Perceptual and Quantitative Assessment of Dysphonia Across Vowel Categories

*Supraja Anand, †Mark D. Skowronski, ‡Rahul Shrivastav, and *David A. Eddins, *Tampa, and †Boca Raton, Florida, and ‡Athens, Georgia

Summary: Objectives. This study aims to determine the sensitivity of perceptual and computational correlates of breathy and rough voice quality (VQ) across multiple vowel categories using single-variable matching tasks (SVMTs).
Methods. Sustained phonations of /a/, /i/, and /u/ from 20 dysphonic talkers (10 with primarily breathy voices and 10 with primarily rough voices) were selected from the University of Florida Dysphonic Voice Database. For primarily breathy voices, perceived breathiness was judged, and for primarily rough voices, perceived roughness was judged by the same group of 10 listeners using an SVMT with five replicates per condition. Measures of pitch strength, cepstral peak, and autocorrelation peak were applied to models of the perceptual data.
Results. Intra- and inter-rater reliability were high for both the breathiness and the roughness perceptual tasks. For breathiness judgments, the effect of vowel was small. Averaged over all talkers and listeners, breathiness judgments for /a/, /i/, and /u/ were −11.6, −11.2, and −12.2 dB noise-to-signal ratio, respectively. For roughness judgments, the effect of vowel was larger. The perceived roughness of /a/ was higher than /i/ or /u/ by 3 dB modulation depth. Pitch strength was the most accurate predictor of breathiness matching ($r^2 = 0.84-0.94$ across vowels), and log-transformed autocorrelation peak was the most accurate predictor of roughness matching ($r^2 = 0.59-0.83$ across vowels).
Conclusions. Breathiness is more consistently represented across vowels for dysphonic voices than roughness. This work represents a critical step in advancing studies of voice quality perception from single vowels to running speech.

Key Words: Voice quality−Single-variable matching task (SVMT)−Vowel category−Cepstral peak−Pitch strength.

INTRODUCTION

Annually, 1 of 13 adults in the United States develop voice disorders,[1] with changes in voice quality (VQ) being one of the primary indicators of the presence of an underlying organic or functional problem.[2] Consequently, evaluation of VQ is a vital component of clinical diagnostics and VQ is an important outcome measure for surgical, pharmacological and/or behavioral treatment approaches. Breathiness and roughness are two major dimensions of VQ,[3] often assessed and monitored using a combination of perceptual and acoustic methods. Breathiness may be defined as an audible air escape in the voice, and roughness may be defined as the perceived irregularity of vocal fold vibrations.[3,4] Much of the VQ literature including our programmatic work on development of computational models of VQ perception using *matching* tasks has generally relied on sustained vowel phonations (specifically vowel /a/) for a variety of reasons. Sustained vowels are easy to produce, synthesize, analyze, and replicate across clinics or laboratories. Cognitive-linguistic processing demands are lower in vowel phonations and they provide stationary or "steady-state" stimuli that eliminate contextual articulatory (eg, consonant, dialect) and prosodic (eg, stress, rate) effects. The purpose of the current study was to examine the effects of vowel category (/a/, /i/, /u/) on perception of breathy, rough VQ using *matching* tasks[5,6] and their corresponding quantitative or computational correlates. It was hypothesized that VQ transcends phonetic information and can be judged reliably by listeners across vowels. This study was an intermediary step toward generalization of existing computational models of VQ perception to multiple phonemes and connected speech.

Vowel category and perception of VQ

Many research studies and clinical protocols use the Grade, Roughness, Breathiness, Asthenia, Strain (GRBAS)[3] or the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V)[4] to obtain perceptual judgments of the dysphonic voice. Studies using the GRBAS scale often have used vowels /a/ and/or /i/ along with connected speech.[7−9] Similarly, sustained vowels /a/ and /i/ are recorded in combination with six sentences and conversational speech in the CAPE-V protocol.[4] Despite the selection of multiple stimuli (especially multiple vowels) in both of these clinical tools, perceptual VQ ratings reported in the literature are often representative of "averaged ratings" across speech stimuli.[10,11] Other laboratory experiments on perception of breathy and rough VQ frequently have focused on the vowel /a/ to the exclusion of other speech sounds.[12−20] To date, only one study has examined perceived breathiness across vowels.[21] Although the results did not reveal significant differences across vowels—/a/, /i/, /æ/, and /o/, it is important to note that this study was conducted on healthy adults imitating different levels of breathiness. Similarly, perceived roughness across vowels (/u/, /i/, /ʌ/, and /æ/) has been investigated only on healthy male and female talkers.[22,23] Although an effect of vowel category was reported, the

Accepted for publication December 21, 2017.
Presented at the Voice Foundation 45th Annual Symposium: Care of the Professional Voice, Philadelphia, PA, June 2, 2016.

From the *Department of Communication Sciences and Disorders, University of South Florida, Tampa, Florida; †AventuSoft LLC, Boca Raton, Florida; and the ‡Office of the Vice President for Instruction, University of Georgia, Athens, Georgia.
Address correspondence and reprint requests to Supraja Anand, Department of Communication Sciences and Disorders, University of South Florida, 4202 East Fowler Avenue, PCD 1017, Tampa, FL 33620. E-mail: suprajaanand@usf.edu

Journal of Voice, Vol. 33, No. 4, pp. 473−481
0892-1997

https://doi.org/10.1016/j.jvoice.2017.12.018

direction of this effect was gender dependent. Vowel /a/ from male talkers was perceived rougher compared with other vowels,[23] and an opposite effect was observed in female talkers, with vowel /a/ being perceived as least rough among the vowels.[22] The effects of vowel category have not been reported for perceived breathiness or roughness in dysphonic talkers despite recordings of /a/ and /i/ in many protocols.[24–26]

These studies on the perception of VQ used n-point rating scales (eg, GRBAS), visual analog scales (eg, CAPE-V), and magnitude estimation tasks (eg, Hillenbrand et al[21]) to obtain listener judgments of breathiness and roughness. In a series of experiments, we have shown that such conventional approaches (1) result in arbitrary numbers (2) can easily be biased by context and other extraneous variables, and (3) suffer from the limitations of working with ordinal data.[18,27] An attempt to address these factors led to the development of single-variable matching tasks (SVMTs) to index vocal breathiness and roughness.[5,6,20,28] Unlike the multiparameter synthetic matching task,[16] SVMT requires listeners to manipulate a single parameter of a synthetic comparison stimulus until it matches the dysphonic voice. The value of this parameter at the point of subjective equality is considered as an index of the specific VQ magnitude. For breathiness, the noise-to-signal ratio (NSR) is varied, whereas for roughness, the amplitude modulation depth is varied to establish a perceptual match between the standard voice and the synthetic comparison stimulus. Thus, the SVMT provides context-independent and ratio-level data supporting precise quantification of change in both magnitude and direction of VQ. Although we have established the efficacy of SVMT for both breathiness and roughness perception, we utilized phonation samples of the vowel /a/ elicited from dysphonic speakers. The primary goal of this study was to determine the nature of any dependencies of breathy and rough VQ perception on vowel category or characteristics. VQ was assessed using vowels /a/, /i/, and /u/ that represent maximum deviations in the acoustic space and are most likely to lead to VQ variations, if any.

Vowel category and computational analysis of VQ

The use of vowel /a/ also extends to computational analysis of breathy and rough VQ dimensions.[15,17,29–34] Such strong preference for this vowel could be attributed to various factors such as (1) /a/ is an "open" vowel and is hence easily recognizable, (2) it is common across languages and its production minimally differs across languages or dialects, (3) it has normative values through conventional software (eg, *KAYPENTAX*; PENTAX Medical Americas, Montvale, NJ), and (4) it is used to explain perceptual variance of breathy and rough VQ where data are largely collected using sustained productions of /a/. Among the few studies that have investigated vowel category, inconsistencies across vowels have been reported for measures of acoustic perturbation (ie, jitter, shimmer) and measures of noise (eg, signal-to-noise ratio). Furthermore, these studies were performed on healthy voices.[35–42] Awan et al[43] analyzed cepstral peak prominence (CPP) from sustained vowels of /a/, /i/, /u/, and /æ/ using Hillenbrand's smoothed CPP measure. Although this study focused on vowel productions from healthy adults, there were significant differences between vowels—mean CPP values for low vowels such as /a/ were higher than high vowels /i/ and /u/. Acoustic measures extracted from multiple vowels produced by dysphonic talkers have not been reported with the exception of Solomon et al.[44] In that study, jitter and shimmer did not vary across vowels, but low-to-high ratio of spectral energy and CPP were significantly greater for /a/ compared with /i/. Spectral noise level was also greater for the low vowel (as demonstrated by low harmonics-to-noise ratio).

Accordingly, a secondary goal of this study was to determine the impact of vowel category on three computational analysis methods related to the breathy and rough VQ dimensions: (1) pitch strength (PS) from the auditory sawtooth waveform inspired pitch estimator prime (Aud-SWIPE′),[45] (2) interpolated cepstral peak (ICP),[46] and (3) log-transformed autocorrelation peak (ACP).[47] PS, analogous to tonality,[48] is related to the salience of pitch, varying on a scale from faint to strong,[48,49] and is strongly correlated with VQ judgments.[20,50] Aud-SWIPE′ is a biologically inspired pitch estimator that provides robust estimates of PS for severely dysphonic (ie, type 3) voices.[51,52] The cepstral peak prominence (CPP) is quantified as the normalized peak value in the spectrum of the log spectrum of a stimulus and this peak, associated with the fundamental stimulus period, is also related to VQ perception.[21,53] In the current study, cepstral peak estimated using an interpolated algorithm. Computational differences between conventional CPP and ICP are detailed in the Methods section. Finally, the autocorrelation function has a long history of use in studies of VQ.[21,37,47,54,55] The function is a time-domain method of self-similarity that returns a value of unity for perfectly periodic stimulus and lower values as periodicity deviates due to noise or deterministic variations. Thus, these measures focus on temporal (ACP), spectral (ICP), and tonality (PS) characteristics of the acoustic signal. To date, these three measures have not been used in concert to evaluate the same speech stimuli or to evaluate both the breathiness and the roughness VQ dimensions. One might expect the ACP characteristic to be related more closely to rough stimuli that have irregular variation in amplitude over time. Similarly, the ICP measure might be more closely related to breathiness and the corresponding changes in spectral shape.[34,56] Previous work revealed that PS was strongly correlated with perceived breathiness and moderately correlated with perceived roughness,[55] and thus one might expect that this measure will be related to the co-occurrence of the breathy and rough percepts in the same stimuli. Regression models were developed to examine which of these measures better predicted listener judgments of breathiness and roughness obtained via SVMTs.

METHODS

Sustained vowels

Sustained productions of the vowels /a/, /i/, and /u/ were selected from a large database of dysphonic voices recorded from the Ear, Nose, and Throat clinic at the University of Florida (University of Florida Dysphonic Voice Database). This database contains recorded samples of the vowel phonations along with read and spontaneous speech from 193 talkers with dysphonia (73 male and 120 female) resulting from various etiologies (eg, hyperfunctional voice disorders; vocal fold paralysis; spasmodic dysphonia; and presbyphonia). For the current experiment, 20 talkers, 10 with primarily breathy voices (5 male and 5 female; 68.5 ± 9.6 years) and 10 with primarily rough voices (5 male and 5 female; 62.0 ± 9.0 years), were selected using a stratified random sampling procedure conducted by three expert listeners for the vowel /a/ phonation. Stimuli were first categorized into two groups: primarily breathy and primarily rough. These were obtained through consensus judgements by all three experts. Next, the voices in each group were rated on a five-point scale to indicate the magnitude of breathiness and roughness. This allowed the selection of stimuli that spanned a wide range along the breathiness and roughness VQ continua. Samples of /i/ and /u/ from the same set of breathy and rough talkers were included in the current analysis. All stimuli were sampled at 20 kHz, 16-bit amplitude resolution, and 500-millisecond duration with 20-millisecond raised cosine onset/offset ramps.

Listeners

Ten listeners (6 male and 4 female; 23 ± 1.8 years) were recruited to participate in each of the perceptual experiments cited in the following paragraph and each participant consented to participation according to procedures approved by an institutional review board. All listeners were native speakers of American English and passed a hearing screening test (pure tone thresholds below 20-dB HL from 250 to 8000 Hz, ANSI, 2010). Stimulus presentation was controlled using the TDT *SykofizX* software and TDT System 3 hardware (Tucker-Davis Technologies, Inc., Alachua, FL). Listeners were seated in a sound-treated booth, and stimuli were presented monaurally using ER-2 insert earphones (Etymotic Research Inc, Elk Grove Village, IL) at 85-dB sound pressure level.

Perceptual tasks

Breathiness

The perception of breathy voice quality was measured using the SVMT.[5] This matching task compared a sustained vowel with a synthetic comparison stimulus (ie, a noisy sawtooth waveform with NSR as the single variable parameter or independent variable). The sawtooth stimulus, which contains all even and odd harmonics of the fundamental frequency (f_0), was constructed with an f_0 of 151 Hz and the white noise was filtered with a second-order low-pass filter (cut-off frequency: 151 Hz) to match the spectral slope of the sawtooth waveform. The range of possible NSR values was −40 to 10 dB NSR in 2-dB steps. On a given trial, each sustained vowel (1 of 10 talkers) was presented first, followed by a 500-millisecond silent gap, followed by the comparison stimulus at a particular NSR. Using an up-down adaptive tracking procedure, listeners were instructed to make an adjustment to the NSR until the perceived breathiness of the two stimuli was judged to be equal. The first trial of each adaptive track started with an initial NSR value of either −30 dB (low) or 0 dB (high), and results for the high- and low-initial values were averaged for each of five replicates of each stimulus. Replicates were tested consecutively for each initial value.

Roughness

The perception of rough voice quality was measured using a similar SVMT procedure.[6] In this SVMT, dysphonic sustained vowels were compared with an amplitude-modulated noisy sawtooth waveform where the single variable parameter was amplitude modulation depth. The sawtooth stimulus was constructed as described previously with an NSR of −20 dB for naturalness. The sawtooth stimulus was multiplied by the modulation function (Eq. 1) to impart the perception of roughness onto the reference waveform:

$$H(t) = 1 + m{*}\sin(2\pi ft + \phi)\hat{}4. \tag{1}$$

Here, $m \in [0, 1]$ is the modulation depth, f is the modulation frequency (25 Hz), and ϕ is the modulation phase (set to 0 radians). Modulation depth in dB: $m_{\mathrm{dB}} = 20\log_{10}m$. A fourth-order sinusoid was chosen for the modulation function to create a sufficiently rough reference stimulus compared with lower order sinusoids.[6] The range of possible modulation depths was −40 to 0 dB in 2-dB steps. As with breathiness, the standard vowel stimulus was compared with a comparison stimulus on each trial, and the modulation depth of the comparison stimulus was varied using an up-down adaptive tracking procedure until the roughness between the two stimuli was judged to be a perceptual match. The first trial of each adaptive track started with an initial modulation depth value of either −30 dB (low) or 0 dB (high), and the matching results were averaged across five replicates of each adaptive track.

For both the matching tasks, talker order was randomized. However, vowels were always presented in the same order for each talker (/a/ followed by /i/ and then /u/). For example, if the randomized talker order was T01, T05, T03, etc., then, /a/, /i/, and /u/ from T01 were presented first, followed by all vowels from T05 and then T03, etc., By presenting vowels for each talker in a sequence rather than presenting a single vowel from all talkers, any bias in perceptual measurements due to an "order effect" was minimized. For each VQ dimension, data were collected over two sessions on different days and was no longer than 2 hours.

Computational analyses

Pitch strength (PS)

PS estimates were obtained with the Aud-SWIPE′ algorithm.[45] In applying this algorithm to voice samples, as in Eddins et al, 2016[20] the spectrum of the stimulus is passed through an auditory front end that filters the spectrum analogous to the outer and middle ear transfer functions. Next, the algorithm processes the spectrum with a filter-bank analogous to the place map of the cochlea. Following principles detailed by Moore et al,[57] the spectrum is converted to specific loudness on an equivalent rectangular bandwidth frequency scale to transform the stimulus from an acoustic to a perceptual representation. Next, a series of sawtooth kernel functions are created that span a range of f_0 values, termed "pitch candidates" (48 candidates per octave, range 70−260 Hz). An analysis window (Hamming) of eight pitch periods in length is used to estimate the windowed stimulus spectrum. The pitch-synchronous spectral analysis reduces the interaction of window length and pitch period on the spectral estimate. The resulting specific loudness function is correlated with a sawtooth waveform representation (limited to only the prime-numbered harmonics). The normalized correlation value is referred to as PS, and the pitch candidate with the largest PS is referred to as the pitch height. Pitch height and PS were estimated regularly at a frame rate of 100 frames/s (10-millisecond frame offset), and the median PS over all frames of a stimulus was used to model the perceptual data.

Interpolated Cepstral Peak (ICP)

The interpolated CP (ICP) algorithm is similar to the CPP computation[21] with three key differences. First, the ICP algorithm uses a Hann window (instead of Hamming or rectangular windows) which tapers to zero at the end points, causing the spectrum of the window to roll off at a steeper rate than other windows. With a steep roll off of the windowed spectrum, the spectral leakage associated with a finite-length window is contained locally around the frequencies of individual harmonics in the stimulus spectrum. Thus, prominent spectral peaks do not mask fainter harmonics with significant spectral leakage which better preserves the dynamic range of harmonics in the stimulus spectrum. Second, the ICP algorithm zero-pads the log spectrum before calculating the spectrum of the log spectrum via an inverse fast Fourier transform (FFT). By zero-padding the log spectrum before the inverse FFT, the resultant cepstrum has higher resolution than without zero-padding and effectively interpolates the cepstrum around the narrow cepstral peak associated with the fundamental period of the stimulus. The ICP reduces the sensitivity of the peak to f_0. Third, the ICP algorithm does not normalize the cepstral peak value relative to a linear regression function of the cepstrum as in CPP. The primary purpose of linear regression normalization in CPP is to compensate for scaling issues associated with window length, FFT size, and whether an FFT or inverse FFT is used on the log spectrum. By properly accounting for such factors, the ICP algorithm may be compared directly to theoretic values of CP without the need for normalization.[46] In the current study, the following parameters were used: data resampled to 20 kHz, f_0 search range: 70−260 Hz, frame rate: 100 frames/s, 40-millisecond Hann window length, FFT size: 2^{11}, log spectral nulls clipped at 100 dB below the peak of the log spectrum, inverse FFT zero-padded to eight times FFT size. The median CP value over all frames was used to model the perceptual data.

Log-transformed Autocorrelation Peak (ACP)

In the current study, the autocorrelation function of each 500-millisecond sustained vowel stimulus was calculated and scaled to account for autocorrelation lag. The peak of the function was measured in the range of lags corresponding to the f_0 70−260 Hz and normalized by the zero-lag value. The normalized peak value P was log-transformed to account for ceiling effects as shown in Eq. (2):[47]

$$\mathrm{ACP} = -\log(1-P). \qquad (2)$$

ACP was used to model the perceptual data, providing a comparison of temporal, cepstral, and perceptually motivated acoustic measures.

Statistical analysis

Intraclass correlation was used to assess both intra- (comparison across the five replicates) and inter- (between participants) rater reliability[58,59] for the perceptual data. Repeated measures analysis of variance (ANOVA) was performed for individual voice quality dimensions to determine the effect of talkers (between-subject factor) and vowels (within-subject factor) on perceived breathiness or roughness. Similarly, for each voice quality dimension, any significant differences in computational measures between the vowels were examined via Pearson r correlation and a set of univariate ANOVA. Further, linear regressions were performed to model the perceptual data with each of the computational measures, and a one-way analysis of covariance was computed for analysis of the slopes of regression functions. Parametric analyses were performed owing to the normal distribution of the data (ascertained via skewness and kurtosis values).

RESULTS

Rater reliability

Intra- and inter-rater reliability measured using intraclass correlation for breathy and rough VQ are shown in Table 1.

TABLE 1.
Intra- and Inter-rater Reliability Described by Intraclass Correlation (2, K)

Reliability	Intra-rater (Mean ± SD)		Inter-rater (Mean)	
Vowel/VQ	Breathiness	Roughness	Breathiness	Roughness
/a/	0.982 ± 0.02	0.959 ± 0.04	0.953	0.841
/i/	0.934 ± 0.16	0.947 ± 0.08	0.936	0.843
/u/	0.946 ± 0.12	0.979 ± 0.01	0.930	0.875

For intra-rater reliability, K = 5 replicates, and for inter-rater reliability, K = 10 listeners.
Abbreviation: SD, standard deviation; VQ, voice quality.

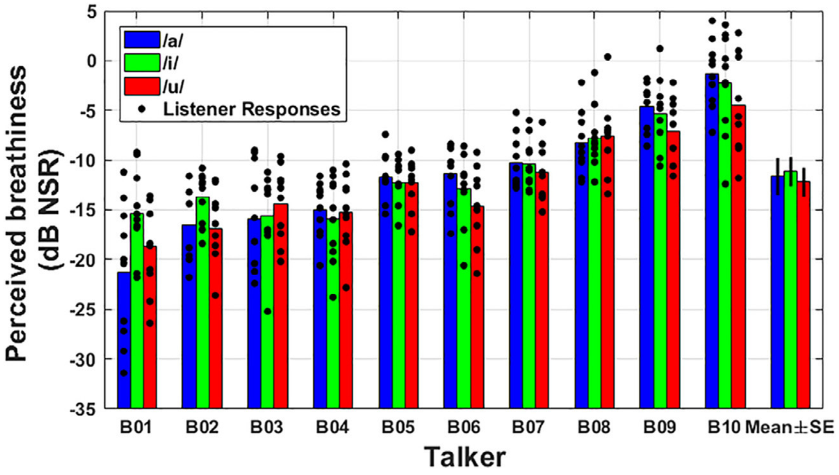

FIGURE 1. Perceived breathiness (dB NSR) versus talker. Bars represent mean judgments over all listeners, ordered from low to high for vowel /a/, and dots represent individual listener responses. The right-most bars represent mean ± SE over all talkers. NSR, noise-to-signal ratio; SE, standard error.

Intra- and inter-rater reliability were high for all vowels and for both VQ dimensions.

Comparison of VQ judgments across vowels

The breathy matching results are shown in Figure 1, ordered by judgments for the vowel /a/ among talkers. Inspection of the judgments of individual listeners (dots) shows that the variation among listeners is similar for all vowels and talkers with the exception of talker B01. This talker had the least amount of breathiness and therefore showed a wider variation among listeners (ie, listeners find it difficult to match the "breathiness" of a continuous comparison stimulus when there is little breathiness to match in the vowel stimulus). Breathy judgments for the vowel /a/ had a wider range of NSR values across talkers (B01 to B10) compared with the other vowels (the range of NSR values was −21.3 to −1.4 dB for /a/, −16.0 to −2.3 dB for /i/, and −18.7 to −4.5 dB for /u/). Despite the range differences, judgments were highly correlated among vowels: /a/—/i/: Pearson $r = 0.935$, /a/—/u/: $r = 0.962$, /i/—/u/: $r = 0.940$. Thus, it appears that breathiness was equally expressed among vowels (right-most bars; mean ± standard error [SE]). A repeated measures ANOVA confirmed that the within-subject factor "vowel" had a significant but small effect on breathy judgments: $F_{2,18} = 5.62$, $P_{GG} = 0.016$, $\eta^2 = 3.1\%$ (P_{GG} is the P value with Greenhouse-Geisser correction for sphericity). The between-subject factor "talker" had the largest effect on breathy judgments ($F_{9,81} = 45.96$, $P_{GG} < 0.0001$, $\eta^2 = 68.9\%$), and the interaction of the factors "talker" and "vowel" had a moderate effect ($F_{18,162} = 4.70$, $P_{GG} = 0.002$, $\eta^2 = 12.7\%$), which is in agreement with speaker-specific differences between breathiness judgments across vowels.

The rough matching results are shown in Figure 2, ordered by judgments for the vowel /a/ among talkers. The dots represent the judgments of individual listeners and show that the variation among listeners is similar for all vowels and talkers. The right-most bars (mean ± SE) indicate that perceived roughness was significantly higher for /a/ compared with the other vowels, although the rough judgment ranges were similar: /a/: −25.2 to −14.8 dB, /i/: −25.4 to −16.0 dB, /u/: −25.7 to −14.3 dB. Relative to breathiness

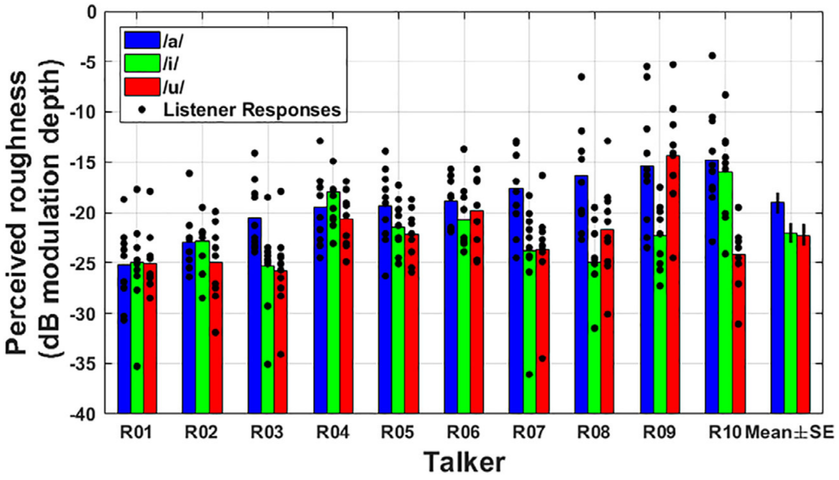

FIGURE 2. Perceived roughness (dB modulation depth) versus talker. Bars represent mean judgments over all listeners, ordered from low to high for vowel /a/, and dots represent individual listener responses. The right-most bars represent mean ± SE over all talkers. SE, standard error.

TABLE 2.
Mean ± SD Values for Each Computational Measure and Vowel for Both VQ Dimensions

Computational Analysis	Breathy			Rough		
Vowel →	/a/	/i/	/u/	/a/	/i/	/u/
PS	0.39 ± 0.19	0.38 ± 0.18	0.44 ± 0.20	0.24 ± 0.15	0.39 ± 0.17	0.43 ± 0.16
ICP	−15.96 ± 4.97	−16.64 ± 4.35	−15.87 ± 4.64	−18.57 ± 2.83	−14.41 ± 5.52	−16.62 ± 3.46
ACP	3.02 ± 1.41	3.69 ± 1.56	4.29 ± 1.78	1.63 ± 1.37	3.81 ± 1.90	3.95 ± 1.63

Abbreviation: SD, standard deviation.

judgments, correlations of roughness judgments were weaker across vowels /a/, /i/, and /u/: Pearson $r = 0.422$, /a/ −/u/: $r = 0.531$, /i/−/u/: $r = 0.205$. This indicates that the degree of roughness was not judged to be the same across vowels for all talkers. A repeated measures ANOVA indicated that the within-subject factor "vowel" had a significant effect on roughness judgments: $F_{2,18} = 18.15$, $P_{GG} = 0.0002$, $\eta^2 = 24.2\%$. The between-subject factor "talker" had the largest effect on breathy judgments ($F_{9,81} = 12.01$, $P_{GG} < 0.0001$, $\eta^2 = 36.6\%$), and the interaction of the factors "talker" and "vowel" had a large effect as well ($F_{18,162} = 11.74$, $P_{GG} < 0.0001$, $\eta^2 = 31.3\%$).

Comparison of computational analyses across vowels

Raw values of the three computational measures (PS, ICP, and ACP) for each vowel and VQ dimension are reported in Table 2. Further, a set of univariate ANOVA was used to examine the differences between the vowels /a/, /i/, and /u/ for each of the VQ dimensions.

There were no significant effects of breathiness across vowel category for any of the computational analyses (PS: $F_{2,29} = 0.277$, $P = 0.760$; ICP: $F_{2,29} = 0.081$, $P = 0.922$; ACP: $F_{2,29} = 1.589$, $P = 0.223$). Alternatively, for rough voices, there was a significant effect of vowels for PS and ACP measures ($F_{2,29} = 3.682$, $P = 0.04$; $F_{2,29} = 6.243$, $P = 0.006$), and no significant effect of vowels was found for ICP measure ($F_{2,29} = 2.571$, $P = 0.095$). Post hoc analyses (Bonferroni correction) revealed marginal significance between vowels /a/ and /u/ for PS ($P = 0.05$). Significant differences were observed between vowel /a/ and vowel /i/ ($P = 0.019$) and between vowel /a/ and vowel /u/ ($P = 0.012$) for ACP. Vowel /a/ had lower values of PS and ACP compared with vowels /i/ and /u/.

TABLE 3.
Pearson Correlation Coefficients (*r*) Grouped by Each Vowel Pair and VQ Dimension for Computational Analysis Methods

Computational Analysis	Breathy			Rough		
Vowel Pair→	/a-i/	/a-u/	/i-u/	/a-i/	/a-u/	/i-u/
PS	0.93*	0.90*	0.96*	0.38	0.67†	−0.03
ICP	0.84*	0.93*	0.89*	0.34	0.40	0.29
ACP	0.78*	0.69†	0.89*	0.47	0.46	0.16

* Significant at the 0.01 level (two-tailed).
† Significant at the 0.05 level (two-tailed).

For breathy and rough VQ, Pearson correlation between each vowel pair is provided for each of the computational measures (PS, ICP, and ACP) in Table 3. For breathy talkers, vowels /a/, /i/, and /u/ were highly correlated with each other ($r = 0.69-0.96$) for all three measures. On the contrary, for rough talkers, there was weak correlation among most of the vowel pairs for PS, ICP, and ACP ($r = -0.03$ -0.67). Likewise, the perceptual judgements of roughness varied markedly across the three vowel categories (ie, lower correlation among vowels for perceived roughness).

Population models

Population models, relating each computational analysis (PS, ICP, and ACP) to the mean judgment of each stimulus averaged over all listeners, were trained using linear regression functions and the results are shown in Figure 3. Individual models were fit to each vowel for each perceptual task and each computational analysis method. A one-way analysis of covariance for each task and measure indicated that all of the slopes of all of the regression functions in Figure 3 were significantly different from zero ($P < 0.001$), and the slopes among the three vowels were not significantly different ($P > 0.05$) from each other. For models of breathy judgments, PS produced the highest coefficient of determination (r^2: 0.849−0.946 among vowels), and ICP produced nearly as high goodness-of-fit terms (r^2: 0.644−0.822 among vowels). Note that both methods are sensitive to harmonic intensity in the frequency domain. ACP, on the contrary, produced much lower goodness-of-fit terms (r^2: 0.406 −0.714 among vowels) and is more sensitive to time domain stimulus variations. For models of roughness judgments, ACP produced the highest goodness of fit (r^2: 0.590−0.833 among vowels), and PS produced the second highest goodness of fit (r^2: 0.504−0.768 among vowels). ICP produced a much lower goodness of fit (r^2: 0.369−0.565 among vowels). For all tasks and acoustic measures, model fits were most accurate for the vowel /a/.

DISCUSSION

The current study investigated whether or not VQ perception generalizes across multiple steady-state vowels spoken

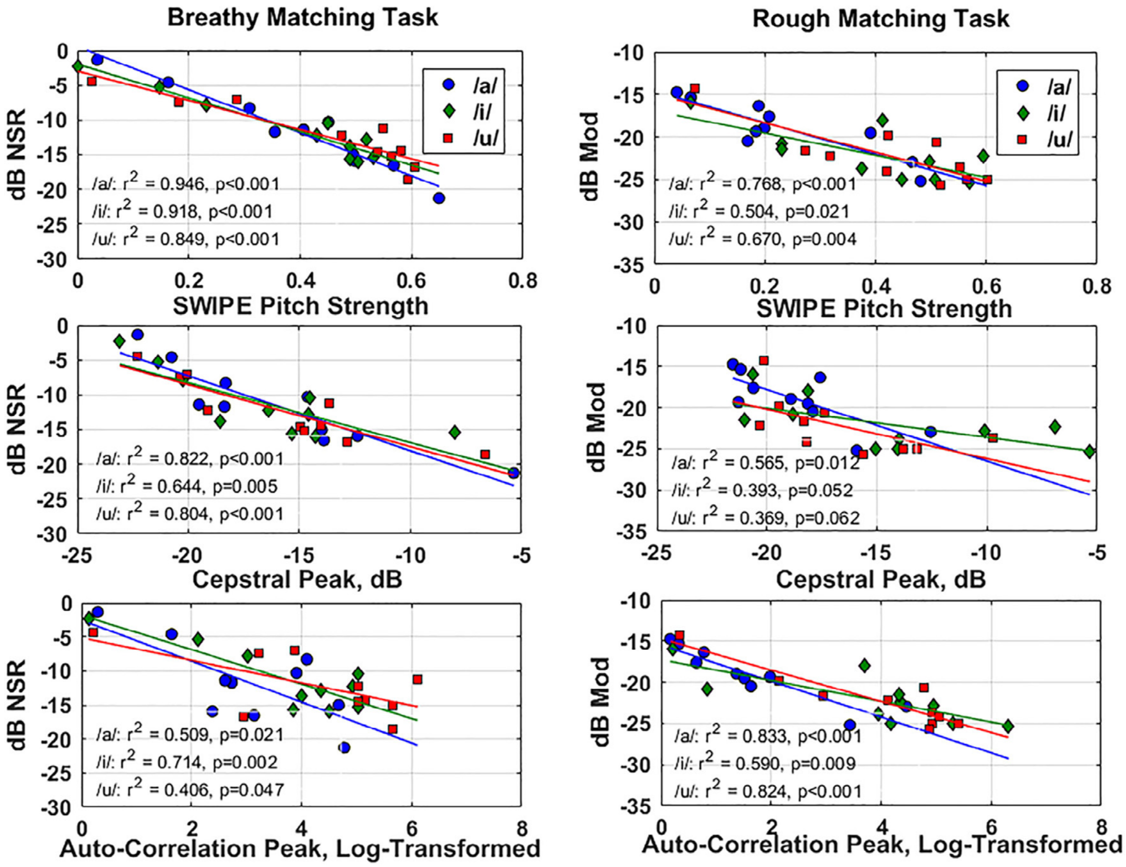

FIGURE 3. Perceived breathiness (dB NSR, left column) and perceived roughness (dB Mod-modulation depth, right column) versus computational analysis method. Top row: pitch strength; middle row: cepstral peak; bottom row: log-transformed autocorrelation peak. Points represent mean judgment for each talker, averaged over all listeners, and lines represent linear regression functions for each vowel. Inset text within each panel shows the goodness of fit for each vowel. NSR, noise-to-signal ratio; SE, standard error.

by talkers with a wide range of dysphonia when judged using an SVMT and whether or not vowel category impacts computational correlates of VQ perception. Perceptual judgments of breathiness were highly correlated among the three corner vowels and vowel category had minimal effect on VQ perception. These results are consistent with Hillenbrand et al and can be supported by a physiological explanation. Production of breathy VQ is fundamentally governed by modifications of the glottal source (ie, incomplete closure of the vocal folds) resulting in turbulent airflow. Accordingly, the changes in vocal tract configurations brought about by different vowels may have negligible effects on perceived breathiness. Unlike breathiness, perceived roughness was dependent on vowel category. Vowel /a/ was perceived to be significantly rougher compared with vowels /i/ and /u/. Further, there was a low correlation in perceived roughness among vowels. Studies on perception of roughness are limited *per se*. Similar to the current study, Sansone and Emanuel[23] showed that vowel /a/ was perceived to be rougher compared with other vowels (/u/, /i/, /ʌ/ , and /æ/) for their male talkers. These results could indicate a potential inverse relationship between tongue height and perceived roughness. On the contrary, vowel /a/ was perceived to be least rough for the female talkers in a similar experimental paradigm reported by Lively and Emanuel.[22] While the exact reasons for the gender differences are somewhat unclear, it is evident that perceived roughness may be affected by vowel category.

Although vowels /a/ and /i/ were recorded in a previous study,[10] acoustic measures extracted from multiple vowels of dysphonic talkers have not been reported except for Solomon et al.[44] The increase in mean CPP value for vowel /a/ compared with vowel /i/ over the wide range of dysphonia evaluated in their study is consistent with CPP values for healthy talker productions as reported by Awan et al. The authors attributed the increase in mean CPP to the emphasis of low frequency energy in the vowel /a/ relative to the other vowels evaluated, as well as greater sound pressure radiation resulting from open or increased size of the oral cavity (low tongue and jaw position) when producing /a/. In the current study, there were no effects of vowel category on any of the computational analyses for breathy talkers. On the contrary, the /a/ phonations of rough talkers had lower PS and ACP values compared with vowels /i/ and /u/. The variation in perceived roughness and corresponding variation in acoustic analyses across vowel category could potentially be a by-product of the stimulus selection process. The original association of voice quality dimensions to talkers was based on /a/ vowel phonations. The /i/ and /u/ phonations were sampled from the same talkers but may have differed from /a/ in voice quality.

Regression analyses indicated that perceived breathiness was more accurately modeled by PS estimates compared with ICP and ACP (higher r^2 for all vowels). These results are in agreement with previous literature that supports a high correlation between PS and perceived breathiness.[20,55] This is consistent with the suggestion that perceived breathiness of voiced speech is analogous to the degree of tonality of nonspeech sounds, both of which can be indexed by degree using PS. Perceived roughness was more accurately modeled by ACP compared with other two measures. Linear models appear to sufficiently relate the computational analyses to matching task judgments, using log transforms for CP and ACP measures, and are simpler compared with power law.[19,60] It is not surprising, given the temporal basis of ACP, that it provides the most robust correlation with perceived roughness, presumed to be a temporal envelope-based phenomenon. Overall, perceived breathiness was more accurately modeled compared with perceived roughness, likely due to higher intra- and inter-rater reliability and to the factors underlying the percepts. Breathy VQ is largely related to a single degree of freedom—airflow turbulence, whereas rough VQ can be associated with variations in multiple parameters such as amplitude modulations, bifurcations, jitter, and shimmer (although the latter two rarely provide a strong account for perceptual judgments). Thus, roughness may be considered to depend on more parameters than breathiness in terms of production, perception, and modeling.

There are several potential limitations of the current work worthy of consideration. First, a potential limitation common to many studies of dysphonic VQ is that several computational measures (eg, PS, CP, and ACP) may overlap in the individual VQ dimensions that they can capture. Furthermore, dysphonic voices can have covarying VQ dimensions with each dimension being cued by multiple acoustic or computational measures. Therefore, a modeling-based approach in future work may be appropriate to determine the discriminatory power of such computational measures in differentiating VQ dimensions. Furthermore, such modeling might also be used to examine the possible relationship between VQ severity and acoustic or computational measures. A logical next step would be to extend such VQ models to connected speech to determine any effects that complex articulatory (eg, consonant) and prosodic (eg, rate) properties may have on VQ perception and related computational measures.

CONCLUSIONS

Perceptual judgments of breathiness did not vary markedly across vowel category. Those judgments were more accurately modeled for vowel /a/ than /i/ or /u/ by the computational measures compared. Further, the high r^2 values (>0.80) for /i/ and /u/ obtained when comparing the breathiness judgments with pitch strength estimates indicate that an existing model of breathiness perception[20] based on pitch strength estimates will generalize to multiple vowels. Perceptual judgments of roughness did vary significantly across vowel category, with /a/ being judged as most rough among the three categories evaluated. The computational measures revealed similar differences among vowel categories and the temporal-based autocorrelation peak yielded the strongest model predictions among the computational measures evaluated.

Acknowledgments

This work was supported by the National Institute for Deafness and other Communicative Disorders R01DC009029 (RS and DAE). The authors would like to thank Taylor Traficant for assistance with data collection. The authors would also like to thank two anonymous reviewers for their suggestions.

REFERENCES

1. Bhattacharyya N. The prevalence of voice problems among adults in the United States. *Laryngoscope*. 2014;124:2359–2362.
2. Colton R, Casper J. *Understanding Voice Problems: A Physiological Perspective for Diagnosis and Treatment*. Baltimore: Williams and Wilkins; 1996.
3. Hirano M. *Clinical Examination of Voice*. New York: Springer-Verlag; 1981.
4. Kempster GB, Gerratt BR, Abbott KV, et al. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. *Am J Speech Lang Pathol*. 2009;18:124–132.
5. Patel S, Shrivastav R, Eddins DA. Developing a single comparison stimulus for matching breathy voice quality. *J Speech Lang Hear Res*. 2012;55:639–647.
6. Patel S, Shrivastav R, Eddins DA. Identifying a comparison for matching rough voice quality. *J Speech Lang Hear Res*. 2012;55:1407–1422.
7. De Bodt MS, Wuyts FL, Van de Heyning PH, et al. Test-retest study of the GRBAS scale: influence of experience and professional background on perceptual rating of voice quality. *J Voice*. 1997;11:74–80.
8. Wuyts FL, De Bodt MS, Van de Heyning PH. Is the reliability of a visual analog scale higher than an ordinal scale? An experiment with the GRBAS scale for the perceptual evaluation of dysphonia. *J Voice*. 1999;13:508–517.
9. Nemr K, Simoes-Zenari M, Cordeiro GF, et al. GRBAS and CAPE-V scales: high reliability and consensus when applied at different times. *J Voice*. 2012;26:812 e17−e22.
10. Wolfe V, Martin D. Acoustic correlates of dysphonia: type and severity. *J Commun Disord*. 1997;30:403–416.
11. Helou LB, Solomon NP, Henry LR, et al. The role of listener experience on Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) ratings of postthyroidectomy voice. *Am J Speech Lang Pathol*. 2010;19:248–258.
12. Klatt DH, Klatt LC. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J Acoust Soc Am*. 1990;87:820–857.
13. Kreiman J, Gerratt BR, Precoda K. Listener experience and perception of voice quality. *J Speech Lang Hear Res*. 1990;33:103–115.
14. Kreiman J, Gerratt BR, Precoda K, et al. Individual differences in voice quality perception. *J Speech Lang Hear Res*. 1992;35:512–520.
15. Hillenbrand J, Houde RA. Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. *J Speech Lang Hear Res*. 1996;39:311–321.
16. Gerratt BR, Kreiman J. Measuring vocal quality with speech synthesis. *J Acoust Soc Am*. 2001;110:2560–2566.

17. Shrivastav R, Sapienza CM. Objective measures of breathy voice quality obtained using an auditory model. *J Acoust Soc Am*. 2003;114:2217–2224.
18. Patel S, Shrivastav R, Eddins DA. Perceptual distances of breathy voice quality: a comparison of psychophysical methods. *J Voice*. 2010;24:168–177.
19. Shrivastav R, Camacho A, Patel S, et al. A model for the prediction of breathiness in vowels. *J Acoust Soc Am*. 2011;129:1605–1615.
20. Eddins DA, Anand S, Camacho A, et al. Modeling of breathy voice quality using pitch-strength estimates. *J Voice*. 2016;30:774.e1–774.e7.
21. Hillenbrand J, Cleveland RA, Erickson RL. Acoustic correlates of breathy vocal quality. *J Speech Lang Hear Res*. 1994;37:769–778.
22. Lively MA, Emanuel FW. Spectral noise levels and roughness severity ratings for normal and simulated rough vowels produced by adult females. *J Speech Hear Res*. 1970;13:503–517.
23. Sansone F Jr, Emanuel FW. Spectral noise levels and roughness severity ratings for normal and simulated rough vowels produced by adult males. *J Speech Hear Res*. 1970;13:489–502.
24. Deal RE, Emanuel FW. Some waveform and spectral features of vowel roughness. *J Speech Hear Res*. 1978;21:250–264.
25. Toner MA, Emanuel FW. Direct magnitude estimation and equal appearing interval scaling of vowel roughness. *J Speech Hear Res*. 1989;32:78–82.
26. Eskenazi L, Childers DG, Hicks DM. Acoustic correlates of vocal quality. *J Speech Lang Hear Res*. 1990;33:298–306.
27. Shrivastav R, Sapienza CM, Nandur V. Application of psychometric theory to the measurement of voice quality using rating scales. *J Speech Lang Hear Res*. 2005;48:323–335.
28. Eddins DA, Shrivastav R. Psychometric properties associated with perceived vocal roughness using a matching task. *J Acoust Soc Am*. 2013;134:EL294–EL300.
29. Herzel H, Berry D, Titze IR, et al. Analysis of vocal disorders with methods from nonlinear dynamics. *J Speech Lang Hear Res*. 1994;37:1008–1019.
30. de Krom G. Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments. *J Speech Lang Hear Res*. 1995;38:794–811.
31. Qi Y, Hillman RE. Temporal and spectral estimations of harmonics-to-noise ratio in human voice signals. *J Acoust Soc Am*. 1997;102:537–543.
32. Carding P, Steen I, Webb A, et al. The reliability and sensitivity to change of acoustic measures of voice quality. *Clin Otolaryngol*. 2004;29:538–544.
33. Awan SN, Roy N. Toward the development of an objective index of dysphonia severity: a four-factor acoustic model. *Clin Linguist Phon*. 2006;20:35–49.
34. Samlan RA, Story BH, Bunton K. Relation of perceived breathiness to laryngeal kinematics and acoustic measures based on computational modeling. *J Speech Lang Hear Res*. 2013;56:1209–1223.
35. Horii Y. Jitter and shimmer differences among sustained vowel phonations. *J Speech Hear Res*. 1982;25:12–14.
36. Sorensen D, Horii Y. Frequency and amplitude perturbation in the voices of female speakers. *J Commun Disord*. 1983;16:57–61.
37. Milenkovic P. Least mean square measures of voice perturbation. *J Speech Hear Res*. 1987;30:529–538.
38. Sussman JE, Sapienza C. Articulatory, developmental, and gender effects on measures of fundamental frequency and jitter. *J Voice*. 1994;8:145–156.
39. Gelfer MP. Fundamental frequency, intensity, and vowel selection: effects on measures of phonatory stability. *J Speech Lang Hear Res*. 1995;38:1189–1198.
40. Orlikoff RF. Vocal stability and vocal tract configuration: an acoustic and electroglottographic investigation. *J Voice*. 1995;9:173–181.
41. MacCallum JK, Zhang Y, Jiang JJ. Vowel selection and its effects on perturbation and nonlinear dynamic measures. *Folia Phoniatr Logop*. 2011;63:88–97.
42. Franca MC. Acoustic comparison of vowel sounds among adult females. *J Voice*. 2012;26:671 e9−e17.
43. Awan SN, Giovinco A, Owens J. Effects of vocal intensity and vowel type on cepstral analysis of voice. *J Voice*. 2012;26:670 e15−e20.
44. Solomon NP, Awan SN, Helou LB, et al. Acoustic analyses of thyroidectomy-related changes in vowel phonation. *J Voice*. 2012;26:711–720.
45. Camacho A. On the use of auditory models' elements to enhance a sawtooth waveform inspired pitch estimator on telephone-quality signals. *In Information Science, Signal Processing and their Applications (ISSPA)* 11th International Conference on (pp. 1080−1085); IEEE.
46. Skowronski MD, Shrivastav R, Hunter EJ. Cepstral peak sensitivity: A theoretic analysis and comparison of several implementations. *J Voice*. 2015;29(6):670–681.
47. Wolfe VI, Martin DP, Palmer CI. Perception of dysphonic voice quality by naive listeners. *J Speech Lang Hear Res*. 2000;43:697–705.
48. Zwicker E, Fastl H. *Psychoacoustics: Facts and Models*. New York: Springer-Verlag; 1990.
49. Fastl H, Zwicker E. *Psychoacoustics: Facts and Models*. Berlin: Springer; 2007.
50. Eddins DA, Vera-Rodriguez A, Skowronski MD, et al. Behavioral and computational estimates of breathiness and roughness over a wide range of dysphonic severity. *J Acoust Soc Am*. 2015;138:1809.
51. Eddins DA, Shrivastav R, Skowronski M, et al. *Using pitch and fundamental frequency to characterize dysphonic voices*. Paper presented at the The American Speech-Language and Hearing Association (ASHA) Convention; Orlando, Florida; 2014.
52. Kopf L, Shrivastav R, Eddins D, et al. *A comparison of voice signal typing and pitch strength Paper presented at the The American Speech-Language and Hearing Association (ASHA) Convention*. Orlando, FL; 2014.
53. Awan SN, Roy N. Acoustic prediction of voice type in women with functional dysphonia. *J Voice*. 2005;19:268–282.
54. Deliyski D. Acoustic model and evaluation of pathological voice production. *Proceedings of 3rd Conference on Speech Communication and Technology, EuroSpeech'93, Berlin, Germany*. 1993;3:1969–1972.
55. Shrivastav R, Eddins DA, Anand S. Pitch strength of normal and dysphonic voices. *J Acoust Soc Am*. 2012;131:2261–2269.
56. Heman-Ackah YD, Michael DD, Goding GS. The relationship between cepstral peak prominence and selected parameters of dysphonia. *J Voice*. 2002;16:20–27.
57. Moore BC, Glasberg BR, Baer T. A model for the prediction of thresholds, loudness, and partial loudness. *J Audio Eng Soc*. 1997;45:224–240.
58. Scheffe H. *The Analysis of Variance*. New York: Joh Wiley & Sons; 1959.
59. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86:420.
60. Shrivastav R, Camacho A. A computational model to predict changes in breathiness resulting from variations in aspiration noise level. *J dVoice*. 2010;24:395–405.