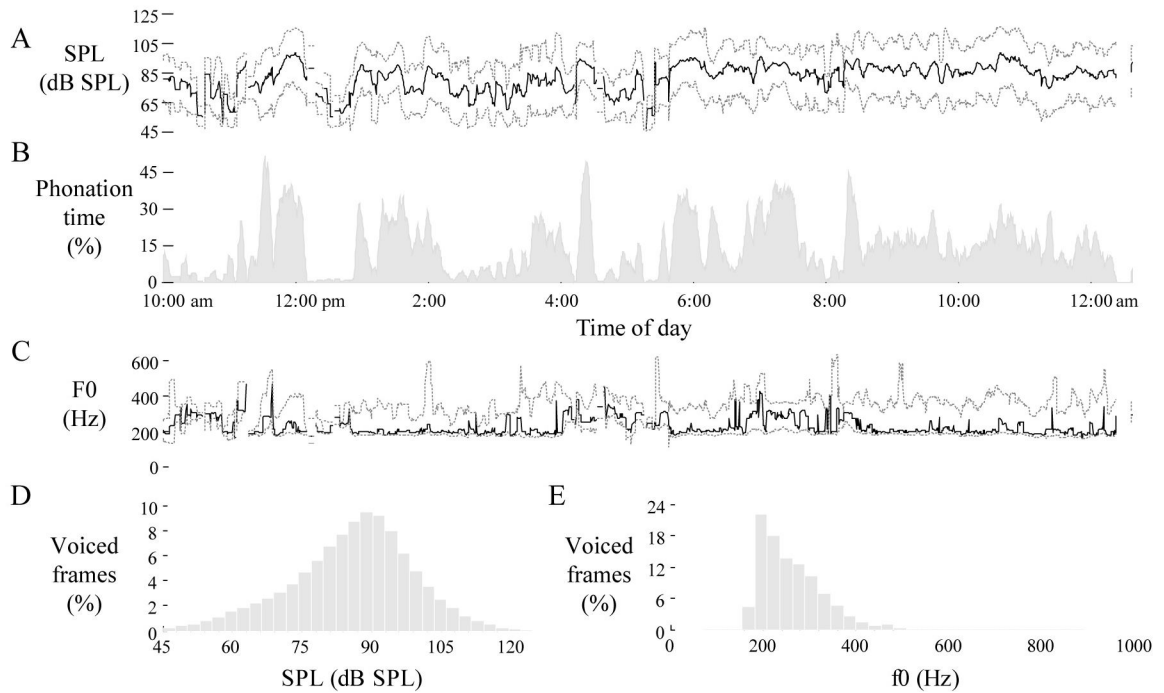
43. Mehta DD, Zañartu M, Feng SW, Cheyne HA II, Hillman RE. Mobile voice health monitoring using a wearable accelerometer sensor and a smartphone platform. *IEEE Trans Biomed Eng.* 2012; 59(11):3090–3096. [PubMed: 22875236]
44. Carullo A, Vallan A, Astolfi A. Design issues for a portable vocal analyzer. *IEEE Trans Instrum Meas.* 2013; 62(5):1084–1093.
45. Cheyne HA, Hanson HM, Genereux RP, Stevens KN, Hillman RE. Development and testing of a portable vocal accumulator. *J Speech Lang Hear Res.* 2003; 46(6):1457–1468. [PubMed: 14700368]
46. Van Stan JH, Gustafsson J, Schalling E, Hillman RE. Direct comparison of three commercially available devices for voice ambulatory monitoring and biofeedback. *Perspect Voice Voice Disord.* 2014; 24(2):80–86.
47. Rantala L, Vilkmán E. Relationship between subjective voice complaints and acoustic parameters in female teachers' voices. *J Voice.* 1999; 13(4):484–495. [PubMed: 10622515]
48. Hunter EJ, Titze IR. Variations in intensity, fundamental frequency, and voicing for teachers in occupational versus nonoccupational settings. *J Speech Lang Hear Res.* 2010; 53(4):862–876. [PubMed: 20689046]
49. Cantarella G, Iofrida E, Boria P, et al. Ambulatory phonation monitoring in a sample of 92 call center operators. *J Voice.* 2013; 28(3):393.e1–393.e6. [PubMed: 24321583]
50. Lyberg Åhlander V, García DP, Whitling S, Rydell R, Löfqvist A. Teachers' voice use in teaching environments: A field study using Ambulatory Phonation Monitor. *J Voice.* 2014 in press.
51. Rantala L, Vilkmán E, Bloigu R. Voice changes during work: Subjective complaints and objective measurements for female primary and secondary schoolteachers. *J Voice.* 2002; 16(3):344–355. [PubMed: 12395987]
52. Nacci A, Fattori B, Mancini V, et al. The use and role of the Ambulatory Phonation Monitor (APM) in voice assessment. *Acta Otorhinolaryngol Ital.* 2013; 33(1):49–55. [PubMed: 23620641]
53. Ghassemi M, Van Stan JH, Mehta DD, Zañartu M, Cheyne HA II, Hillman RE. Learning to detect vocal hyperfunction from ambulatory neck-skin acceleration features: Initial results for vocal fold nodules. *IEEE Trans Biomed Eng.* 2014; 61(6):1668–1675. [PubMed: 24845276]
54. Ward PD, Thibeault SL, Gray SD. Hyaluronic acid: Its role in voice. *J Voice.* 2002; 16(3):303–309. [PubMed: 12395982]
55. Hogikyan ND, Sethuraman G. Validation of an instrument to measure voice-related quality of life (V-RQOL). *J Voice.* 1999; 13(4):557–569. [PubMed: 10622521]
56. Kempster GB, Gerratt BR, Verdolini Abbott K, Barkmeier-Kraemer J, Hillman RE. Consensus Auditory-Perceptual Evaluation of Voice: Development of a standardized clinical protocol. *Am J Speech Lang Pathol.* 2009; 18(2):124–132. [PubMed: 18930908]
57. Verdolini K, Ramig LO. Review: Occupational risks for voice problems. *Logoped Phonatr Vocol.* 2001; 26(1):37–46.
58. Švec JG, Titze IR, Popolo PS. Estimation of sound pressure levels of voiced speech from skin vibration of the neck. *J Acoust Soc Am.* 2005; 117(3):1386–1394. [PubMed: 15807026]
59. Titze IR, Hunter EJ, Švec JG. Voicing and silence periods in daily and weekly vocalizations of teachers. *J Acoust Soc Am.* 2007; 121(1):469–478. [PubMed: 17297801]
60. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences. 2.* Hillsdale: Lawrence Erlbaum; 1988.
61. Hanson HM, Stevens KN, Kuo H-KJ, Chen MY, Slifka J. Towards models of phonation. *J Phon.* 2001; 29(4):451–480.
62. Verdolini K, Druker DG, Palmer PM, Samawi H. Laryngeal adduction in resonant voice. *J Voice.* 1998; 12(3):315–327. [PubMed: 9763181]
63. Tao C, Jiang JJ. Mechanical stress during phonation in a self-oscillating finite-element vocal fold model. *J Biomech.* 2007; 40(10):2191–2198. [PubMed: 17187805]
64. Hunter EJ, Titze IR. Quantifying vocal fatigue recovery: Dynamic vocal recovery trajectories after a vocal loading exercise. *Ann Otol Rhinol Otolaryngol.* 2009; 118(6):449–460.

65. Stepp CE, Merchant GR, Heaton JT, Hillman RE. Effects of voice therapy on relative fundamental frequency during voicing offset and onset in patients with vocal hyperfunction. *J Speech Lang Hear Res.* 2011; 54(5):1260–1267. [PubMed: 21498578]
66. Lien Y-AS, Stepp CE. Comparison of voice relative fundamental frequency estimates derived from an accelerometer signal and low-pass filtered and unprocessed microphone signals. *J Acoust Soc Am.* 2014; 135(5):2977–2985. [PubMed: 24815277]
67. Zañartu M, Ho JC, Mehta DD, Hillman RE, Wodicka GR. Subglottal impedance-based inverse filtering of voiced sounds using neck surface acceleration. *IEEE Trans Audio Speech Lang Processing.* 2013; 21(9):1929–1939.
68. Zañartu, M.; Espinoza, V.; Mehta, DD., et al. Toward and objective aerodynamic assessment of vocal hyperfunction using a voice health monitor. 8th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications; Firenze, Italy. 2013.
69. Holmberg EB, Hillman RE, Perkell JS, Gress C. Relationships between intra-speaker variation in aerodynamic measures of voice production and variation in SPL across repeated recordings. *J Speech Hear Res.* 1994; 37(3):484–495. [PubMed: 8084180]
70. Hunter EJ. Teacher response to ambulatory monitoring of voice. *Logoped Phoniatr Vocol.* 2012; 37(3):133–135. [PubMed: 22443101]

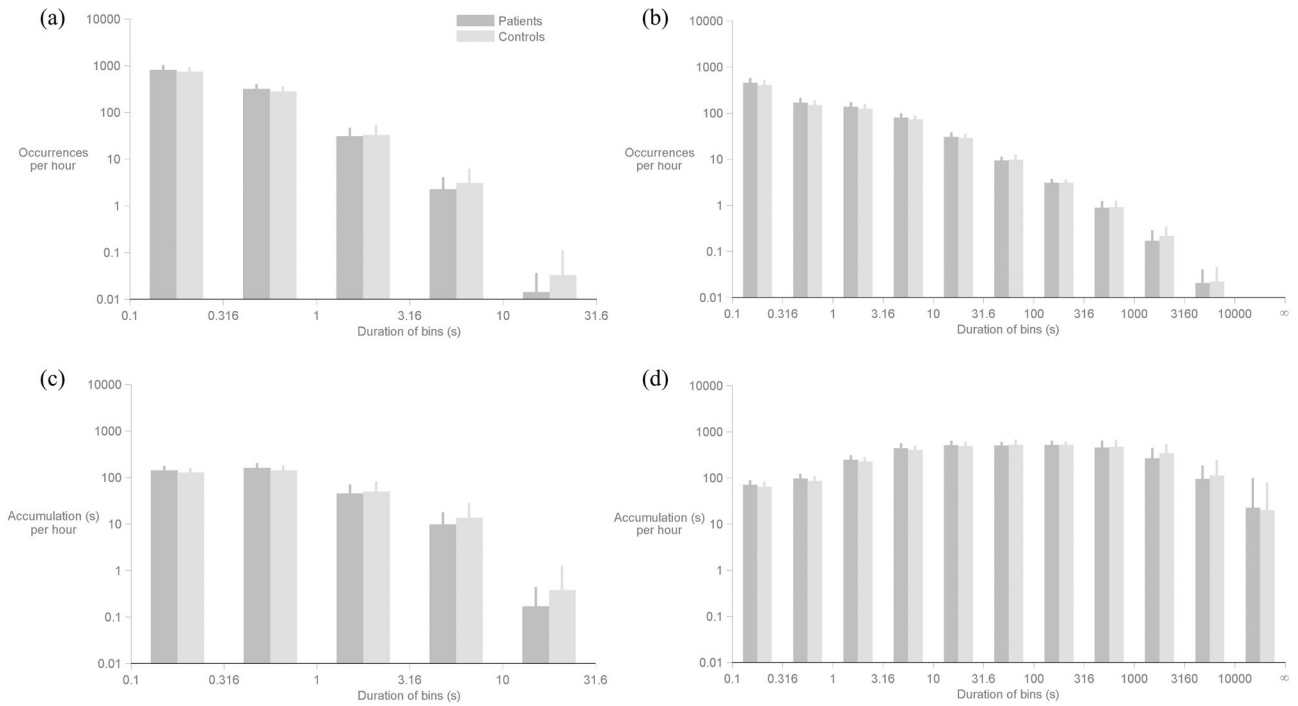




**Figure 1.** Illustration of the accelerometer-based ambulatory voice monitor: (A) Smartphone, accelerometer sensor, and interface cable with circuit encased in epoxy and (B) wired accelerometer mounted on a silicone pad affixed to the anterior neck surface midway between the thyroid prominence and the suprasternal notch.



**Figure 2.** Daylong voice use profile for an 18-year-old voice student with vocal fold nodules. Values are plotted across five-minute sliding windows for (A) sound pressure level (SPL) at 15 cm, (B) phonation time, and (C) fundamental frequency (f0). Dotted lines represent minimum (5<sup>th</sup> percentile) and maximum (95<sup>th</sup> percentile) values. Solid lines represent mean and mode statistics for SPL and f0, respectively. Histograms are also displayed for (D) SPL and (E) f0 as percentages of voiced frames over the entire day.



**Figure 3.** Comparison of occurrence and accumulation histograms between patient (dark gray) and control (light gray) groups. Occurrence histograms are shown separately for (a) voiced and (b) non-voiced segments. Accumulation histograms are also shown separately for (c) voiced and (d) non-voiced segments. Error bars denote one standard deviation above the mean. No statistically significant differences ( $p < 0.003$ ) were found between patient-control pairs within any bin.

**Table 1**

Occupations of all patient and matched-control participants in the study (35 pairs). Diagnoses for the patient group are also listed for each occupation.

Occupation	No. Subject Pairs	Patient Diagnoses
Singer	28	Nodules (25), Polyp (3)
Media relations	1	Nodules
Registered nurse	1	Polyp
Recruiter	2	Nodules
Psychotherapist	1	Nodules
Teacher	2	Nodules

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2**

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. Mean and standard deviation (SD) reported within the patient group (n = 35).

	<b>Mean ± SD</b>
<i>V-RQOL</i>	
Social-Emotional	76.0 ± 20.9
Physical Functioning	77.7 ± 16.1
Total Score	77.0 ± 16.1
<i>CAPE-V</i>	
Overall Severity	27.0 ± 12.4
Roughness	16.5 ± 13.0
Breathiness	14.7 ± 10.6
Strain	16.5 ± 10.7
Pitch	5.9 ± 9.7
Loudness	3.7 ± 8.0

**Table 3**

Group-based values (mean  $\pm$  SD) for voice use summary statistics of sound pressure level (SPL), fundamental frequency ( $f_0$ ), and vocal dose measures for weeklong data collected from the patient and matched-control groups ( $n = 35$  pairs). The first row reports the duration of monitoring during the week in hours/minutes/seconds (hh:mm:ss). Test statistics exhibiting statistically significant differences between patient-control pairs are labeled with an asterisk (\*,  $p < 0.003$ ). Paired t-test statistics are derived from the pairwise comparison of each summary statistic for control values minus their matched patient values.

Voice use summary statistic	Patient group	Paired control group	Paired t-test statistic	KS test statistic
Monitored duration (hh:mm:ss)	80:24:31 $\pm$ 14:49:48	81:32:21 $\pm$ 11:36:35	0.37	0.14
<i>SPL (dB SPL re 15 cm)</i>				
Mean	84.4 $\pm$ 4.5	83.6 $\pm$ 4.8	-0.91	0.11
Standard deviation	12.1 $\pm$ 2.0	12.9 $\pm$ 2.2	1.63	0.23
Minimum	63.4 $\pm$ 5.5	61.6 $\pm$ 5.8	-1.40	0.23
Maximum	103.6 $\pm$ 6.9	104.6 $\pm$ 6.7	0.75	0.17
Range	40.2 $\pm$ 7.3	43.0 $\pm$ 7.9	1.55	0.23
<i><math>f_0</math> (Hz)</i>				
Mode	198.1 $\pm$ 22.6	202.9 $\pm$ 18.1	1.15	0.20
Standard deviation	76.1 $\pm$ 17.6	88.0 $\pm$ 18.9	3.48*	0.31
Minimum	167.5 $\pm$ 17.4	170.9 $\pm$ 14.2	1.10	0.23
Maximum	396.2 $\pm$ 67.1	437.2 $\pm$ 61.8	3.49*	0.34
Range	228.6 $\pm$ 57.9	266.3 $\pm$ 59.8	3.45*	0.31
<i>Phonation time</i>				
Cumulative (hh:mm:ss)	08:01:02 $\pm$ 02:14:43	07:33:39 $\pm$ 02:32:51	-0.87	0.23
Normalized (%)	10.0 $\pm$ 2.3	9.3 $\pm$ 2.7	-1.37	0.20
<i>Cycle dose</i>				
Cumulative (millions of cycles)	7.208 $\pm$ 2.278	7.324 $\pm$ 2.851	0.22	0.17
Normalized (cycles/hr)	89,624 $\pm$ 23,879	89,759 $\pm$ 30,384	0.02	0.14
<i>Distance dose</i>				
Cumulative (m)	27,701 $\pm$ 9,303	27,519 $\pm$ 12,259	-0.09	0.14
Normalized (m/hr)	345.3 $\pm$ 98.1	337.0 $\pm$ 134.5	-0.37	0.17