JSLHR

Article

The Relationship Between Perception of Vocal Effort and Relative Fundamental Frequency During Voicing Offset and Onset

Cara E. Stepp,[a] Devon E. Sawin,[b] and Tanya L. Eadie[b]

Purpose: In this study, the authors aimed to determine the relationship between relative fundamental frequency (RFF) and listener perception of vocal effort in individuals with varying degrees of vocal hyperfunction.
Method: Thirty women diagnosed with voice disorders commonly associated with vocal hyperfunction and 10 healthy women provided speech samples that were used to obtain parameters of RFF. Twelve listeners judged the speech samples for overall severity and vocal effort (VE) using rating scales.
Results: Significant but relatively weak negative correlations were found between perceptual measures and offset RFF parameters. Although offset RFF was increased in healthy participants relative to speakers with voice disorders, no differences were seen in RFF as a function of severity of VE in individuals with voice disorders.
Conclusions: Although a statistically significant correlation between offset RFF and VE was found, examination of data as a function of both VE and health status indicated that RFF more accurately classifies the presence of a voice disorder than does severity of voice quality or VE. There is a need for further research to investigate the clinical utility of RFF measures for assessment of rehabilitation progress.

Key Words: vocal hyperfunction, fundamental frequency, vocal effort

Voice disorders have a devastating impact on individuals' lives through both economic and social effects (Ramig & Verdolini, 1998), and vocal hyperfunction accounts for 10%–40% of individuals who are referred to multidisciplinary voice clinics (Roy, 2003). *Vocal hyperfunction* has been defined as including "conditions of abuse and/or misuse of the vocal mechanism due to excessive and/or 'imbalanced' muscular forces" (Hillman, Holmberg, Perkell, Walsh, & Vaughan, 1989, p. 373) and excessive laryngeal and paralaryngeal tension (e.g., Aronson, 1980; Dworkin, Meleca, & Abkarian, 2000; Koufman & Blalock, 1991; Morrison, Rammage, Belisle, Pullan, & Nichol, 1983; Roy, 2008), and it is a common cause of and accompaniment to many types of voice disorders.

Individuals diagnosed with vocal hyperfunction have been shown to respond to behavioral intervention (Holmberg, Hillman, Hammarberg, Sodersten, & Doyle, 2001). However, the valid determination of outcomes depends on the therapist's ability to accurately measure the construct. Current assessment of vocal hyperfunction in clinical practice still relies primarily on subjective interpretation of patient history and physical examination of the neck and surrounding tissues (e.g., neck tension palpation rating systems) as well as both visual (e.g., laryngovideostroboscopic) and auditory–perceptual measures. Of these measures, listener perception of vocal effort (VE; strain) is the most common auditory–perceptual quality attributed to vocal hyperfunction.

Although auditory–perceptual measures may still be considered "gold standard" measures for defining hyperfunction-related voice disorders, they may be subject to error and poor reliability (Kreiman, Gerratt, Kempster, Erman, & Berke, 1993). As a result, objective measures such as those derived from acoustics, aerodynamics, and electromyographic signals have been important areas of inquiry for assessing vocal hyperfunction. Acoustic measures are generally computed during middle segments of sustained vowels or, more recently, throughout running speech (Awan & Roy, 2009; Bhuta, Patrick, &

---

[a]Boston University
[b]University of Washington, Seattle
Correspondence to Cara E. Stepp: cstepp@bu.edu

Editor: Anne Smith
Associate Editor: Shaheen Awan
Received October 30, 2011
Revision received March 24, 2012
Accepted April 30, 2012
DOI: 10.1044/1092-4388(2012/11-0294)

Garnett, 2004; Heman-Ackah, Michael, & Goding, 2002). A number of studies have shown relatively strong correlations between acoustic measures and overall severity (OS) of dysphonia, breathiness, and roughness (Bhuta et al., 2004; Heman-Ackah et al., 2002; Shrivastav & Sapienza, 2003). As yet, however, no single measure has been shown to be highly correlated with the specific perceptual attributes of VE or strain (e.g., Bhuta et al., 2004).

Recent work suggests potential for the use of relative fundamental frequency (RFF) changes surrounding voiceless consonant production as an assay of vocal hyperfunction (Stepp, Hillman, & Heaton, 2010; Stepp, Merchant, Heaton, & Hillman, 2011). *RFF* has been defined as the fundamental frequency of the cycles immediately before and after production of a voiceless consonant, normalized by the "steady-state" fundamental frequencies of the voicing preceding and subsequent to the consonant (Stepp et al., 2010). Normalization in semitones (STs) relative to the steady-state instantaneous fundamental frequency allows the changes in fundamental frequency to be compared across individuals with differing typical fundamental frequencies. Several studies have shown that, for healthy speakers, the vocal cycles immediately after voiceless consonant production (vowel onset) show increased RFF (Goberman & Blomgren, 2008; Ohde, 1984; Robb & Smith, 2002; Watson, 1998). The RFF immediately prior to voiceless consonant production (vowel offset) also shows characteristic patterns. Young healthy speakers have relatively stable offset RFF, whereas older healthy speakers show a slight decrease in RFF leading into the voiceless consonant (Goberman & Blomgren, 2008; Robb, Chen, Gilbert, & Lerman, 2001; Watson, 1998).

It has been hypothesized that the characteristic physiological behaviors that occur during the production of vowels before and after a voiceless consonant underlie these distinctive changes in the RFF. These physiological behaviors include changes in tension (Löfqvist, Baer, McGarr, & Story, 1989; Stevens, 1977), transitions to and from vocal fold abduction/adduction (Fukui & Hirose, 1983), and changes in airflow (Löfqvist, Koenig, & McGowan, 1995; Löfqvist & McGowan, 1992).

In particular, tension is thought to be increased before, during, and immediately following voiceless consonant production (Löfqvist et al., 1989; Stevens, 1977), a phenomenon that may be used to inhibit voicing (e.g., [+stiff]). Carryover of this short-term increase in tension into surrounding vowels could contribute to an increase in both offset and onset RFF (Halle & Stevens, 1971; Stevens, 1977). We previously postulated that high-baseline laryngeal muscle tension might reduce individuals' ability to create these short-term variations in tension during devoicing (Stepp et al., 2010, 2011), leading to the lowered RFF seen in populations with high laryngeal tension (i.e., individuals with Parkinson's disease and vocal hyperfunction; Goberman & Blomgren, 2008; Stepp et al., 2010). Although acoustic measures using vowel onsets have been previously shown to be more predictive of listener ratings of roughness (de Krom, 1995), the use of vowel onsets and offsets for acoustic correlates of VE or strain has not yet been examined.

RFF has been shown to (a) discriminate between individuals with and without voice disorders associated with vocal hyperfunction (Stepp et al., 2010) and (b) normalize after successful voice therapy toward values seen in speakers with no vocal impairment (Stepp et al., 2011). Although these results seem promising, it is yet unknown how RFF relates to perceived VE or OS among individuals with varying degrees of hyperfunction. Characterizing the nature of this relationship is critical for determining the appropriate role for use of RFF measures. Consequently, the objective of this study was to determine the relationship between RFF and listener perception of VE and OS of voice in individuals with varying degrees of vocal hyperfunction. Specifically, on the basis of prior work (Stepp et al., 2010; Stepp & Eadie, 2011), we hypothesized that RFF values would be negatively correlated with the perception of VE, with voice samples rated with high VE showing lower values of RFF.

Method

Participants

Speakers. Speech samples were selected from an archived database of samples recorded under identical conditions for clinical research. Selected samples included 30 adult women with a variety of voice disorders commonly associated with vocal hyperfunction (age range: 18–63 years; *M* = 31, *SD* = 15) as well as 10 healthy adult women with normal voices (age range: 18–53 years; *M* = 25, *SD* = 11). The samples were selected by the authors—all of whom are experienced voice researchers—to primarily represent a broad spectrum of perceived VE. All participants were native English speakers, and none reported speech, language, or hearing problems beyond voice complaints for individuals in the voice disorders group.

All speakers were evaluated either at the University of Washington Speech and Hearing Clinic or in the Laryngology Clinic at the University of Washington Medical Center. All speakers were assessed by an experienced laryngologist and one or more certified speech-language pathologists. Diagnoses (normal voice or voice disorder) for both speaker groups were made on the basis of comprehensive voice evaluation procedures that included videostroboscopy, a careful case history, and auditory–perceptual measures. All speakers with voice disorders reported voice complaints as the primary reason for their evaluation, whereas no control speakers reported any vocal complaints. The diagnoses of participants with voice disorders are shown in Table 1. The healthy speakers were evaluated as part of a volunteer voice screening program at the University of Washington Speech and Hearing Clinic.

Table 1. Participant (P) diagnoses.

Participant	Age (yrs.)	Diagnosis
P1	26	Vocal fold nodules
P2	51	Interarytenoid granulation (granuloma)
P3	47	Cyst, chronic cough
P4	63	Reinke's edema
P5	31	Vocal fold nodules
P6	44	Vocal cysts
P7	56	Vocal fold nodules
P8	26	Reinke's edema
P9	60	Right vocal fold paresis
P10	26	Vocal fold nodules
P11	27	Pseudocysts
P12	52	Reinke's edema
P13	52	Vocal fold edema
P14	18	Left vocal fold paresis; left vocal fold vascular lesion
P15	18	Vocal fold nodules
P16	20	Left vocal fold polyp and right vocal fold reactive lesion
P17	19	Left vocal fold pseudocyst and right vocal fold reactive lesion
P18	20	Vocal fold nodules
P19	23	Subepithelial vocal fold lesion
P20	18	Right vocal fold intracordal lesion and left vocal fold reactive lesion
P21	18	Right vocal fold cyst and left reactive lesion
P22	20	Right vocal fold paresis and left vocal fold pseudocyst
P23	22	Muscle tension dysphonia
P24	51	Muscle tension dysphonia
P25	21	Vocal fold edema
P26	18	Reinke's edema
P27	19	Left vocal fold pseudocyst
P28	18	Vocal fold edema; posterior arytenoid erythema
P29	19	Vocal fold edema; vascular lesion
P30	19	Left vocal fold microlesion and right vocal fold companion lesion
C1	35	Healthy normal voice
C2	53	Healthy normal voice
C3	18	Healthy normal voice
C4	28	Healthy normal voice
C5	20	Healthy normal voice
C6	19	Healthy normal voice
C7	19	Healthy normal voice
C8	19	Healthy normal voice
C9	19	Healthy normal voice
C10	19	Healthy normal voice

Note. C = control participant.

Listeners. Listeners were four men and eight women between the ages of 20 and 36 years ($M = 25$, $SD = 4$) with no history of speech, language, or hearing disorders. Listeners passed hearing screening tests at 20 dB SPL for the octave frequencies of 250 Hz through 8000 Hz. These participants were recruited from the student population and broader community at the University of Washington and had no prior experience with or exposure to voice disorders. Participants were native English speakers who were paid for their participation. Procedures were approved by the University of Washington Human Subjects Committee.

Speech Collection and Stimulus Preparation

All speakers provided demographic information and then were asked to provide speech samples of the first paragraph of "The Rainbow Passage" (Fairbanks, 1960) using their typical pitch and loudness used in everyday conversation. The speech samples were recorded with a headset microphone (AKG C420) routed to a digital audiotape recorder (Tascam DAP1) at a sampling rate of 44.1 kHz in a quiet environment with low amounts of ambient noise.

Once we recorded the speech samples, we transferred them from a digital audio recorder (Sony PCM-R500) to a desktop computer with a specialized sound card and converted them into digital .wav files using acoustic software at a sampling rate of 44.1 kHz and 16 bits of resolution (Sony Soundforge Pro 10.0). For the listener protocol, we extracted the fifth and sixth sentences of "The Rainbow Passage," and we normalized the intensity of each speech sample for peak energy using sound-editing software (Sony Soundforge Pro 10.0) to ensure that each sample was presented to listeners at the same relative intensity level. The samples were then entered into a custom software program (using the Ruby on Rails web platform; http://rubyonrails.org/) that randomly generates speaker order, presents rating scales, and records responses. To assess intrarater reliability, we randomly repeated approximately 20% ($n = 8$) of the samples for each dimension (OS and perceived VE), for a total of 96 speech samples judged by each listener ($n = 40 + 8$ samples per dimension × 2 dimensions = 96 samples).

Acoustic Data Analysis

Three voiced–voiceless–voiced combinations were selected from the fifth and sixth sentences of the passage for analysis: "ever finds" (/ər/-/f/-/aɪ/) and two combinations from "looking for" (/ʊ/-/k/-/ɪŋ/ and /ɪŋ/-/f/-/ɔr/). We selected these three instances because they contain a voiceless obstruent surrounded on each side by vowels that are produced by most speakers with a relatively long duration. A single investigator (the first author) performed acoustic analysis by displaying the time waveforms of the samples in Praat acoustic analysis software (Boersma & Weenink, 2008) and measuring the 10 periods of vibration prior to (offset) and after (onset) the voiceless consonant using the pulse function (see Figure 1). We used 10 periods for analysis in order to be consistent with prior work and to permit comparison of results (Goberman & Blomgren, 2008; Stepp et al., 2010; Watson, 1998). The instantaneous fundamental frequency was calculated as the inverse of each period, *T* (see Figure 1). All frequencies were then converted to STs relative to the points in the voicing farthest from the voiceless consonant—the first cycle in the 10 cycles prior to voicing offset (Offset Cycle 1) and the final cycle in the 10 cycles following voicing onset (Onset Cycle 10), as in the following equation (Baken, 1987, p. 127):

$$\mathrm{ST} = 39.86 \times \log_{10}(f/f_{ref}). \tag{1}$$

We averaged the RFF for each speaker across all three voiceless consonant productions that were studied to provide a more stable estimate of RFF. Previous work did not find significant differences among these three voiceless consonant productions (Stepp et al., 2010). During some productions, *glottalization,* or a lack of periodicity prior or following the voiceless consonant production, made it impossible to reliably determine RFF. When this occurred, RFF values from that production were excluded, and only the remaining productions contributed to the average for that speaker. We used an average of 2.6 (*SD* = 0.6) productions for each participant's offset averages and an average of 2.4 (*SD* = 0.7) productions for each participant's onset averages.

The first author reevaluated approximately 20% of the samples approximately 9 months after the initial evaluation to assess intrarater reliability (Pearson's $r = .91$); the average difference for all RFF values in this sample was –0.04 ST, which was not a statistically significant difference, Student's paired t test, $t(159) = 0.72$, $p = .36$, two-tailed. A second trained researcher (see Acknowledgments) analyzed approximately 20% of the samples to assess interrater reliability (Pearson's $r = .93$); the average difference for all RFF values between the two raters in this sample was –0.04 ST, which was not a statistically significant difference, Student's paired t test, $t(159) = 0.75$, $p = .32$, two-tailed.

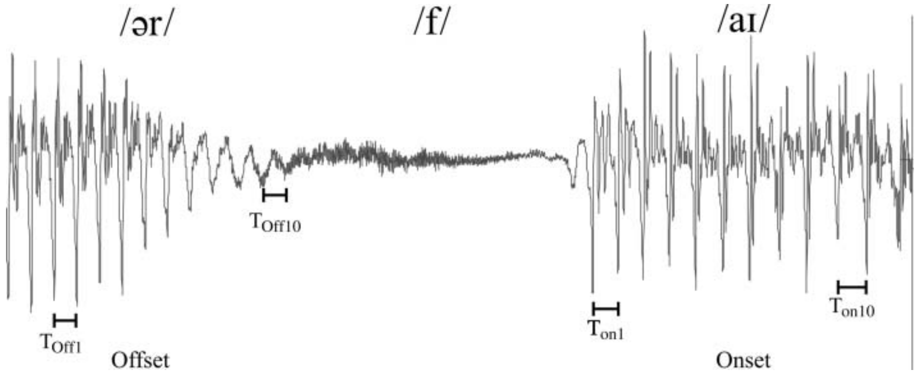
Listening Procedures

Before performing the rating tasks, listeners were given definitions of OS and perceived VE and were familiarized with a 100-mm visual analog scale (VAS). *OS* was defined as "a comprehensive measure of how 'good' or 'poor' the voice sample is judged to be by the listener" (Eadie & Doyle, 2002, p. 3017), and *VE* was defined as "the perceived effort during phonation" (Verdolini, Titze, & Fennell, 1994, p. 1001).

Listeners first listened to two speech samples—those of one male speaker and one female speaker who represented the approximate midpoint on the VAS for OS or VE—to familiarize themselves with each dimension. Listeners were then asked to make judgments of OS and VE for all speech samples using 100-mm VAS. Stimuli were presented to listeners at comfortable loudness levels through headphones (Samson RH600). For each dimension, listeners made judgments after a single presentation of each stimulus. The order of perceptual dimension (OS, VE) was counterbalanced, and speaker order was randomized across listeners. The entire listening session lasted approximately 30 min.

After data collection was complete, group means of listener ratings for OS and VE were calculated for each speaker. Listeners demonstrated adequate reliability;

Figure 1. Example of an acoustic waveform of voice offset and onset used for analysis. We used the periods, T, of the 1st through the 10th cycles of vibration to estimate the instantaneous fundamental frequency as a function of cycle.



the mean Pearson product–moment correlations between original and repeated samples were .90 ($SD = .10$) for OS and .82 ($SD = .14$) for VE. Interrater reliability was analyzed using group mean intraclass correlation coefficients (ICCs), with demonstrated reliability for both OS (ICC = .97) and VE (ICC = .96).

Statistical Analysis

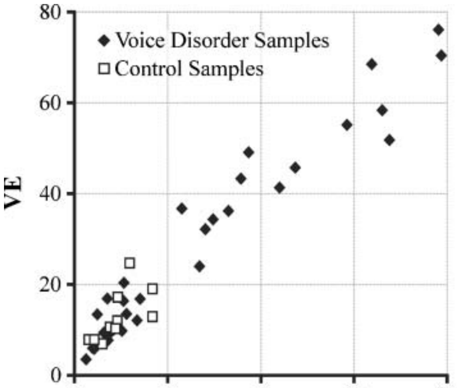
All statistical analyses were completed using Minitab Statistical Software (Minitab Inc.). To determine the relationships between perceptual and RFF parameters, we calculated Pearson product–moment correlations between perceptual mean scores for each speaker with two RFF parameters: (a) RFF at Offset Cycle 10 and (b) RFF at Onset Cycle 1. These two measures were selected as being most likely to differentiate among levels of vocal hyperfunction, given the patterns seen previously in this population (Stepp et al., 2010, 2011) and because they were the points farthest from those used for normalization (Offset Cycle 1 and Onset Cycle 10). To determine whether RFF differentiated levels of perceived hyperfunction without assuming a linear relationship between measures, we conducted a one-way analysis of variance (ANOVA) for each of the two RFF parameters on the basis of VE severity and vocal disorder status (healthy voice vs. voice disorder). Speakers were collapsed into four groups: one group of speakers with no vocal impairment (controls) and three groups of individuals with voice disorders further grouped using VE severity such that *normal* was associated with VE scores from 0 through 11 mm ($n = 8$), *mild* was associated with VE scores from 12 through 35 mm ($n = 10$), and *moderate-to-severe* was associated with VE scores from 36 through 100 mm ($n = 12$) on the 100-mm VAS based on the structure of the Consensus Auditory–Perceptual Evaluation of Voice (Kempster, Gerratt, Verdolini Abbott, Barkmeier-Kraemer, & Hillman, 2009). We hypothesized that controls and individuals with voice disorders and normal VE would have the highest RFF values, with individuals with voice disorders and mild VE showing decreased RFF and individuals with voice disorders and moderate-to-severe VE showing the lowest RFF values. We used Student's *t* tests to assess potential differences in VE ratings and the two RFF parameters between individuals with voice disorders and control participants.

Finally, we determined measures of sensitivity and specificity and associated parameters for the RFF value at Offset Cycle 10 and the RFF value at Onset Cycle 1 to ascertain how accurately the RFF measures could classify individuals with voice disorders from control participants. *Sensitivity* (the true-positive rate) is the proportion of actual positives that are correctly identified as positive (i.e., how good the measure is at identifying individuals with voice disorders). *Specificity* is the proportion of negatives that are correctly identified as negative (the true-negative rate; i.e., how good the measure is at identifying individuals with healthy voices; Dollaghan, 2007). The relationship between sensitivity and specificity was displayed graphically using a receiver operating characteristic (ROC) curve. The ROC curve was generated by calculating sensitivity and specificity at all potential cutoff points with a resolution of 0.01 ST for discriminating between participants with voice disorder ($n = 30$) and healthy participants ($n = 10$). Based on the ROC, the area under the curve (AUC) was determined via numerical integration (trapezoidal rule) for each detection scheme. Positive likelihood ratios were calculated as sensitivity/(1 – specificity); negative likelihood ratios were calculated as 1 – sensitivity/specificity. Both measures provide additional information regarding the diagnostic value of each parameter.

Results

The perceptual ratings of OS and VE for the 40 samples showed a mean OS of 23.5 ($SD = 22.8$, range: 2.5–79.2) and mean VE of 25.8 ($SD = 20.5$, range: 3.6–76.8). In the ratings for the 30 voice samples from individuals with voice disorders, the mean OS was 28.2 ($SD = 24.5$, range: 2.5–79.2), and the mean VE was 30.1 ($SD = 22.0$, range: 3.6–76.8), suggesting that most voices were in the mild to moderate range for both OS and VE. Average values for the voice samples from control participants showed a mean OS of 9.4 ($SD = 4.7$, range: 3.1–16.9) and a mean VE of 13.1 ($SD = 5.8$, range: 7.1–25.0). Figure 2 shows individuals' values of OS and VE as a function of vocal status (participant with voice disorder or healthy participant). Overall, ratings of OS and VE were highly correlated ($R^2 = .94$, $p < .001$), indicating that the

Figure 2. Vocal effort (VE) ratings for participants with voice disorder and control participants as a function of overall severity (OS) ratings.



perception of VE explains a large amount of the variability in the perception of OS of voices in this sample (see Figure 2). A Student's *t* test (two-sided, $df = 37$) revealed that VE ratings were statistically significantly higher in the 30 samples from individuals with voice disorder compared with the 10 control voices ($T = -3.8$, $p < .001$, $d = -0.88$), but that there was also considerable overlap between VE ratings of some individuals with voice disorders and those of the control voices (see Figure 2).

The RFF values for the 40 samples showed a mean Offset Cycle 10 value of –1.30 ST ($SD = 1.21$, range: –3.64 to 2.12) and mean Onset Cycle 1 value of 2.54 ST ($SD = 0.99$, range: –0.36 to 4.31). The 30 voice samples from individuals with voice disorders had a mean Offset Cycle 10 value of –1.75 ST ($SD = 0.91$, range: –3.64 to –0.40) and a mean Onset Cycle 1 value of 2.47 ST ($SD = 1.03$, range: –0.36 to 4.31). The 10 samples from control participants had a mean Offset Cycle 10 value of 0.05 ($SD = 1.00$, range: –1.78 to 2.12) and a mean Onset Cycle 1 value of 2.77 ($SD = 0.87$, range: 1.39 to 3.89). Figure 3 shows the mean RFF of participants as a function of vocal status (30 individuals with voice disorders or 10 healthy participants). A Student's *t* test (two-sided) found that Offset Cycle 10 RFF was statistically significantly lower in the 30 samples from individuals with voice disorders compared with the 10 control voices, $t(14) = 5.04$, $p < .001$, $d = 1.93$, but did not find a significant difference between Onset Cycle 1 RFF, $T(18) = 0.92$, $p = .37$, $d = 0.31$. We found statistically significant ($p < .05$) but weak negative correlations between perceptual measures (OS, VE) and Offset Cycle 10 RFF but not between either perceptual measure and Onset Cycle 1 RFF (see Table 2). The RFF parameters were able to explain roughly 7% to 19% (R^2) of the variance seen in the perceptual data.

Table 2. Pearson's correlation coefficients among relative fundamental frequency (RFF) and perceptual measures.

RFF	OS	VE
Offset 10	–0.43 ($p = .005$)	–0.44 ($p = .005$)
Onset 1	–0.27 ($p = .10$)	–0.27 ($p = .09$)

Note. OS = overall severity; VE = vocal effort.

Figure 3. Mean values of relative fundamental frequency (RFF) for participants with voice disorder ($n = 30$) and control participants ($n = 10$). Error bars indicate standard error. A Student's *t* test (two-sided) revealed that Offset Cycle 10 RFF was statistically significantly lower in the individuals with voice disorders compared with the 10 control voices. No significant difference in Onset Cycle 1 RFF was found between individuals with voice disorders and control speakers. ST = semitone.

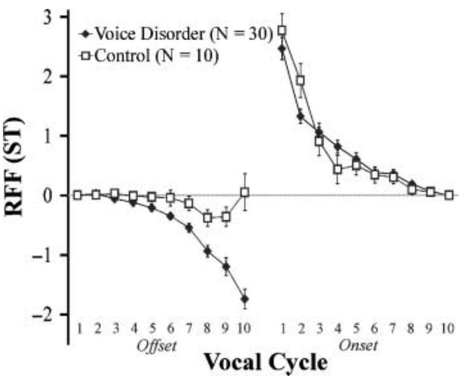
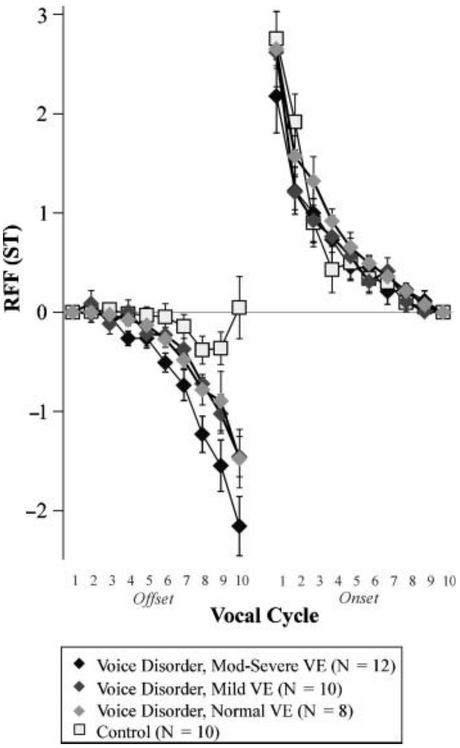


Figure 4 shows the mean RFF of speakers as a function of VE grouping (healthy participants, participants with voice disorder with normal VE, participants with voice disorder with mild VE, and participants with

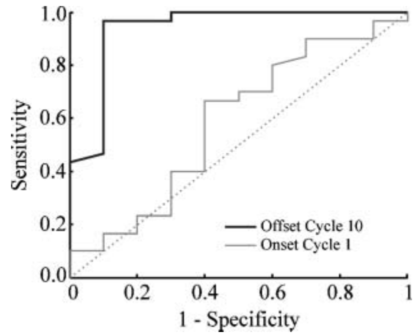
Figure 4. Mean values of RFF by group: control, voice disorder with normal VE, voice disorder with mild VE, and voice disorder with moderate VE. Voices were grouped using VE severity (*normal* = VE scores from 0 through 11, *mild* = VE scores from 12 through 35, *moderate-to-severe* = VE scores from 36 through 100). Error bars indicate standard error. A one-way analysis of variance showed a statistically significant effect of group on the Offset Cycle 10 RFF but not on Onset Cycle 1 RFF. Post hoc testing of Offset Cycle 10 RFF indicated that RFF was significantly increased in controls relative to all groups of speakers with voice disorder. No significant differences in Offset Cycle 10 RFF were seen among the groups of individuals with voice disorders.



voice disorder with moderate-to-severe VE). Although VE ratings were comparable for the 10 controls and the eight participants with voice disorder and VE ratings in the normal range, mean RFF values between the two groups show distinctly different patterns. The control group shows higher RFF, whereas all the groups of speakers with a vocal disorder, regardless of their VE ratings, showed lower RFF (see Figure 4). However, mean offset RFF values for the participants with voice disorder did show some stratification when plotted as a function of their VE grouping (see Figure 4), with a trend for decreased offset RFF in individuals with voice disorder and moderate-to-severe VE compared with other individuals with voice disorders. A one-way ANOVA revealed a statistically significant effect of VE grouping (control participants, participants with voice disorder with normal VE, participants with voice disorder with mild VE, and participants with voice disorder with moderate-to-severe VE) on Offset Cycle 10 RFF, $F(3) = 11.3$, $p < .001$, $\eta_p^2 = .48$, but not on Onset Cycle 1 RFF, $F(3) = 0.73$, $p = .543$, $\eta_p^2 = .06$. Post hoc testing of Offset Cycle 10 RFF using Tukey's simultaneous *t* tests indicated that RFF was significantly increased ($p < .01$) in healthy participants relative to all groups of speakers with voice disorder but that no significant differences in Offset Cycle 10 RFF were seen among the groups of individuals with voice disorders.

The ROC curve is shown in Figure 5 and reveals the sensitivity and specificity of offset and onset RFF in discriminating individuals with voice disorders from healthy speakers. Onset Cycle 1 is very close to the chance detection rate (dotted gray line), but Offset Cycle 10 shows excellent discrimination, reaching close to perfect detection (upper left corner). The AUC was 0.94 for Offset Cycle 10 and 0.58 for Onset Cycle 1. In addition to calculating sensitivity and specificity, we also calculated positive and negative likelihood ratios (LRs). A *positive LR* (LR+) reflects the level of confidence that a positive test score truly indicates the presence of a disorder, whereas a *negative LR* (LR–) suggests how well a normal score represents the condition of no disorder (Dollaghan, 2007). The higher the LR+, the greater the probability that an individual who obtained an abnormal score has the disorder, with values ≥ 10 representing a strong result (Dollaghan, 2007). The lower the LR–, the greater the probability that an individual who obtained a normal score does not have the disorder, with values < .10 indicating a strong result. Results from this study showed moderately strong results for Offset Cycle 10 RFF: The maximum LR+ was 9.7 (occurring at the –0.56 ST threshold [sensitivity = .97, specificity = .90]), and the associated LR– was 0.04. Results were much weaker for Onset Cycle 1 RFF: The maximum LR+ was 1.67 (occurring at the 1.76 ST threshold [sensitivity = .17, specificity = .90]), with an associated LR– of 0.92.

Figure 5. Receiver operating characteristic curves for Offset Cycle 10 (thick black line) and Onset Cycle 1 RFF (thin gray line). The dotted gray line on the diagonal from the lower left to the upper right corners shows when the true-positive rate is equal to the false-positive rate (chance).



Discussion

The goal of this study was to characterize the relationship between RFF and listener perception of VE and OS in individuals with varying degrees of vocal hyperfunction. Our hypothesis was that RFF values would be negatively correlated with the perception of VE. Overall, although we did find statistically significant negative correlations between offset RFF and both VE and OS, these relationships were weak. Instead, we unexpectedly found that individuals with voice disorders had lower offset RFF, even when their VE ratings were in the normal range. In addition, we found that onset RFF was not sensitive to voice disorders or to VE ratings.

RFF Discriminates Between Voices With Disorder and Voices in the Normal Range

Samples from individuals with voice disorders ($n = 30$) were found to have an average offset RFF value of –1.75 ST at Cycle 10 and an average onset RFF value of 2.47 ST at Cycle 1. These values are consistent with those found by Stepp et al. (2010), who studied female participants with hyperfunction-related voice disorders and reported average offset RFF values ranging from –1.017 to –1.61 ST at Cycle 10 and average onset RFF values ranging from 2.12 to 2.48 at Cycle 1. Samples from control voices ($n = 10$) in this study were found to have an average offset RFF value of 0.05 ST and an average onset RFF value of 2.77 ST. These, too, are consistent with the previous literature, in which individuals with healthy voices showed average offset RFF at Cycle 10 of –0.33 and average onset RFF at Cycle 1 of 3.82 (Stepp et al., 2010), albeit with onset values slightly lower than those previously reported.

When considered together with the findings of previous studies, our results also support the contention that offset RFF differs in individuals with those voice disorders that are associated with vocal hyperfunction relative to individuals with healthy voices. In addition, the results support the possibility of using RFF as a sensitive indicator for those voice disorders that are commonly associated with vocal hyperfunction. As shown in the ROC curve in Figure 5, detection using Offset Cycle 10 RFF appears qualitatively most similar to ideal detection. Offset Cycle 10 has a high AUC (0.94), suggesting very high diagnostic accuracy. In addition, it has a maximum LR+ that is moderately strong (9.7, with strong being ≥ 10) and an associated LR– that is strong (0.04, with strong being $< .10$). The Offset Cycle 10 RFF values lower than –0.56 ST are nearly 10 times more likely to be seen in individuals with voice disorders than in individuals without voice disorders. The sensitivity seen in Offset Cycle 10 is quite promising, even in light of the performance of multiparameter approaches (e.g., using mostly spectral-based measures of connected speech samples) using far more data points for analysis (e.g., Awan, Roy, Jette, Meltzner, & Hillman, 2010). However, to interpret its clinical significance and role as a possible outcome measure, the acoustic measures used in this study must be interpreted in light of their relationship with auditory–perceptual measures.

Relationship Between RFF Measures and OS and VE

We anticipated finding negative correlations between onset and offset RFF and the perception of OS of voice as well as VE, with high VE (and OS) corresponding to lowered RFF. Although we did see significant negative relationships between offset RFF and the perceptions of VE and OS, these relationships were quite weak (see Table 2). Furthermore, we did not find statistically significant correlations between onset RFF and VE or OS. These correlations are far weaker than those found for other acoustic measures that have been examined as potential correlates of OS of dysphonia (e.g., Awan & Roy, 2009). If there is a relationship between VE and offset RFF, it may not be linear.

Although RFF appears to be associated with voice disorder, it does not appear to be sensitive to changes in the magnitude of vocal hyperfunction perceived by other listeners, such as the differences between normal and mildly rated VE (see Figure 4). Furthermore, even though VE ratings for control speakers were similar to those of the individuals with voice disorders with normal VE, these two groups had very different RFF profiles. Thus, the greatest strength of RFF as a clinical assessment tool may be as a general identifier of individuals with voice disorders (including possible vocal hyperfunction), even at the mild end of the spectrum. This is ideal because typical listeners struggle to reliably perceive differences in VE and OS at the mild level (Kreiman et al., 1993). In these cases, RFF may serve as a subperceptual tool, which might have clinical utility for tracking individuals at risk.

RFF Measures and Laryngeal Tension: Summary and Future Work

Past work has shown that RFF discriminates between individuals with vocal hyperfunction and individuals with healthy voices (Stepp et al., 2010) and normalizes in individuals with vocal hyperfunction after a successful course of voice therapy (Stepp et al., 2011). Although the current study showed statistically significant correlations between offset RFF and perceptual measures of VE and OS, examination of data as a function of both VE and health status (voice disorder or control) indicated that the measure is more strongly associated with voice disorder than as a severity indicator of VE. These results indicate that RFF parameters may be most informative clinically as a diagnostic tool or measure of vocal risk but that they may be less informative as a feedback modality for voice patients during rehabilitation. However, until changes in RFF and perception of VE are examined within participants over time and over a greater spectrum of VE, it is unclear whether RFF can be a useful measure of rehabilitation progress. In addition, future work should examine the relationship between RFF measures and an individual with dysphonia's own judgment of VE because it is possible that other listeners (e.g., clinicians or inexperienced listeners) may not differentiate these judgments from OS of dysphonia. For example, in this study, listeners' judgments of perceived VE were strongly related to OS ($r^2 = .94$). Eadie et al. (2010) found that although experienced clinicians' and inexperienced listeners' judgments of perceived VE for dysphonic speakers were strongly correlated ($r = .91$), their ratings were only moderately correlated with the speakers' own judgments of VE ($r = .51$–$.56$). Because speakers have access to kinesthetic cues that might also relate to vocal hyperfunction, it is possible that self-judged VE might relate more strongly to measures of RFF. Future prospective studies incorporating RFF and both listener- and self-rated measures of VE throughout courses of therapy will allow determination of the utility of changes in RFF as a rehabilitation tool.

Although we have found that RFF measures are highly sensitive to voice disorders, the clinical utility is hampered by the current manual nature of RFF estimation. Currently, it is not automatic and requires time-consuming, cycle-by-cycle confirmation of period. We are currently examining specially designed stimuli for RFF elicitation that will allow development of automated RFF estimation techniques.

Our current model of RFF production is based on the linear superposition of tension, abduction effects, and aerodynamic factors (Stepp et al., 2011). It is possible that changes in tension do not scale linearly with changes in offset RFF, which could explain the obvious discrimination in offset RFF between voices with disorder and voices that are healthy, without a strong linear correlation between RFF and VE. As seen in Figure 4, onset RFF did not show obvious trends for stratification as a function of severity of perceived VE or disease state. This finding does not fit with our current model—that high-baseline laryngeal muscle tension might reduce the ability to create short-term variations in tension during devoicing, leading to a reduction in both offset and onset RFF in individuals with voice disorders. Instead, it is possible that changes in vocal fold tissue or kinematics in individuals with voice disorders are truly responsible for the differences in offset RFF. For instance, older speakers have also shown decreases in offset RFF (Watson, 1998) relative to younger speakers, but they have shown similar patterns of onset RFF. We do not believe that age is a major factor in the results of our present work because participants with a range of ages were part of both the voice disorder and control groups. However, it is possible that morphological changes or the changes to vocal fold kinematics that are present in older speakers may also contribute to changes in individuals with voice disorders across the life span. Further elucidation of this topic will be contingent on careful study of the physiological mechanisms of RFF in both healthy voices and voices with disorder at a variety of ages. One further explanation for decreased offset RFF would be systematic use of glottalization during vocal offset. Use of glottalization can result in lower F_0 values and has been associated with voice disorders (Ylitalo & Hammarberg, 2000). Although glottalization was noted in several samples, these samples were not used for analysis. It is not likely that systematic use of glottalization is a primary contributor to the differences in RFF found between control speakers and individuals with voice disorders. To understand the potential role of tension in RFF production, future studies—in which researchers collect RFF and conduct simultaneous electromyography of intrinsic and extrinsic laryngeal muscles under differing vocal conditions—are necessary in individuals with and without voice disorders.

Acknowledgments

We acknowledge funding support from the Department of Speech and Hearing Sciences at the University of Washington. We thank Gabrielle Hands of Boston University's Sensorimotor Rehabilitation Laboratory for acoustic analysis for interrater reliability of RFF and Jody Norwood for her help in the perceptual protocol in the Vocal Function Laboratory at the University of Washington.

References

Aronson, A. E. (1980). *Clinical voice disorders: An interdisciplinary approach*. New York, NY: Thieme-Stratton.

Awan, S. N., & Roy, N. (2009). Outcomes measurement in voice disorders: Application of an acoustic index of dysphonia severity. *Journal of Speech, Language, and Hearing Research, 52,* 482–499.

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

Baken, R. J. (1987). *Clinical measurement of speech and voice*. Austin, TX: Pro-Ed.

Bhuta, T., Patrick, L., & Garnett, J. D. (2004). Perceptual evaluation of voice quality and its correlation with acoustic measurements. *Journal of Voice, 18,* 299–304.

Boersma, P., & Weenink, D. (2008). Praat: Doing phonetics by computer (Version 5.0.20) [Computer software]. Retrieved from http://www.praat.org/.

de Krom, G. (1995). Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments. *Journal of Speech, Language, and Hearing Research, 38,* 794–812.

Dollaghan, C. A. (2007). *The handbook for evidence-based practice in communication disorders*. Baltimore, MD: Brookes.

Dworkin, J. P., Meleca, R. J., & Abkarian, G. G. (2000). Muscle tension dysphonia. *Current Opinion in Otolaryngology & Head and Neck Surgery, 8,* 169–173.

Eadie, T. L., & Doyle, P. C. (2002). Direct magnitude estimation and interval scaling of pleasantness and severity in dysphonic and normal speakers. *Journal of Speech, Language, and Hearing Research, 112,* 3014–3021.

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

Fairbanks, G. (1960). *Voice and articulation drillbook* (2nd ed.). New York, NY: Harper & Row.

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

Goberman, A. M., & Blomgren, M. (2008). Fundamental frequency change during offset and onset of voicing in individuals with Parkinson disease. *Journal of Voice, 22,* 178–191.

Halle, M., & Stevens, K. N. (1971). A note on laryngeal features. *MIT Research Laboratory of Electronics Quarterly Progress Report, 101,* 198–213.

Heman-Ackah, Y. D., Michael, D. D., & Goding, G. S., Jr. (2002). The relationship between cepstral peak prominence and selected parameters of dysphonia. *Journal of Voice, 16,* 20–27.

Hillman, R. E., Holmberg, E. B., Perkell, J. S., Walsh, M., & Vaughan, C. (1989). Objective assessment of vocal hyperfunction: An experimental framework and initial results. *Journal of Speech and Hearing Research, 32,* 373–392.

Holmberg, E. B., Hillman, R. E., Hammarberg, B., Sodersten, M., & Doyle, P. (2001). Efficacy of a behaviorally based voice therapy protocol for vocal nodules. *Journal of Voice, 15,* 395–412.

Kempster, G. B., Gerratt, B. R., Verdolini Abbott, K., Barkmeier-Kraemer, J., & Hillman, R. E. (2009). Consensus auditory-perceptual evaluation of voice: Development of a standardized clinical protocol. *American Journal of Speech Language Pathology, 18,* 124–132.

Koufman, J. A., & Blalock, P. D. (1991). Functional voice disorders. *Otolaryngologic Clinics of North America, 24,* 1059–1073.

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

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

Löfqvist, A., Koenig, L. L., & McGowan, R. S. (1995). Vocal tract aerodynamics in /aCa/ utterances: Measurements. *Speech Communication, 16,* 49–66.

Löfqvist, A., & McGowan, R. S. (1992). Influence of consonantal environment on voice source aerodynamics. *Journal of Phonetics, 20,* 93–110.

Morrison, M. D., Rammage, L. A., Belisle, G. M., Pullan, C. B., & Nichol, H. (1983). Muscular tension dysphonia. *Journal of Otolaryngology, 12,* 302–306.

Ohde, R. N. (1984). Fundamental frequency as an acoustic correlate of stop consonant voicing. *The Journal of the Acoustical Society of America, 75,* 224–230.

Ramig, L. O., & Verdolini, K. (1998). Treatment efficacy: Voice disorders. *Journal of Speech, Language, and Hearing Research, 41,* S101–S116.

Robb, M. P., Chen, Y., Gilbert, H. R., & Lerman, J. W. (2001). Acoustic comparison of vowel articulation in normal and reverse phonation. *Journal of Speech, Language, and Hearing Research, 44,* 118–127.

Robb, M. P., & Smith, A. B. (2002). Fundamental frequency onset and offset behavior: A comparative study of children and adults. *Journal of Speech, Language, and Hearing Research, 45,* 446–456.

Roy, N. (2003). Functional dysphonia. *Current Opinion in Otolaryngology & Head and Neck Surgery, 11,* 144–148.

Roy, N. (2008). Assessment and treatment of musculoskeletal tension in hyperfunctional voice disorders. *International Journal of Speech-Language Pathology, 10,* 195–209.

Shrivastav, R., & Sapienza, C. M. (2003). Objective measures of breathy voice quality obtained using an auditory model. *The Journal of the Acoustical Society of America, 114*(4, Pt. 1), 2217–2224.

Stepp, C. E., & Eadie, T. L. (2011, May). *Relative fundamental frequency as an acoustic correlate of vocal effort in spasmodic dysphonia*. Paper presented at the 161st meeting of the Acoustical Society of America, Seattle, WA.

Stepp, C. E., Hillman, R. E., & Heaton, J. T. (2010). The impact of vocal hyperfunction on relative fundamental frequency during voicing offset and onset. *Journal of Speech, Language, and Hearing Research, 53,* 1220–1226.

Stepp, C. E., Merchant, G. R., Heaton, J. T., & Hillman, R. E. (2011). Effects of voice therapy on relative fundamental frequency during voicing offset and onset in patients with vocal hyperfunction. *Journal of Speech, Language, and Hearing Research, 54,* 1260–1266.

Stevens, K. N. (1977). Physics of laryngeal behavior and larynx modes. *Phonetica, 34,* 264–279.

Verdolini, K., Titze, I. R., & Fennell, A. (1994). Dependence of phonatory effort on hydration level. *Journal of Speech and Hearing Research, 37,* 1001–1007.

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

Ylitalo, R., & Hammarberg, B. (2000). Voice characteristics, effects of voice therapy, and long-term follow-up of contact granuloma patients. *Journal of Voice, 14,* 557–566.