Effects of Adventitious Acute Vocal Trauma: Relative Fundamental Frequency and Listener Perception

*Elizabeth S. Heller Murray, *Gabrielle L. Hands, *Carolyn R. Calabrese, and *,†Cara E. Stepp, *†Boston, Massachusetts

Summary: Objective. High voice users (individuals who demonstrate excessive or loud vocal use) are at risk for developing voice disorders. The objective of this study was to examine, both acoustically and perceptually, vocal changes in healthy speakers after an acute period of high voice use.

Methods. Members of a university women's volleyball team (n = 12) were recorded a week before (pre) and week after (post) the 10-week spring season; n = 6 control speakers were recorded over the same time period for comparison. Speakers read four sentences, which were analyzed for relative fundamental frequency (RFF). Eight naïve listeners participated in an auditory-perceptual visual sort and rate (VSR) task, in which they rated each voice sample's overall severity and strain.

Results. No significant differences were found as a function of time point in the VSR ratings for the volleyball group. Onset cycle 1 RFF values were significantly lower ($P = 0.04$) in the postrecordings of the volleyball participants compared with prerecordings, but there was no significant difference ($P = 0.20$) in offset cycle 10 RFF values. Receiver operating characteristic analyses indicated moderate sensitivity and specificity of onset cycle 1 RFF for discrimination between the volleyball and control participants. Changes were not apparent in the control group as a function of time for either, onset cycle 1 RFF, offset cycle 10 RFF, or either vocal attribute.

Conclusions. Onset cycle 1 RFF may be an effective marker for detecting vocal changes over an acute high voice use period of time before perceptual changes are noted.

Key Words: High voice use–Relative fundamental frequency–Dysphonia–Acoustic analysis–Auditory-perceptual judgment–Strain.

INTRODUCTION

Voice disorders, which are often caused or exacerbated by improper voice use, can have negative social and emotional consequences.[1] Many individuals who present at voice clinics because of vocal misuse are diagnosed with vocal hyperfunction.[2] Vocal hyperfunction is a "condition of abuse and/or misuse of the vocal mechanism due to excessive and/or imbalanced muscular forces."[3](p373) This presence of heightened muscle tension often causes the voice to be perceived as strained[3] and can be especially problematic for individuals who rely heavily on their voices throughout the day, such as teachers, singers, aerobics instructors, and lawyers.[1,2,4–6] In fact, approximately 30% of the working population in the United States has an occupation that requires a substantial amount of voice use, and approximately 3% of the population has an occupation in which adequate vocal abilities are important for public safety.[7] High voice users are at an increased risk for developing a voice disorder.[1,2,4–6] Much of the previous work has focused on identifying individuals who are at high risk for developing voice disorders, but less work has examined the vocal changes after acute periods of such high voice use (ie, excessive or loud vocal use over a period of time). The goal of this study was to examine acoustic and perceptual changes in the vocal characteristics of members of a university women's volleyball team, who display a unique model of high voice use over a relatively short period of time.

Teaching is a high voice use profession that has received a large amount of attention with respect to the risk of developing voice disorders.[1,2,4,5,8–12] Specifically, teachers have been shown to have a higher risk for experiencing voice disorders than nonteachers,[4] and many teachers report that their profession has an adverse impact on their voices.[12] In addition to the high vocal load required for teaching, the suboptimal acoustics of the classroom can also have a detrimental impact on teachers' voices. An increase in environmental noise has been associated with an increase in vocal symptoms, as teachers are required to speak louder to be heard over the noise.[8,10] Not surprisingly, in comparison with classroom teachers, physical education teachers have a further increased probability of developing a voice disorder, even independent of gender, age, and daily number of teaching hours.[12] The indoor physical education settings have a substantial amount of environmental noise, over which physical education teachers are required to constantly project their voices to conduct their classes.[13] One study following physical education student teachers across a semester found that ratings of voice quality and vocal fatigue were both increased in the middle and the end of the semester compared with the beginning.[14] Like physical education teachers, aerobics instructors also conduct classes in loud environments and have reported experiencing both acute and chronic vocal difficulties.[15,16] A survey of aerobics and group fitness instructors revealed an association between duration of exercise instruction and vocal problems.[17] It is

Accepted for publication April 8, 2015.
From the *Department of Speech, Language, and Hearing Sciences, Boston University, Boston, Massachusetts; and the †Department of Biomedical Engineering, Boston University, Boston, Massachusetts.
Address correspondence and reprint requests to Elizabeth S. Heller Murray, Department of Speech, Language, and Hearing Sciences, Boston University, Boston, MA. E-mail: ehmurray@bu.edu
Journal of Voice, Vol. 30, No. 2, pp. 177-185
0892-1997/$36.00

http://dx.doi.org/10.1016/j.jvoice.2015.04.005

hypothesized that these instructors are at risk for developing voice disorders due to the necessity to raise their voices, potentially in an inefficient manner, while teaching their classes.[18]

The participants in sporting events are a group of high voice users who have received less attention in the literature. Like teachers and instructors, members of athletic teams must regularly communicate in high noise environments to coordinate game play. In addition, some sporting activities, such as a tennis serve or a golf drive, are often accompanied by repetitive yelling and/or grunting, and this forceful adduction of the vocal folds places a large demand on the vocal mechanism that occurs over an acute period of time in a repetitive fashion.[1,19] For example, club tennis players may have a higher concentration of voice use during their tennis seasons than they do during other parts of the year. Because of their risk for developing voice disorders, cheerleaders are one group of athletes who have received attention in the literature. Surveys of high-school cheerleaders noted that 32–86% reported difficulty with their voices, ranging from mild dysphonia to aphonia at the time of the cheerleading event.[20] Relative to their peers, cheerleaders are more likely to experience chronic vocal problems and to be diagnosed with voice disorders.[21]

The benefit of examining athletes' vocal changes rests in the acute nature of their vocal trauma. Athletes who use their voices during sporting events have temporary but consistent patterns of voicing. This provides an advantage over examining teachers, who may exhibit more chronic patterns of potential vocal misuse and/or abuse. The present study examined members of a university club sports team before and after an athletic season, allowing examination of potential changes in acoustic and perceptual measures over a short 10-week time period. Acoustic analysis focused on changes in relative fundamental frequency (RFF), which has been shown to correlate with subjective ratings of strain[22,23] and has been implicated in individuals with voice disorders to be associated with vocal hyperfunction.[24] RFF is measured from a vowel-voiceless consonant-vowel speech sample, focusing on the 10 cycles of the vowel waveform directly preceding the voiceless consonant (offset) and the 10 cycles of the vowel waveform directly following the voiceless consonant (onset). Previous research has indicated that the 10th offset cycle (offset cycle 10) and first onset cycle (onset cycle 1) RFF values in individuals with vocal hyperfunction are lower than healthy controls.[23,24] However, previous work has not examined the utilization of RFF to detect small changes in the voices of participants undergoing high voice use. Perceptual assessment will focus on whether listeners can detect changes in the overall severity and strain. Thus, this study will examine both RFF and subjective impressions of overall severity and strain in individuals before (pre) and after (post) a 10-week indoor volleyball season.

We hypothesized that, because of the high voice use that occurred during the volleyball season, (1) the offset cycle 10 and onset cycle 1 of players would have a lower RFF in the recordings taken after the season than in recordings taken before the season and (2) listeners would rate the vocal samples recorded from players after the season as having higher overall severity and strain than those recorded before the season.

METHODS

Participants

All participants completed written consent in compliance with the Boston University Institutional Review Board and were compensated for their participation.

Speakers. Participants were 18 young adults (15 women); all were native speakers of American English. Twelve participants (volleyball group) were members of the Boston University women's volleyball club team (mean [M], 20.0 years; standard deviation [SD], 1.5 years; 11 players and 1 coach). One speaker reported seeing a speech-language pathologist as a child because of difficulty saying the phoneme /r/. Another speaker reported that it was previously recommended that she receives an evaluation for a possible voice disorder, but as her voice did not bother her, she did not follow the recommendation. As neither of these participants had a diagnosed voice disorder, they were included in the analyses. No other participants reported any speech, language, or hearing disorders. All participants in the volleyball group were recorded the week before the beginning of the spring volleyball season (pre) and returned approximately 1 week after the completion of the 10-week season (post; M, 81.8 days; SD, 9.2 days). Although many speakers in the volleyball group reported experiencing episodes of acute phonotrauma (eg, severe hoarseness and/or aphonia) during the period between the prerecordings and postrecordings, no speakers were experiencing these symptoms at the time of either recording. For qualitative comparison, six additional individuals (three women; M, 20.8 years; SD, 1.5 years) who were not participants in club sports formed a control group and were also recorded over the same time period as the volleyball season. They returned for a postrecording an average of 85.0 days after the prerecording (SD, 17.4 days).

Listeners. Eight young adults (four women; M, 19.8 years; SD, 6.4 years) acted as listeners. The listeners were native speakers of American English and reported no prior history of speech, language, or hearing disorders. Clinical training for perceptual judgment of vocal qualities can result in listeners who have their own approaches to rating voices because of their own personal experiences, and training with previous research indicating that the perceptual judgments between naïve raters are more consistent as compared with raters who are experienced listeners.[25] Therefore, listeners who had no prior experience with or coursework in voice disorders, formal exposure to individuals with voice disorders, or experience using rating scales for judging dysphonia were recruited. All listeners passed a hearing screening at 25 dB for the frequencies 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz in both ears.

Stimuli collection and preparation

The speakers were recorded reading four sentences at two time points (pre and post), for a total of 144 speech samples (18 speakers $\times$ 4 sentences $\times$ 2 time points = 144 speech samples). The speech samples were recorded in a sound-treated room using a headset microphone (Shure model WH20, Niles, IL) placed at 45° angle from the mouth, with one of two soundcards, a PreSonus Firepod FP10 (PreSonus Audio Electronics, Inc., Baton Rouge, LA) or a Komplete Audio 6 (Native Instruments, Los Angeles, CA). All samples were recorded at a sampling frequency of 44 100 Hz using *Audacity* (Audacity Software, Pittsburgh, PA).[26] After the acquisition of all the speech samples, nine speech samples were discarded because of misarticulations. The remaining 135 speech samples were normalized by peak amplitude using *MATLAB* (MathWorks, Natick, MA)[27] and used as stimuli for the listening task.

Listening task

Listeners participated in an auditory-perceptual task, rating overall severity of dysphonia and strain using a visual sort and rate (VSR) paradigm.[28] During the VSR task, listeners were shown sets of eight or nine dots on a screen, with each dot corresponding to a speaker saying one sentence (Figure 1, *upper*). Participants were instructed to listen to all the speech samples and drag the dots vertically along a scale from 0 to 100, with 0 representing the least severe (or strained) voice they could imagine and 100 representing the most severe (or strained) voice they could imagine. Listeners were encouraged to first sort all the dots into the general area they believed the speech sample belonged, facilitating a natural grouping of speech samples with similar severities. Then listeners were asked to rate all the speech samples by moving the dots vertically along the 0–100 scale (Figure 1, *lower*).

In preparation for the auditory-perceptual task, four pilot listeners who had experience listening to disordered voices rated the overall severity and strain for each speech sample using the

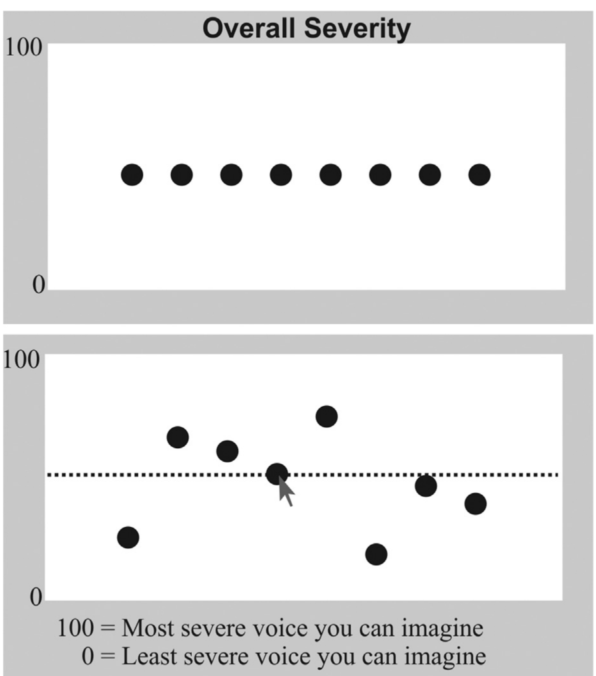

FIGURE 1. Example auditory-perceptual assessment of overall severity using a visual sort and rate task. *Upper*: Initial presentation of the eight sound clips with the vertical scale from 0 to 100. *Lower*: Listeners clicked on the dots to listen to the speech samples and then sorted and rated the dots along the 0 to 100 scale to indicate the severity of the speech samples.

Consensus Auditory-Perceptual Evaluation of Voice.[29] These preliminary ratings ranged from 0.5 to 18.6 for overall severity and 0 to 15.3 for strain. Using these preliminary ratings for overall severity, the speech samples were divided into three groups, characterized qualitatively as (1) 67 mild overall severity samples (range 0.5–3.9), (2) 61 mild-moderate overall severity samples (range 4.0–9.8), and (3) 34 moderate overall severity samples (range 10.0–18.6). To assess intrarater reliability in the perceptual task, 20% of the speech samples were duplicated, resulting in a total of 162 speech samples (135 speech samples + 27 speech samples repeated for reliability). The 162 speech samples were then divided into 20 sets to be used for the overall severity listening task: 18 sets of eight speech samples and two sets of nine speech samples. Sets were designed so that each one would mirror the qualitative overall severity distributions (mild, mild-moderate, moderate) of the entire group. This same procedure was followed using the pilot strain ratings to obtain 20 sets for the strain listening task. Similar to overall severity, the strain speech samples were divided into three groups, characterized qualitatively as (1) 79 mild strain samples (range 0–1.9), (2) 56 mild-moderate strain samples (range 2.0–5.9), and (3) 27 moderate strain samples (range 6.3–15.3).

On constructing of the sets, the perceptual experiment was conducted. Eight listeners were given operational definitions of the two vocal attributes, overall severity of dysphonia and strain. Overall severity was defined as “a comprehensive measure of how good or poor the voice is”[30] (p3017) and strain was defined as “the amount of vocal effort perceived in the voice sample.”[31] (p1001) To familiarize participants to the two vocal attributes before the listening task, listeners were provided with exemplars[32] of mild and severe speech samples for both overall severity and strain. All exemplars and the VSR stimuli were presented at a comfortable loudness through headphones (Sennheiser model HD280 pro, Old Lyme, CT). After listening to the exemplars, listeners participated in the VSR task[28] using the sets created from the pilot listeners. Participants could listen to the speech samples as many times as they felt necessary and could take a break at any point if they needed to. The presentation of the overall severity sets and the strain sets was counterbalanced, with each group of sets taking approximately 30–45 minutes. Overall severity ratings and strain ratings for each speech sample were averaged across all eight listeners, resulting in one overall severity rating and one strain rating per speech sample. Pearson correlation coefficients were calculated to determine a measure of listener intrarater reliability for both overall severity (average r, 0.67; range r, 0.37–0.93) and for strain (average r, 0.55; range r, 0.23–0.79). A type (2, k) interclass correlation[33] was calculated as a measure of interrater reliability; it was 0.67 for overall severity and 0.66 for strain.

Acoustic analysis

The RFF was measured for each speech sample. An RFF instance is defined as a vowel, followed by a voiceless consonant, followed by a vowel (Figure 2). RFF analysis can be conducted on a variety of voiceless consonants; however, previous work has indicated that sentences with the /ʃ/ and /f/ phonemes

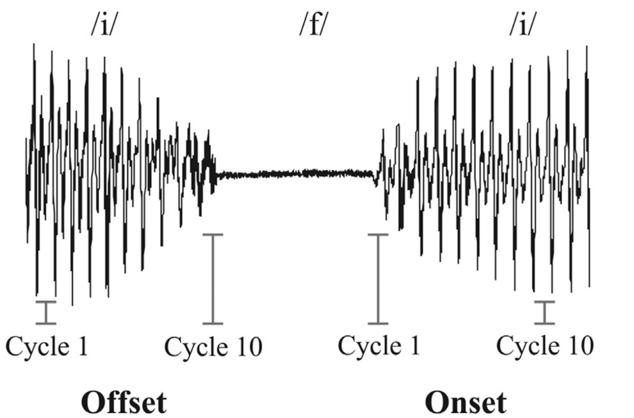

FIGURE 2. Waveform of the speech segment /ifi/, with the reference cycles (offset cycle 1 and onset cycle 10) and the cycles used for analysis (offset 10 and onset cycle 1) for relative fundamental frequency (RFF) indicated.

have the lowest level of intraspeaker variability, and therefore, those phonemes are the most appropriate to use for RFF estimation.[34] Thus, the four sentences that were chosen as stimuli each contained three RFF instances of interest: two sentences had RFF instances with the center phoneme /ʃ/ and two had RFF instances with the center phoneme /f/ (Table 1). Averaging across a minimum of six RFF instances is necessary to provide a stable estimation of RFF.[34] Therefore, this study examined 12 RFF instances per speaker at each time point (three RFF instances from each of the four sentences).[22] Analysis of RFF was consistent with previous work[34] and was conducted by the first author (E.H.M.). The acoustic waveform of the RFF instance was viewed in *Praat* (Praat Software, Amsterdam, the Netherlands),[35] and the 10 vocal cycles of the vowel preceding the voiceless consonant (offset) and the 10 vocal cycles of the vowel following the voiceless consonant (onset) were selected (Figure 2).

For each of those 20 vocal cycles, the instantaneous fundamental frequency was calculated by taking the inverse of the period and converting it to semitones (ST) relative to the reference frequency. This conversion to ST was accomplished using

TABLE 1.
Stimuli With RFF Instances Bolded

Stimuli	International Phonetic Alphabet (IPA) Transliteration
W**e sho**wed Nell m**y shi**ny n**ew shoe** bin	/iʃoʊ/, /aɪʃaɪ/, /uʃu/
The d**ew shi**mmered over m**y shi**ny bl**ue she**ll again	/uʃI/, /aɪʃaɪ/, /uʃɛ/
Nell**y fou**nd n**ew fa**bric while R**ay fe**ll down	/ifaʊ/, /ufæ/, /eɪfɛ/
Only w**e fee**l you d**o fai**l in n**ew fa**llen dew	/ifi/, /ufe/, /ufɔ/

Equation 1,[36] in which f is the frequency of the measured cycle and f_{ref} is the reference frequency. As in previous studies, the reference frequency selected was the cycle furthest away from the voiceless consonant (eg, offset cycle 1 for all offset cycles or onset cycle 10 for all onset cycles), as those cycles were most likely to capture a steady state portion of the vowel. These reference frequencies were used to normalize the values, to allow for comparison across individuals with different base fundamental frequencies. RFF instances were excluded if there were glottalizations or less than 10 vocal cycles. Additionally, RFF instances were excluded if the RFF values for offset cycle 2 and onset cycle 9 (the vocal cycles directly adjacent to the reference cycle) had an absolute value >0.8 ST, indicating that the reference cycle was not located in a steady state portion of the surrounding voicing. After exclusions, there were an average of 9.8 RFF instances per participant, per time period, which were used for data analysis. To calculate a Pearson correlation coefficient for both intrarater and interrater reliability of RFF, the first author (E.H.M.) repeated 20% of the samples (intrarater, $r = 0.97$) 1 month later, and the third author (C.R.C.) analyzed the same speech samples (interrater, $r = 0.91$).

$$\mathrm{RFF(ST)} = 39.86 \times \log_{10}\left(\frac{f}{f_{ref}}\right) \qquad \text{Equation 1}$$

Statistical analysis

Statistical analyses were conducted in *Minitab* (Minitab Inc., State College, PA).[37] RFF changes in the control group, which were obtained over the same time period as the volleyball group, were examined qualitatively. To examine whether listeners perceived changes in vocal quality from the prerecordings to the postrecordings for the speakers in the volleyball group, two paired sample t tests were conducted on the VSR ratings, one for overall severity and one for strain. Similarly, to examine the changes in RFF values from prerecordings to postrecordings in the members of the volleyball group, two paired sample t tests were conducted on RFF offset cycle 10 and onset cycle 1 values, the two cycles furthest away from the reference cycles (the steady state portion of the vowel), which have been previously shown to differentiate between individuals with vocal hyperfunction and healthy controls.[23,24]

Receiver operating characteristic (ROC) curves were generated from four different measures to look at how effectively changes (post-pre) could be used to correctly classify participants as members of either the volleyball group and the control group: (1) RFF values for offset cycle 10, (2) RFF values for onset cycle 1, (3) VSR ratings for overall severity, and (4) VSR ratings for strain. ROC curves for RFF measures were generated by finding the sensitivity and specificity for thresholds of RFF changes (post-pre) between −4 ST and 3 ST at every 0.001 ST. The ROC curves that used VSR ratings were generated by finding the sensitivity and specificity thresholds of perceptual changes (post-pre) between −8 and 13, using a step size of 0.001. Sensitivity was defined as the true positive rate, or the proportion of people in the volleyball group correctly identified as belonging to the volleyball group. Specificity was defined as the true negative rate, or the proportion of people in the control group who were correctly identified as not belonging to the volleyball group. Using sensitivity and specificity values, the maximum positive likelihood ratio (LR+) and its associated negative likelihood ratio (LR−) were calculated to determine the potential for using changes in the four different measures to differentiate speakers as members of the volleyball or control group. An LR+, calculated as $\text{Sensitivity}/1-\text{Specificity}$, is the confidence that a change in a measure across a given threshold indicates a member of the volleyball group. An LR−, calculated as $1-\text{Sensitivity}/\text{Specificity}$, is the confidence that a change in a measure across a given threshold indicates a member of the control group. Finally, calculation of ROC curves resulted in area under the curve (AUC) values, providing additional information about how effectively each measure could discriminate between individuals in the volleyball group and individuals in the control group, with an AUC of 0.5 reflecting chance.[38]

RESULTS

Onset cycle 1 RFF values from the volleyball group were significantly lower ($T = 2.3$, $P = 0.04$) in the postrecordings (M, 2.5 ST; SD, 1.6 ST) compared with prerecordings (M, 3.2 ST; SD, 1.6 ST; Figure 3). However, there was no significant difference ($T = -1.4$; $P = 0.20$) in RFF offset cycle 10 values between post (M, −0.6 ST; SD, 1.0 ST) and pre (M, −0.9 ST; SD, 1.2 ST). The control group participants displayed very minimal changes as function of time point in RFF for both onset cycle 1 (pre: M, 2.9 ST; SD, 0.7 ST; post: M, 2.9 ST; SD, 0.4 ST) and offset cycle 10 (pre: M, −1.0 ST; SD, 0.8 ST; post: M, −0.7 ST; SD, 1.1 ST; Figure 3).

Perceptual ratings of overall severity and strain were highly correlated ($r = 0.84$). Paired sample t tests were conducted to examine differences as function of time point in the perceptual ratings of overall severity and strain in the volleyball group. There was no significant difference between pre (M, 40.3; SD, 9.3) and post (M, 44.0; SD, 9.5) for overall severity ($T = -1.7$; $P = 0.12$) or between pre (M, 40.8; SD, 10.9) and post (M, 42.3; SD, 9.4) for strain ($T = -1.3$; $P = 0.29$). Although no statistical comparisons were made, both overall severity and strain in the control group displayed very minimal changes as a function of time point (Figure 4).

ROC curves were generated for changes (post-pre) in onset cycle 1 and offset cycle 10 RFF (Figure 5, *upper*). For changes in onset cycle 1 RFF, the AUC was 0.71, which is greater than chance (0.50). The maximum LR+ was 3.5 (occurring at −0.68 ST, with 0.83% specificity and 0.58% sensitivity), and the associated LR− was 0.50. Therefore an RFF change (decrease) in onset cycle 1 of −0.68 ST from pre to post was the optimal threshold to differentiate members of the volleyball and control groups. The ROC curve generated for offset cycle 10 RFF changes indicated weaker discrimination (Figure 5, *upper*). The AUC was approximately chance (0.56), with an LR+ of 2.5 (occurring at −0.11 ST, with 0.83% specificity and 0.42% sensitivity) and an associated LR− of 0.70. ROC curves were also generated for changes (post-pre) in VSR ratings for the vocal attributes of overall severity and strain (Figure 5, *lower*). The AUC for both vocal attributes was just over chance, with an AUC of 0.61 for strain and 0.65 for overall severity. The maximum LR+ for strain was 3.0 (occurring at 2.60, with 0.83% specificity and 0.50% sensitivity), and the associated LR− was 0.06. The maximum LR+ for overall severity was 3.5 (occurring at 3.55, with 0.83% specificity and 0.58% sensitivity), and the associated LR− was 0.50.

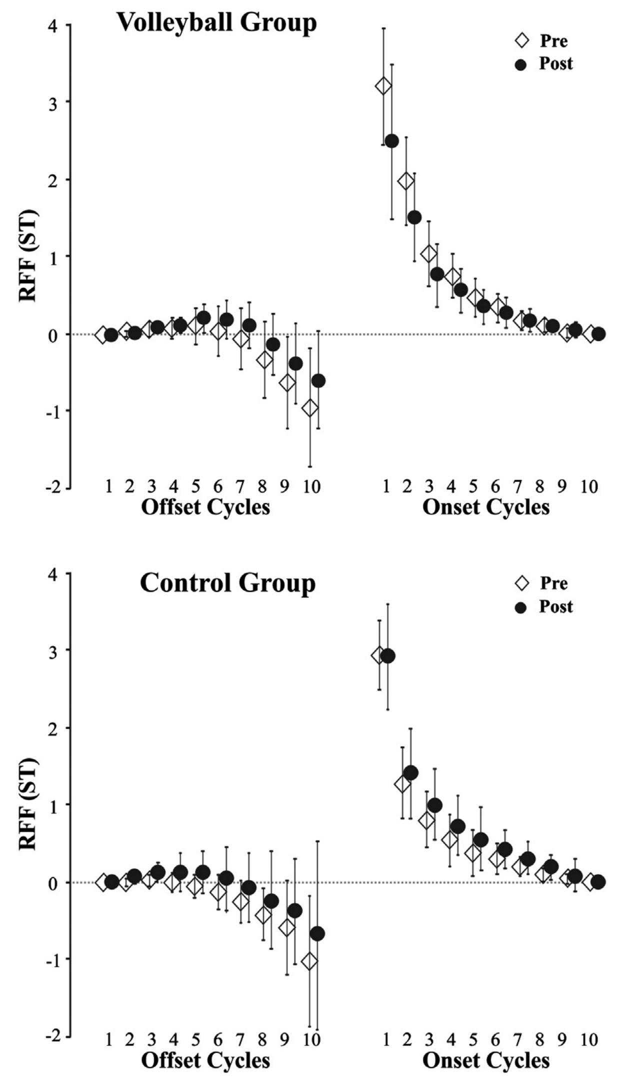

FIGURE 3. Relative fundamental frequency (RFF) averages for each of the 10 offset cycles and 10 onset cycles for both the volleyball group (*upper*) and the control group (*lower*). Averages from the prerecordings are the open diamonds and the postrecordings are the black circles; error bars indicate 95% confidence intervals.

DISCUSSION

Overall, this study demonstrated that onset cycle 1 RFF values decreased after an acute period of high voice use in healthy individuals. These acoustic changes were noted in the absence of clear perceptual changes. Results from this study suggest that onset cycle 1 RFF may be used as a marker for high voice

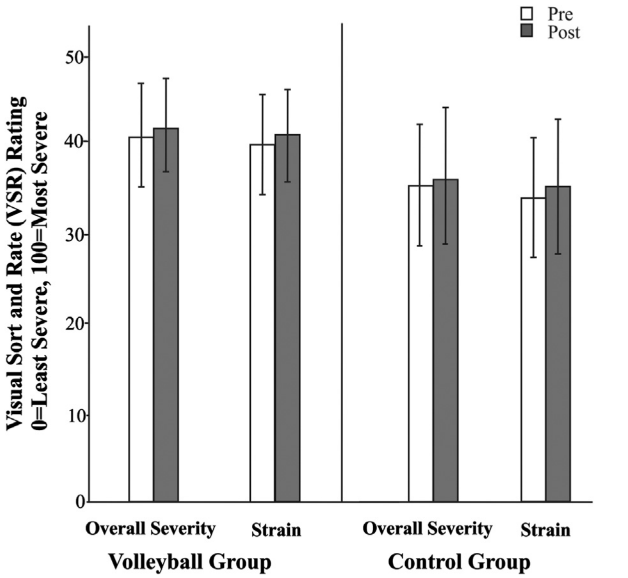

FIGURE 4. Average visual sort and rate (VSR) ratings for both overall severity and strain from the prerecordings and postrecordings of the volleyball and control groups. Error bars indicate 95% confidence intervals.

use over an acute period of time and could potentially indicate individuals are misusing and/or abusing their voices even before perceptual changes are noted.

Comparison with previous literature for onset cycle 1

Onset cycle 1 RFF values from prerecordings and postrecordings from the control group and prerecordings of the volleyball group were comparable with previously reported RFF values for healthy participants (2.8 ST–3.8 ST; Figure 6).[23,24] However, the onset cycle 1 RFF values from the postrecordings of the volleyball group (M, 2.5 ST) were lower and more similar to those previously found in individuals with vocal hyperfunction (1.9 ST–2.5 ST).[23,24,39] The changes in onset cycle 1 RFF values in the volleyball group's postrecordings suggest that there may have been increased vocal strain in the volleyball group that did not occur in the control group. These changes may be attributed to the high voice use that took place during the volleyball season. Potentially, individuals who engage in high voice use activities have an increased opportunity to, and therefore increased risk of, abuse and/or misuse of their voices, especially if they are not using their voices effectively and efficiently. This study provides evidence that individuals who are high voice users may have voicing patterns similar to individuals with vocal hyperfunction: onset cycle 1 RFF values from the volleyball group were lower in the postrecordings, with values that were similar to previously reported data from individuals with vocal hyperfunction.

Onset cycle 1 as a marker for high-voice use in a healthy population

Although this is one of the first articles to discuss RFF changes over an acute period of high voice use, there is a precedent for examining RFF over a short time period. Over a relatively brief

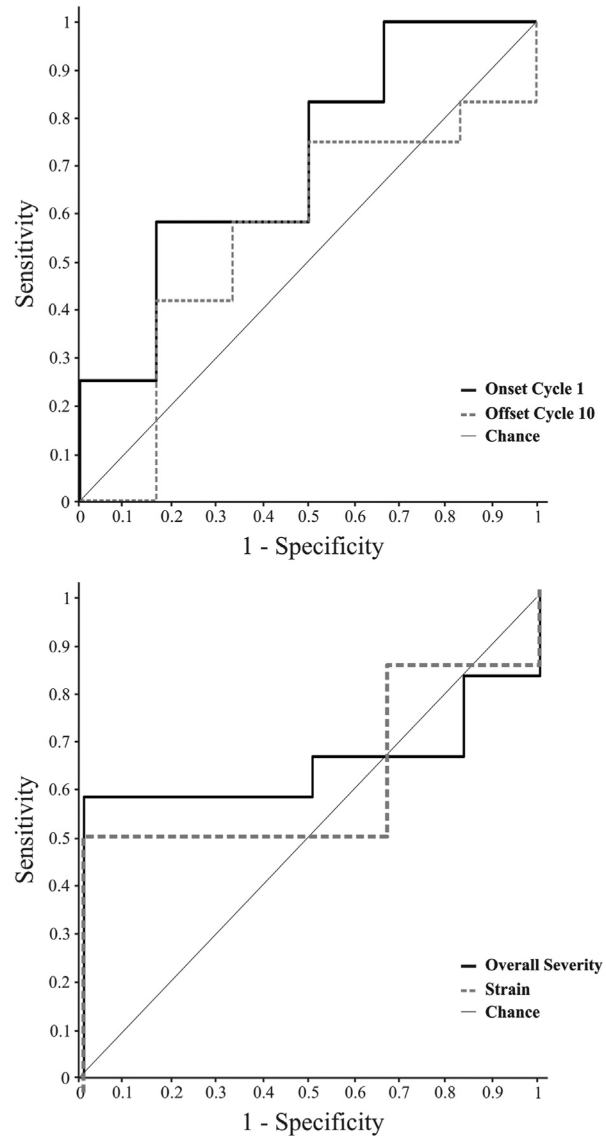

FIGURE 5. *Upper*: Receiver operating characteristic (ROC) curves for the change (post-pre) in onset cycle 1 RFF values (black solid line) and offset cycle 10 RFF values (gray dotted line) of the volleyball group. *Lower*: ROC curves for the changes (post-pre) in VSR ratings for overall severity (black solid line) and strain (gray dotted line) of all speakers. Chance is indicated on each panel by the diagonal gray line.

period of voice therapy (averaging 10 sessions), RFF values from individuals with vocal hyperfunction have been shown to increase in both offset cycle 10 and onset cycle 1.[39] Although onset cycle 1 values increased, they were still lower than those of healthy participants, an effect suggested to be due to the detection of a lingering hyperfunctional pattern.[39] The present study offers further evidence that onset cycle 1 RFF can change over an acute period of time.

The ROC curve for changes (post-pre) in onset cycle 1 RFF had an AUC of 0.71, indicating that using RFF onset cycle 1 values as a measure of change could differentiate between individuals in the volleyball and control groups at a better than chance rate (an AUC of 0.50). The measure of change in RFF onset cycle

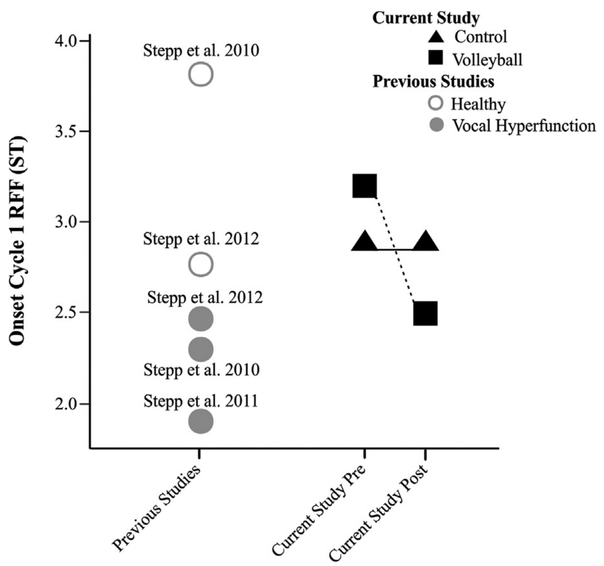

FIGURE 6. Average onset cycle 1 RFF (ST) from previous studies compared with the present study.

1 had an LR+ score of 3.5, indicating moderate confidence that individuals with a decrease in RFF values of at least 0.68 ST between the prerecordings and the postrecordings were a member of the volleyball group. However, the LR− score for the change in RFF onset cycle 1 was 0.5, which indicates insufficient evidence to be confident that an individual with a decreased RFF onset cycle was not a member of the control group.[40] Although the confidence levels were not strong, the ability to differentiate healthy individuals (control group) from healthy individuals with high voice use (volleyball group) provides potentially valuable information. Specifically, even in a group of healthy individuals, onset cycle 1 RFF may be a marker for high voice use over short time periods and could potentially indicate that these individuals are misusing and/or abusing their voices and therefore are at risk for developing voice disorders.

Overall severity and strain judgments over an acute period of voice use

Overall severity and strain perceptual judgments were highly correlated ($r = 0.84$), which indicates that listeners' perceptions of overall severity were heavily related to the strain perceived in the samples. There were no significant differences found between prerecordings and postrecordings for either vocal attribute. However, although no statistical tests were conducted, qualitative differences were noted between the VSR ratings of the control group and the volleyball group, with higher (more severe) average VSR ratings for both vocal attributes for the volleyball group relative to the control group. ROC curves for the perception of changes for both overall severity and strain resulted in AUC curves around 0.6, close to chance. This result indicates that the changes in the perception of overall severity and strain were not effective measures of differentiating between individuals in the volleyball and control groups. Interestingly, significant changes in onset cycle 1 RFF values were noted despite this lack of detection for qualitative differences in vocal quality.

Although perceptual measures are the gold standard for judgments of vocal quality,[41] they may not be sensitive to the changes that occur during acute periods of high voice use and may potentially fail to capture vocal abuse and/or misuse behaviors during these periods. Early detection of these vocally abusive behaviors is extremely beneficial for individuals who demonstrate vocal patterns of abuse and/or misuse, as the earlier behavioral patterns are detected, the easier they are to change during therapeutic interventions.[19] Additionally, if a high voice user does develop a pathology from her vocal misuse and/or abuse, the earlier it is detected, the more pliable and responsive the pathology is to therapy.[19] Therefore, the ability of onset cycle 1 RFF to detect small changes over an acute period of high voice use could have a substantial positive effect on therapeutic outcomes.

Limitations and future work

During the course of this study, the amount of voice use of each participant over the 10-week period between the prerecordings and postrecordings was not controlled. Although the volleyball participants reported increased vocal use during games and practices, voice use was not measured during these or other activities. Individuals in the volleyball group may have coincidentally misused and/or abused their voices in typical daily use in addition to the voice use related to volleyball, although no such behavior was reported. Future work will examine changes in onset cycle 1 RFF longitudinally in a group of self-reported high voice users and a group of self-reported low voice users using ambulatory monitoring devices.[42,43] RFF has been shown to be calculated accurately from a neck-placed accelerometer,[44] and the use of ambulatory monitoring could provide a noninvasive method of documenting vocal loads from both groups of individuals, potentially providing further insight into the relationship between voice use and changes in onset cycle 1 RFF.

This study also unexpectedly demonstrated that offset cycle 10 increased from the prerecordings to the postrecordings. However, this pattern was noted in both groups, and results from the ROC analysis revealed an AUC of 0.56, indicating that the changes (post-pre) of offset cycle 10 were not a viable measure to differentiate between individuals in the volleyball and control groups. Given that changes in offset cycle 10 were noted in both the volleyball and control groups, it is unlikely to be related to the differences in voice use between the two but rather may be because of potential environmental or social changes that could have affected all participants. For instance, the prerecordings were conducted in winter, when there is generally higher heat and less humidity indoors due to heating systems, and the postrecordings were conducted in spring. It is possible that the changes in offset cycle 10 RFF are mirroring the differences in hydration level on the basis of the season of the recording, with overall lower RFF values noted in the dryer climates associated with indoors during the winter season. These external drying agents in winter can be dehydrating, which can have negative phonatory effects.[31,45,46] Specifically, dehydration and exposure to low relative humidity have been shown to raise phonation threshold pressure, which is a measure of the subglottal pressure that is required for vocal fold vibration during phonation.[31,45–48] Additionally, hydration has been shown to have an effect on perceived phonatory effort, with an increase in hydration corresponding to a decrease in perceived phonatory effort.[31,48] Future research to examine potential changes in offset cycle 10 RFF at varying hydration levels in both individuals with voice disorders, and healthy participants will facilitate better characterization of any potential relationship between hydration and RFF.

CONCLUSIONS

This study was the first to characterize RFF changes over an acute period of high voice use. Results do not indicate that offset cycle 10 is an effective maker for vocal misuse and/or abuse over a relatively short period of time. However, small changes in onset cycle 1 were detected in the absence of perceived qualitative changes of overall severity or strain by listeners. Group data and ROC analysis at the individual level suggest the possibility of onset cycle 1 acting as an effective marker for vocal misuse and/or abuse over an acute high voice use time period. As individuals with vocal hyperfunction often display vocal abuse and/or misuse.[3] RFF is a promising measure to identify individuals who may be at increased risk for developing a voice disorder.

Acknowledgments

The authors thank Yu-An Stephanie Lien and Timothy Streeter. This work was supported by grants DC012651 and DC04663 from the National Institute on Deafness and Other Communication Disorders (NIDCD).

REFERENCES

1. Ramig LO, Verdolini K. Treatment efficacy: voice disorders. *J Speech Lang Hear Res*. 1998;41:S101–S116.
2. Roy N, Merrill RM, Gray SD, Smith EM. Voice disorders in the general population: prevalence, risk factors, and occupational impact. *Laryngoscope*. 2005;115:1988–1995.
3. Hillman RE, Holmberg EB, Perkell JS, Walsh M, Vaughan C. Objective assessment of vocal hyperfunction: an experimental framework and initial results. *J Speech Lang Hear Res*. 1989;32:373–392.
4. Roy N, Merrill RM, Thibeault S, Gray SD, Smith EM. Voice disorders in teachers and the general population: effects on work performance, attendance, and future career choices. *J Speech Lang Hear Res*. 2004;47:542–551.
5. Verdolini K, Ramig LO. Review: occupational risks for voice problems. *Logoped Phoniatr Vocol*. 2001;26:37–46.
6. Williams NR. Occupational groups at risk of voice disorders: a review of the literature. *Occup Med*. 2003;53:456–460.
7. Titze IR, Lemke J, Montequin D. Populations in the US workforce who rely on voice as a primary tool of trade: a preliminary report. *J Voice*. 1997;11:254–259.
8. Kristiansen J, Lund SP, Persson R, Shibuya H, Nielsen PM, Scholz M. A study of classroom acoustics and school teachers' noise exposure, voice load and speaking time during teaching, and the effects on vocal and mental fatigue development. *Int Arch Occup Environ Health*. 2014;87:1–10.
9. Martins RH, Pereira ER, Hidalgo CB, Tavares EL. Voice disorders in teachers. A review. *J Voice*. 2014;28:716–724.
10. Simberg S, Sala E, Vehmas K, Laine A. Changes in the prevalence of vocal symptoms among teachers during a twelve-year period. *J Voice*. 2005;19:95–102.
11. Thibeault SL, Merrill RM, Roy N, Gray SD, Smith EM. Occupational risk factors associated with voice disorders among teachers. *Ann Epidemiol*. 2004;14:786–792.
12. Smith E, Kirchner HL, Taylor M, Hoffman H, Lemke JH. Voice problems among teachers: differences by gender and teaching characteristics. *J Voice*. 1998;12:328–334.
13. Ryan S, Mendel LL. Acoustics in physical education settings: the learning roadblock. *Phys Educ Sport Pedag*. 2010;15:71–83.
14. Grillo EU, Fugowski J. Voice characteristics of female physical education student teachers. *J Voice*. 2011;25:e149–e157.
15. Newman C, Kersner M. Voice problems of aerobics instructors: implications for preventative training. *Logoped Phoniatr Vocol*. 1998;23:177–180.
16. Rumbach AF. Vocal problems of group fitness instructors: prevalence of self-reported sensory and auditory-perceptual voice symptoms and the need for preventative education and training. *J Voice*. 2013;27:524.e511–524.e521.
17. Heidel SE, Torgerson JK. Vocal problems among aerobic instructors and aerobic participants. *J Commun Dis*. 1993;26:179–191.
18. Long J, Williford HN, Olson MS, Wolfe V. Voice problems and risk factors among aerobics instructors. *J Voice*. 1998;12:197–207.
19. Colton RH, Casper JK, Leonard R. *Understanding Voice Problems: A Physiological Perspective for Diagnosis and Treatment*. 4th ed. Baltimore, MD: Lippincott Williams & Wilkins; 2011.
20. Reich A, McHenry M, Keaton A. A survey of dysphonic episodes in high-school cheerleaders. *Lang Speech Hear Serv Sch*. 1986;17:63–71.
21. Andrews M, Shank KH. Some observations concerning the cheering behavior of school-girl cheerleaders. *Lang Speech Hear Serv Sch*. 1983;14:150–156.
22. Eadie TL, Stepp CE. Acoustic correlate of vocal effort in spasmodic dysphonia. *Ann Otol Rhinol Laryngol*. 2013;122:169–176.
23. Stepp CE, Sawin DE, Eadie TL. The relationship between perception of vocal effort and relative fundamental frequency during voicing offset and onset. *J Speech Lang Hear Res*. 2012;55:1887–1896.
24. Stepp CE, Hillman RE, Heaton JT. The impact of vocal hyperfunction on relative fundamental frequency during voicing offset and onset. *J Speech Lang Hear Res*. 2010;53:1220–1226.
25. Kreiman J, Gerratt BR, Precoda K, Berke GS. Individual differences in voice quality perception. *J Speech Lang Hear Res*. 1992;35:512–520.
26. Audactiy, the Free, Cross-Platform Sound Editor (Version 2.0.3) [computer program] 2013.
27. *Matlab R2013a [Computer Program]. Version 8.1.0.604*. Natick, MA: The Mathworks, Inc.; 2013.
28. Granqvist S. The visual sort and rate method for perceptual evaluation in listening tests. *Logoped Phoniatr Vocol*. 2003;28:109–116.
29. Kempster GB, Gerratt BR, Abbott KV, Barkmeier-Kraemer J, Hillman RE. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. *Am J Speech Lang Pathol*. 2009;18:124–132.
30. Eadie TL, Doyle PC. Direct magnitude estimation and interval scaling of pleasantness and severity in dysphonic and normal speakers. *J Acoust Soc Am*. 2002;112:3014–3021.
31. Verdolini K, Titze IR, Fennell A. Dependence of phonatory effort on hydration level. *J Speech Lang Hear Res*. 1994;37:1001–1007.
32. *Voice Disorders: Simulations and Games*. Board of Regents of the University of Wisconsin System; 2008. Available at: http://csd.wisc.edu/slpgames/. Accessed March 1, 2013.
33. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86:420.
34. Lien Y-A, Gattuccio CI, Stepp CE. Effects of phonetic context on relative fundamental frequency. *J Speech Lang Hear Res*. 2014;57:1259–1267.
35. Praat: Doing Phonetic by Computer(Version 5.3.82) [Computer Program]. 2013. Available at: http://www.praat.org/2014. Accessed December 1, 2014.
36. Baken RJ. *Clinical Measurement of Speech and Voice*. Austin, Texas: Pro-Ed; 1987.
37. *Minitab Statistical Software [Computer Program]. Version 16.2.3*. State College: Pennsylvania State University; 2012.
38. Portney L, Watkins M. *Foundations of Clinical Research: Applications to Practice*. 3rd ed. Upper Saddle River, New Jersey: Peason Education; 2009.

39. 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: 1260–1266.
40. Dollaghan C. *The Handbook for Evidence-Based Practice in Communication Disorders*. Baltimore, Maryland: Paul H. Brookes Publishing Co., Inc.; 2007.
41. Kreiman J, Gerratt BR, Kempster GB, Erman A, Berke GS. Perceptual evaluation of voice quality: review, tutorial, and a framework for future research. *J Speech Hear Res*. 1993;36:21–40.
42. Hillman RE, Heaton JT, Masaki A, Zeitels SM, Cheyne HA. Ambulatory monitoring of disordered voices. *Ann Otol Rhinol Laryngol*. 2006;115: 795–801.
43. Bottalico P, Pavese L, Astolfi A, Hunter EJ. Voice accumulation and voice disorders in primary school teachers. *J Acoust Soc Am*. 2014;136:2294.
44. Lien Y-A, 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;135:2977–2985.
45. Levendoski EE, Sundarrajan A, Sivasankar MP. Reducing the negative vocal effects of superficial laryngeal dehydration with humidification. *Ann Otol Rhinol Laryngol*. 2014;123:475–781.
46. Sivasankar M, Erickson E, Schneider S, Hawes A. Phonatory effects of airway dehydration: preliminary evidence for impaired compensation to oral breathing in individuals with a history of vocal fatigue. *J Speech Lang Hear Res*. 2008;51:1494–1506.
47. Tanner K, Roy N, Merrill RM, Elstad M. The effects of three nebulized osmotic agents in the dry larynx. *J Speech Lang Hear Res*. 2007;50:635–646.
48. Verdolini-Marston K, Sandage M, Titze IR. Effect of hydration treatments on laryngeal nodules and polyps and related voice measures. *J Voice*. 1994; 8:30–47.