Objective Indices of Perceived Vocal Strain

[*]Supraja Anand, [†]Lisa M. Kopf, [‡]Rahul Shrivastav, and
[*]David A. Eddins, [*]*Tampa, Florida,* [†]*Cedar Falls, Iowa, and* [‡]*Athens, Georgia*

Abstract: Background. A limited number of experiments have investigated the perception of strain compared to the voice qualities of breathiness and roughness despite its widespread occurrence in patients who have hyperfunctional voice disorders, adductor spasmodic dysphonia, and vocal fold paralysis among others.

Objective. The purpose of this study is to determine the perceptual basis of strain through identification and exploration of acoustic and psychoacoustic measures.

Methods. Twelve listeners evaluated the degree of strain for 28 dysphonic phonation samples on a five-point rating scale task. Computational estimates based on cepstrum, sharpness, and spectral moments (linear and transformed with auditory processing front-end) were correlated to the perceptual ratings.

Results. Perceived strain was strongly correlated with cepstral peak prominence, sharpness, and a subset of the spectral metrics. Spectral energy distribution measures from the output of an auditory processing front-end (ie, excitation pattern and specific loudness pattern) accounted for 77−79% of the model variance for strained voices in combination with the cepstral measure.

Conclusions. Modeling the perception of strain using an auditory front-end prior to acoustic analysis provides better characterization of the perceptual ratings of strain, similar to our prior work on breathiness and roughness. Results also provide evidence that the sharpness model of Fastl and Zwicker (2007) is one of the strong predictors of strain perception.

Key Words: Listener perception−Strained voice−Spectral moments−Spectral sharpness−Cepstral peak prominence (CPP).

INTRODUCTION

Strain, along with breathiness and roughness, is one of the three primary dimensions used to describe dysphonic voice quality. These dimensions form the basis for subjective rating scales used in the clinical evaluation of dysphonic voice quality including the Grade, Roughness, Breathiness, Asthenia, Strain;[1] and the Consensus Auditory Perceptual Evaluation of Voice. [2] The voice qualities of breathiness and roughness have been the focus of far more perceptual investigations than strain. Vocal strain (pressedness) has been defined as the listener's perception of increased vocal effort or hyperfunction[1,3] and may be attributed to increased subglottal pressure and an increased degree of vocal fold adduction during speech.[4,5] Furthermore, changes in the epilarynx (eg, constriction) as well as changes in the vocal tract well above the vocal folds may influence the perception of strain.[6−8] The current study reports an initial attempt to identify objective indices that accurately reflect the perception of strain in dysphonic vowels.

To better understand the bases of differences in voice quality perception, objective analyses have been used in addition to or as a substitute for perceptual evaluations. Two common analysis approaches used to predict perceived voice quality dimensions include direct analyses of the voice acoustic signal[9-11] and analyses of the acoustic signal following an auditory processing front-end.[12-15] The latter takes into account several important transformations of the acoustic signal that occur prior to the formation of an internal percept of the sound quality. A number of studies have focused on acoustic and psychoacoustic indices of vocal breathiness and roughness[10,11,12,14,16-19] while only a few have directly investigated acoustic indices of strained voices.[3,5,20,21] Furthermore, to our knowledge, the current study is the first to examine psychoacoustic indices of vocal strain.

Acoustic indices of strain can be categorized into temporal, spectral, and cepstral domain measures. In the temporal domain, it has been shown that jitter and shimmer have significant but low to moderate correlation with perceptual strain ratings.[20-22] In the spectral domain, increased values of the first spectral moment (ie, the spectral center of gravity) have been related to increased perceived vocal strain.[5] Similarly, voice samples that were synthesized with progressively decreasing open quotient (duration of glottal opening divided by the length of the glottal period) resulted in an increase in the amplitude of higher harmonics. This led to a decrease in spectral tilt/slope.[6,23] These results are consistent with the notion that the perception of various degrees of vocal strain may be related to changes in the distribution of intensity across audio frequency. Along these lines, Hirano[1] reported that both an increase in the magnitude of high-frequency harmonics and increase in high-frequency noise results in increased perception of strain. Measures related to a cepstral representation of the signal represent a third class of acoustic measures applied to strain. In the simplest terms, the cepstrum can be used to quantify the pattern intensity

Accepted for publication June 7, 2018.

From the *Department of Communication Sciences and Disorders, University of South Florida, Tampa, Florida; †Department of Communication Sciences and Disorders, University of Northern Iowa, Cedar Falls, Iowa; 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 E. Fowler Avenue, PCD 1017, Tampa, FL 33620. E-mail: suprajaanand@usf.edu

Journal of Voice, Vol. 33, No. 6, pp. 838−845
0892-1997

https://doi.org/10.1016/j.jvoice.2018.06.005

variations across audio frequency. In the cepstral domain, measures of cepstral peak prominence (CPP), standard deviation of CPP (CPP SD), fundamental frequency (F_0), and standard deviation of signal-to-noise ratio were combined into a single index to classify 102 dysphonic voices.[20] Using this composite index, vowels with and without perceived strain could be classified with 80% accuracy. More recently, Lowell et al[3] used discriminant function analysis to compare 23 strained and 23 normal voices for both sustained vowels and connected speech. A three-variable model (CPP, CPP SD, and CPP F_0) could correctly classify vowel productions with 89.1% accuracy and a two-variable model (CPP SD and CPP F_0) could correctly classify sentence productions with 93.5% accuracy. On the contrary, Bhuta et al[24] reported that *none* of the acoustic indices (eg, measures of perturbation and soft-phonation index) extracted from the Multidimensional Voice Program (MDVP, KayPentax, Montvale, NJ) were able to reliably predict strain ratings.

The identification of objective measures of voice quality can provide valuable information to support our understanding of that quality and the acoustic and perceptual attributes that give rise to the quality. Furthermore, accurate objective measures of voice quality perception may be more reliable than human observers. Initial attempts, however, have shown weak associations between perception and the corresponding objective measure. Time-based perturbation measures showed only poor ($r = -0.13$[21]) to moderate ($r = 0.58$[20]) correlations with perceived strain severity. Cepstral measures, while more sensitive to perceived strain than other acoustic measures,[3,20] unfortunately are not specific to strain. Indeed, cepstral measures have been associated with the perception of roughness, breathiness, and the perception of overall dysphonia severity (eg,[9,25-28]). It is possible that such measures are strongly impacted by the frequent occurrence of phonatory breaks and instances of type II and type III phonations[29] associated with vocal strain, precluding accurate estimation using algorithms that are highly time-sensitive or based on estimates of fundamental frequency or periodicity. It is also possible that the inherent relationship between the voice acoustic signal and perception of sound quality necessarily depends on a combination of the linear and nonlinear transformations imposed by the auditory system prior to the formation of a percept of voice quality. Several studies of breathy and rough voice quality have shown that preprocessing vocal signals with an auditory front-end before acoustic analysis led to better predictors of voice quality perception.[13,15,30] Briefly, the application of an auditory processing front-end involves: (1) filtering the acoustic signal in a manner similar to the frequency-specific filtering of the outer ear and middle ear; (2) converting the signal from a linear to a nonlinear frequency scale that mimics the frequency to place map of the cochlea; (3) processing by a bank of band-pass filters to account for the filtering properties of the basilar membrane of the cochlea; and (4) estimation of an excitation pattern that represents the output to the auditory nerve. Following such an approach, Shrivastav and Camacho[30] compared measures of CPP to a measure of partial loudness that incorporated such an auditory front-end and showed that the estimated partial loudness (analogous to an internal representation of the harmonics-to-noise ratio; based on a loudness model by Moore et al.[31]. They demonstrated that estimates of partial loudness accounted for more variance in the perceived breathiness than the CPP measure. Similarly, estimating the perceived vocal strain may benefit from a signal that has been transformed by an auditory processing front-end may provide a more accurate representation of strain perception than directly estimating strain from the acoustic signal (described in the Methods section; Figure 2).

The search for the best objective measures for sound quality is not unique to the study of dysphonia and there is a long history of research into the concept of sound quality in the context of simple tones, complex tones, noise stimuli, and other sounds.[32] In this context, three primary dimensions of sound quality are described as tonality, roughness, and sharpness. The tonality of a sound is described as a continuum that bounded by the tonal or noisy quality of sounds.[33,34] Tonality has been conceptualized as relating to the salience of the pitch percept (ie, pitch strength) resulting from the sound. Similarly, roughness can be described as the perception of amplitude fluctuations of a rate greater than 20 Hz that can be influenced by, among other attributes, the modulation frequency (the fluctuation rate) and modulation depth (the extent of fluctuation). Finally, the perception of the sharpness has been related to the overall spectral envelope.[33,35] Sharpness of a given stimulus (S) is equated to the perceived sharpness of a narrowband noise (1 critical-band wide) with a center frequency of 1 kHz at a 60 dB sound pressure level. The numerical value of sharpness for a stimulus is calculated as the weighted first moment of the specific loudness distribution as formulated in Equation (1). This is conceptually similar to the relationship between perceived strain and the spectral center of gravity observed by Sundberg and Gauffin.[5] The conventional method of predicting sharpness, as described by Fastl and Zwicker,[33] uses a unit called “acum” (which means “sharp” in Latin), where sharpness, S, is quantified as shown in Equation (1).

$$S = 0.11 \frac{\int_0^{24\ Bark} N' g(z) z \, dz}{\int_0^{24\ Bark} N' dz} \ acum \tag{1}$$

Here, N' is the specific loudness for each critical band z on a bark scale and $g(z)$ is the additional weighting factor that is dependent on the critical-band rate and places extra weight upon high center frequencies (over 16 Bark). A detailed description of concepts such as specific loudness, critical-band, and their relation to sharpness is beyond the scope of the current manuscript and the reader is encouraged to review the chapter by Fastl and Zwicker.[33]

It is possible that the primary dimensions of sound quality in general also map onto the primary dimensions of voice quality (eg, tonality → breathiness; roughness → roughness; and sharpness → strain). Prior work has shown that the perceived tonality, quantified as pitch strength, and the perceived breathiness of 21 dysphonic voices were strongly and negatively correlated (Pearson's r = −0.99). The perceived pitch strength decreased as the perceived breathiness increased.[36] In a follow up study, computational estimates of pitch strength obtained from a sawtooth waveform-inspired pitch estimator with an auditory front-end (Aud-SWIPE[37]) successfully modeled the perception of breathiness in both synthetic and natural breathy stimuli.[12] Consider roughness in a similar way. Fastl and Zwicker[33] showed that the perceived fluctuation strength (analogous to roughness) of a set of nonspeech stimuli had a simple relationship to the amplitude modulation depth of the stimulus. A similar relationship between perceived vocal roughness in dysphonic voices and modulation depth was demonstrated by Eddins and Shrivastav.[17] By extension of the sound and voice quality analogy espoused above, vocal strain may be related to the perception of the spectral sharpness. If the perception of strain maps to an increase in intensity at high-frequencies,[5,6] corresponding to an increase in the first spectral moment, then one might view strain as an analog to the perception of sharpness in any acoustic stimulus. While a weighted first moment as shown in Equation (1) may be adequate for relatively simple tone and noise stimuli evaluated previously, sharpness of complex and temporally dynamic sounds such as dysphonic voices may be better described by higher spectral moments.

The current research builds on our previous work with two main objectives: (1) To better understand the perception of the strain voice quality, a characteristic of some dysphonic voices; and (2) To assess analytic methods that may be useful to quantify or predict the magnitude and direction of the strain percept. For the first objective, listeners rank-ordered a large set of dysphonic voices from low to high perceived strain (familiarization). They then judged the degree of perceived strain for a subset of the voices using a rating scale paradigm (experiment). For the second objective, a series of computational analyses were conducted in which: (1) sharpness was computed; and (2) spectral moments were computed from the raw acoustic waveform as well as the internal representation of that waveform at different stages of auditory processing using a multi-stage auditory processing model. Building on the analogy between sound and voice quality, it was hypothesized that the spectral moments computed following processing of the stimulus using an auditory front-end (eg, specific loudness pattern or model of spectral sharpness) would explain more variance and would outperform the conventional acoustic measures in modeling perceived vocal strain.

METHODS

Listeners

Eighteen undergraduate/graduate students from the University of Florida were recruited for the study. All listeners were native speakers of American English and all had hearing thresholds within normal limits (≤20 dB HL between 250 and 4000 Hz).

Stimuli

Forty-three samples of /a/ phonations representing a wide continuum of strain were selected from the Kay Elemetrics Disordered Voice Database (KEDVD; Kay Elemetrics, Inc., Lincoln Park, New Jersey). All stimuli were edited to be 500 milliseconds in duration (middle portion) and were down sampled to 24414 Hz (original sampling rate = 50 kHz) to match the available sampling rate of the TDT (Tucker-Davis Technologies, Inc., Alachua, FL) hardware.

Equipment and procedure

All procedures were approved by the University Institutional Review Board. All listeners consented to participation and were compensated for their time.

Familiarization

Prior to beginning the experimental task, all listeners completed a visual-sort-rank (VSR) task[38] to ensure familiarity with the strain voice quality dimension. The VSR task was conducted using a custom-designed graphical user interface (GUI; MATLAB, The MathWorks; Natick, MA) as shown in Figure 1.

Each button in the GUI depicted a strained voice sample. To hear a sample, a listener could simply use the mouse to press a button and the stimulus corresponding to the button label was played via PC sound card to the headphones (Sennheiser HD205) worn by the listener. At the beginning of the task, all buttons were located on the right side of the interface. Listeners were instructed to listen to the sound samples and to reposition the buttons on the left side of the interface in a rank order that reflected the magnitude of perceived strain from low to high. There was no time limit for completing the task and listeners could listen to each of the voice stimuli and reorder them multiple times before clicking the ‘DONE’ button. To familiarize the listeners with the concept of ranking the perceived strain in dysphonic voices, the VSR task initially involved feedback (Step 2) that indicated to the listener whether each rank-ordered stimulus was in the nominal order (relative to the expert consensus − third and fourth authors along with an additional expert with over 10 years of experience in assessing and treating dysphonic voices) or was ranked lower (indicating more strain − illustrated by pink color in Figure 1) or higher (indicating less strain − illustrated by blue color in Figure 1) than the expert consensus. This protocol allowed the listeners to hear the stimuli again and reorder them. On the next iteration (Step 3), the feedback only highlighted the stimuli that were assigned a different rank by the listener than the expert (illustrated by green color in Figure 1). Listeners could make changes to their responses one last time and provide a final response by clicking the ‘SUBMIT’ button. Listeners failing to achieve a criterion of 70% accuracy were disqualified from participation in the main experiment.

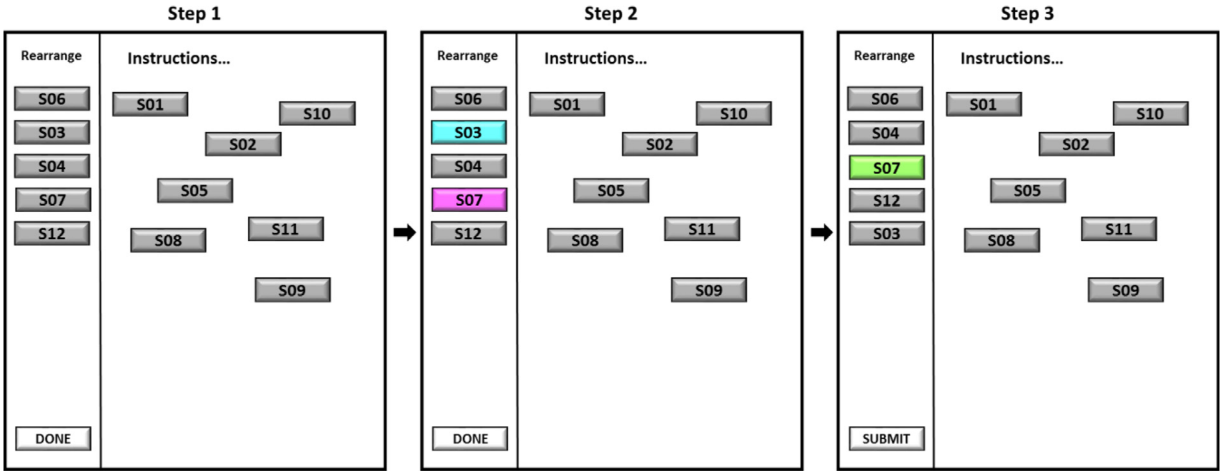

FIGURE 1. A sample interface for visual rank and sort (VSR) task. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

This unusual inclusion criterion was employed because of the difficulty naïve listeners have in understanding the concept of strained voice quality. Out of the 18 recruited listeners who underwent this VSR familiarization task, 12 passed this 70% criterion for participating in the main experiment (11 female, 1 male; mean age: 21.7 years).

Experiment

A subset of 28 phonation samples representing various levels of the strain continuum (low to high) were selected from the familiarization task and presented 10 times in random order for the main experiment. This main experiment was controlled through the TDT System III hardware and software. Hardware included RP2 processor, HB7 headphone buffer and Etymotic ER-2 insert earphones (Etymotic Research, Inc., Elk Grove Village, IL). Stimulus presentation and response acquisition was completed using the TDT SykofizX software application. Perceptual testing was conducted in a single-walled, sound-treated booth and stimuli were presented at 75 dB sound pressure level in the right ear. On each presentation, listeners judged the degree of perceived strain using a seven-point rating scale where 1 indicated "no strain" and 7 indicated "high strain". Ten ratings for each stimulus were averaged and this mean value from each listener was used for further analyses. Averaging allows for reduction of variance in the perceptual data.[39] Short breaks (3–5 minutes) were provided periodically to promote listener attention and to minimize fatigue. This rating scale task was completed by each listener in a single session of less than one hour.

Objective indices

The first, second, third, and fourth spectral moments were computed to identify which moment might best capture the perception of strain. Here, the second moment is related to the variance of the spectral shape, third moment related to the skewness of the distribution of spectral intensity, and fourth moment is related to the kurtosis or the variance of the variance of the spectral shape. Two additional measures were also computed: skewness of the spectral magnitude (normalized third moment) and kurtosis of the spectral magnitude (normalized fourth moment).

While Sundberg and Gauffin[5] computed the first spectral moment of the magnitude spectrum on a linear frequency scale, it may be useful to take into consideration the transformations of the spectrum that take place at various stages of the auditory system. To do so, a simple signal processing front-end commonly invoked in auditory research was considered. In concert, these transformations are used to estimate the internal spectrum, to a first approximation, at the level of the auditory nerve output, thereby including a series of nonlinear transformations mimicking the early stages of auditory processing that are likely to impact the stimulus percept and that would not be captured by traditional acoustic analyses. These steps include the following, as illustrated in Figure 2 and as detailed by Moore et al[31]: (1) computation of the vowel spectrum on a linear frequency axis

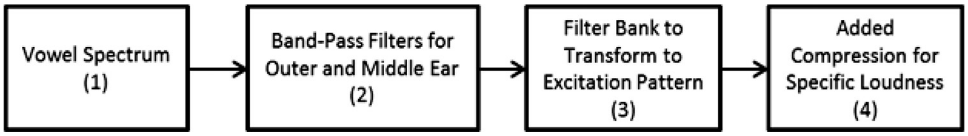

FIGURE 2. Schematic diagram of the auditory processing front-end model.

using a Fast Fourier Transform (FFT); (2) convolution with bandpass functions that model the filtering effects of the outer and middle ears and transformation to a nonlinear Bark scale to mimic the frequency warping when mapping audio frequency to cochlear space; (3) computation of the excitation pattern mimicking basilar membrane excitation; and (4) computation of the specific loudness pattern based on transformations analogous to auditory nerve output. Note that the specific loudness pattern is not an index of stimulus loudness but an estimate of the loudness of each spectral component following the transformations shown in Figure 2. All processing was completed using custom scripts executed in MATLAB (The MathWorks; Natick, MA).

For each stimulus, four different spectral representations were obtained from the analyses described above. The vowel spectrum on a linear scale, the filtered spectrum on a Bark scale, the excitation pattern, and the specific loudness pattern. Each of the six spectral moment measures was calculated following each of the four points of the auditory processing front-end as described above. This yielded a total of 24 moment measures as possible predictors of strain ratings. In addition, spectral sharpness served as the seventh predictor variable for the specific loudness pattern. Spectral sharpness was calculated using a weighting of the first moment of the specific loudness pattern as shown in Equation (1). The weighting accounts for a sharper increase in sharpness for critical-band rates above 16 Bark.[33] Finally, cepstral measures (CPP, CPP F_0) were also computed according to Hillenbrand, Cleveland, and Erickson[40] and compared with subjective ratings of strain.

Statistical analysis

Intra- (comparison across the 10 trials/repetitions) and inter- (among the 12 listeners) judge reliability for the perceptual ratings were assessed using Pearson's product moment correlation coefficient (r). Pearson's r was also used to examine the relationship between the objective indices and perceived strain ratings. Further, a series of stepwise linear regression analyses was computed (SPSS Version 22.0; IBM Corp, Armonk, NY) to determine which of the objective indices (independent variables) best predicted the perception of strain (dependent variable). Stepwise regression was conducted separately for each spectral representation (linear frequency scale, Bark scale, excitation pattern, and specific loudness pattern), with independent variable entry at $P = 0.05$ and exit at $P = 0.10$. For each analysis, the dependent variable was the average perceptual rating of strain across 120 trials (12 listeners, 10 repetitions each).

RESULTS

Perceptual ratings of strain

Figure 3 shows the judgments of perceived strain with average rating values on the ordinate and stimulus/talker on the abscissa (one through 28 ordered from low to high average strain rating). Error bars indicate the standard error of the mean computed across listeners. The distribution illustrates that listeners used the full range of ratings, from 1 to 7, indicating that robust differences in degree of strain were perceived across the stimulus set. Furthermore, the range of ratings was far greater than the range of ratings for any one stimulus, indicating a high degree of sensitivity to perceived strain among these listeners. The average intrajudge reliability was 0.85 and the average interjudge reliability was 0.59.

Identification of objective indices and modeling the perception of strain

Correlations between perceived strain and cepstral measures ranged from moderate to high (CPP F_0: $r = 0.44$, $P < 0.05$; CPP: $r = 0.80$, $P < 0.001$). Table 1 below depicts such correlations, illustrating the potential relationships between the dependent variable (strain ratings) and the other independent variables (spectral distribution metrics/moments and

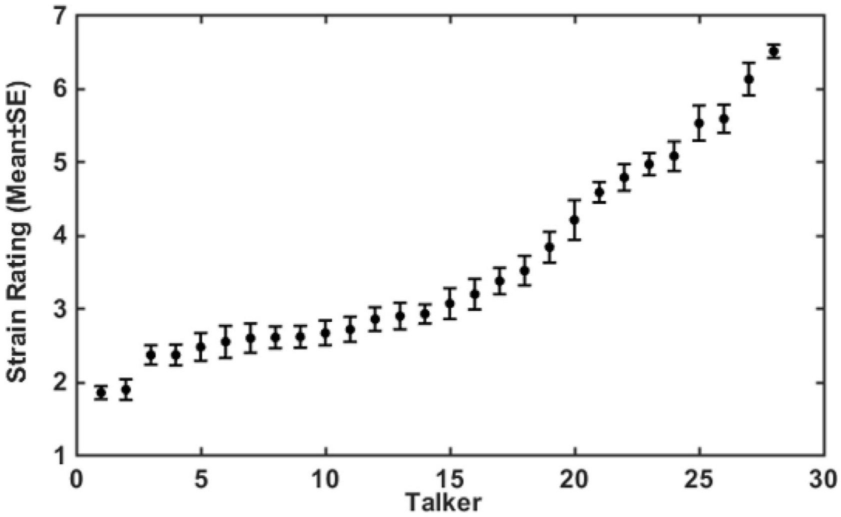

FIGURE 3. Perceived strain for all 28 stimuli/talkers averaged across 10 trials/repetitions per listener and averaged across 12 listeners (bars indicate standard error of the mean).

sharpness) for each of the four spectral transformations/representations. Furthermore, Table 1 also illustrates the results of the stepwise regression analyses where independent variable(s) that entered the stepwise regression analyses along with CPP are in bold and italicized font for each spectral transformations/representations.

For the vowel spectrum represented on a linear frequency scale (first column of Table 1), skewness and kurtosis in combination with CPP explained substantial perceptual variance, $r^2 = 0.79$ ($F_{1,24} = 29.663$, $P = 0.000$). Similarly, for the moments on the Bark scale (second column of Table 1), first moment and CPP resulted in highest $r^2 = 0.75$ ($F_{1,25} = 36.715$, $P = 0.000$). At the next stage of the auditory representation (ie, excitation pattern, third column of Table 1), the stepwise linear regression variables yielding the highest $r^2 = 0.77$ ($F_{1,24} = 27.320$, $P = 0.000$) included first moment, skewness, and CPP. Finally, when representing vowel stimuli in terms of their specific loudness patterns (fourth column of Table 1), first moment and CPP explained the largest amount of perceptual variance with a $r^2 = 0.79$ ($F_{1,25} = 45.624$, $P = 0.000$).

Note that the relationship between perceived strain and CPP was the strongest among all independent variables (Pearson's $r = 0.80$; $P < 0.001$). Accordingly, CPP variable entered all step-wise regression analyses (for all four spectral representations). It is important to note that spectral distribution measures that were significantly correlated with CPP did not enter the stepwise regression analyses because highly correlated predictors do not contribute substantially to the model. For example, the first moment in the linear spectral representation achieved highest correlation with perceptual data (Pearson's $r = 0.57$; $P < 0.01$) and yet did not enter stepwise regression analysis due to high correlations with CPP.

TABLE 1.
Relationships Among Perceived Strain Ratings and Spectral Distribution Measures (Rows) Organized by Spectral Transformation (Columns). For Each Spectral Distribution Measure, the Highest Observed Correlation With Perceptual Strain Rating is in Bold, and the Significance is Indicated by Asterisks Using the Following Code: * = P < 0.05; ** = P < 0.01; *** = P < 0.001. For Each Spectral Distribution Measure, Variable(s) that Entered Stepwise Regression Analyses Along with CPP are in Bold and Italicized Font.

Index	Linear	Bark	Excitation	Loudness
First moment	__0.57**__	___0.33___	___0.23___	___0.62***___
Second moment	0.48**	0.49**	0.44*	0.57**
Third moment	0.48*	0.40*	__0.47*__	0.21
Fourth moment	0.46*	__0.52**__	0.46*	0.43*
Skewness	−___0.43*___	0.22	___0.43*___	−0.50**
Kurtosis	−___0.55**___	−0.01	−0.023	−__0.67***__
Sharpness	-	-	-	__0.67***__

DISCUSSION

While considerable research has been completed to identify acoustic correlates of voice qualities such as breathiness (eg, [9,12,14] only a limited number of studies have directly investigated the perception of strain.[3] To date, there are no well-accepted acoustic or psychoacoustic correlates of strain. The results from this study suggest that listeners are reliable in judging varying degrees of strain. Furthermore, the variation in perceived strain across the 28 stimuli was high, and the correlations between perceptual ratings and spectral distribution measures ranged from low to moderately-high (Pearson's $r = -0.01$ to 0.67). Overall, the highest correlation between perceptual ratings and spectral distribution measures was sharpness along with kurtosis of the specific loudness pattern ($r = 0.67$), accounting for roughly half of the variance (45%) in perceptual ratings. The results provide support for the use of spectral sharpness, subset of spectral distribution measures to describe listener perception of strain severity. Furthermore, they support the transformation of the acoustic spectrum by an auditory processing front-end designed to approximate the frequency warping and nonlinear processes that give rise to an internal auditory representation of the original spectrum. The spectral center of gravity, quantified by the first moment of the spectrum, following transformation of the vowel spectrum into specific loudness pattern along with CPP explained largest amount of variance in perceived strain than other spectral metrics and other stages of the transformation process. The use of correlational analyses in these experiments, while useful, would be bolstered by additional experiments designed to understand the nature of relationship between the first spectral moment of the loudness-based transformation and strain perception.

It is common for? normal and dysphonic voices to have several salient voice quality percepts present simultaneously. For example, most severely rough voices are also breathy. Informal listening indicated that many of the strained voice samples in the current experiment also had salient breathiness and roughness. Similar constraints were also reported by Lowell et al's[3] study that revealed 52% of their "primarily strain" dysphonic samples as having co-occurring breathiness and roughness. It is likely that the presence of such covarying voice qualities may have caused CPP to be a stronger predictor of perceived strain. Therefore, future work in developing objective measures of strain may need to account for the potential influence of breathiness and roughness on strain perception. Once these dimensions are accounted for, it might be easier to determine which of the cepstral/spectral distribution measure(s) are the best predictor(s) of the strain perception *per se*.

Changes in the glottal adduction patterns and subsequent changes in laryngeal aerodynamics (lower airflow[4]) are often and primarily considered to result in the perception of strained or pressed voice and are characterized acoustically by the higher harmonic content (ie, flatter spectral slope). Additional adjustments such as a constriction in the epilaryngeal tube (cross-sectional area that extends from the vocal folds to the aryepiglottic folds) can also lead to similar changes in the acoustic spectrum (ie, increase in high frequency energy) and may likely influence the perception of strain.[6,7] Accordingly, future work combining perception and production might consider the effects of variations in the vocal tract on the perception of strain.

Sharpness is considered one of the primary elements of sound quality perception, along with tonality and roughness. Similar to the strong relationships between pitch strength and perceived vocal breathiness[12,36] as well as amplitude fluctuations and perceived roughness,[17] in this study, a strong relationship between spectral sharpness and the perception of vocal strain is identified. This relationship has not been demonstrated previously. Combined, these relationships between general sound quality perception and voice quality perception offer the opportunity to: (1) leverage existing computational methods that have proven to be beneficial in the world of psychoacoustics; and (2) develop new objective indices to evaluate voice quality that fit within the larger scope of sound quality perception.

In this attempt to understanding the perception of strain as well as determining objective indices, we examined multiple objective indices and identified the one that explained greatest amount of variance in perceived strain ratings. This preliminary study was designed to aid in the development of an appropriate synthetic stimulus to be used in single-variable matching tasks similar to our programmatic line of research on voice quality dimensions of breathiness and roughness. Although sustained phonations are simple and useful towards perceptual evaluation of the dysphonic voice, voice quality associated with connected speech may correspond more closely with perceived handicap and may represent more relevant treatment targets. Thus, a logical extension of this work is to investigate the perception of strain and its objective indices (acoustic and psychoacoustic) in continuous speech.

Acknowledgments

The work was supported by the National Institute for Deafness and other Communicative Disorders R01DC009029 (RS and DAE). Authors would like to thank Nour El-Bashiti for assistance with data collection and Mark Skowronski for assistance with signal processing.

REFERENCES

1. Hirano M. Psycho-acoustic evaluation of voice: GRBAS scale for evaluating the hoarse voice. *Clinican Examination of Voice*. New York: Springer-Verlag; 1981.
2. 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.
3. Lowell SY, Kelley RT, Awan SN, et al. Spectral-and cepstral-based acoustic features of dysphonic, strained voice quality. *Ann Otol Rhinol Laryngol*. 2012;121:539–548.
4. Netsell R, Lotz W, Shaughnessy AL. Laryngeal aerodynamics associated with selected voice disorders. *Am J Otolaryngol*. 1984;5:397–403.
5. Sundberg J, Gauffin J. Waveform and spectrum of the glottal voice source. *Speech Music Hearing Q Progress Status Report*. 1978;19:35–50.
6. Bergan CC, Titze IR, Story B. The perception of two vocal qualities in a synthesized vocal utterance: ring and pressed voice. *J Voice*. 2004;18:305–317.
7. Moisik SR, Esling JH. Modeling the biomechanical influence of epilaryngeal stricture on the vocal folds: A low-dimensional model of vocal−ventricular fold coupling. *J Speech Lang Hear Res*. 2014;57:S687–S704.
8. Titze IR. Nonlinear source−filter coupling in phonation: theory. *J Acoust Soc Am*. 2008;123:1902–1915.
9. Hillenbrand J, Houde RA. Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. *J Speech Lang Hear Res*. 1996;39:311–321.
10. Latoszek BBv, Maryn Y, Gerrits E, et al. The Acoustic Breathiness Index (ABI): a multivariate acoustic model for breathiness. *J Voice*. 2017;31. 511. e511-511. e527.
11. 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.
12. Eddins DA, Anand S, Camacho A, et al. Modeling of breathy voice quality using pitch-strength estimates. *J Voice*. 2016;30. 774. e771-774. e777.
13. Shrivastav R. The use of an auditory model in predicting perceptual ratings of breathy voice quality. *J Voice*. 2003;17:502–512.
14. 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.
15. Shrivastav R, Sapienza CM. Objective measures of breathy voice quality obtained using an auditory model. *J Acoust Soc Am*. 2003;114:2217–2224.
16. Eddins DA, Kopf LM, Shrivastav R. The psychophysics of roughness applied to dysphonic voice. *J Acoust Soc Am*. 2015;138:3820–3825.
17. Eddins DA, Shrivastav R. Psychometric properties associated with perceived vocal roughness using a matching task. *J Acoust Soc Am*. 2013;134:EL294–EL300.
18. Latoszek BBv, De Bodt M, Gerrits E, et al. The exploration of an objective model for roughness with several acoustic markers. *J Voice*. 2018;32:149–161.
19. Latoszek BBv, Maryn Y, Gerrits E, et al. A meta-analysis: acoustic measurement of roughness and breathiness. *J Speech Lang Hear Res*. 2018;61:298–323.
20. Wolfe V, Martin D. Acoustic correlates of dysphonia: type and severity. *J Commun Disord*. 1997;30:403–416.
21. Zwirner P, Murry T, Woodson GE. Perceptual-acoustic relationships in spasmodic dysphonia. *J Voice*. 1993;7:165–171.
22. Dejonckere PH, Remacle M, Fresnel-Elbaz E, et al. Differentiated perceptual evaluation of pathological voice quality: reliability and correlations with acoustic measurements. *Rev Laryngologie-otologie-rhinol*. 1996;117:219–224.
23. Karlsson I. Modelling voice variations in female speech synthesis. *Speech Commun*. 1992;11:491–495.
24. Bhuta T, Patrick L, Garnett JD. Perceptual evaluation of voice quality and its correlation with acoustic measurements. *J Voice*. 2004;18:299–304.
25. Awan SN, Roy N, Jetté ME, et al. Quantifying dysphonia severity using a spectral/cepstral-based acoustic index: comparisons with auditory-perceptual judgements from the CAPE-V. *Clin Linguist Phon*. 2010;24:742–758.
26. Deal RE, Emanuel FW. Some waveform and spectral features of vowel roughness. *J Speech Lang Hear Res*. 1978;21:250–264.
27. Hirano M, Hibi S, Yoshida T, et al. Acoustic analysis of pathological voice: some results of clinical application. *Acta oto-laryngol*. 1988;105:432–438.
28. Prosek RA, Montgomery AA, Walden BE, et al. An evaluation of residue features as correlates of voice disorders. *J Commun Disorders*. 1987;20:105–117.
29. Titze IR. Workshop on acoustic voice analysis: summary statement. *Natl Center Voice Speech* 1994;.

30. Shrivastav R, Camacho A. A computational model to predict changes in breathiness resulting from variations in aspiration noise level. *J Voice*. 2010;24:395–405.
31. 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.
32. Zwicker E, Fastl H. *Psychoacoustics: Facts and Models*. New York: Springer-Verlag; 1990.
33. Fastl H, Zwicker E. *Psychoacoustics: Facts and Models*. 3rd ed. Berlin: Springer; 2007.
34. Zwicker E, Fastl H. Pitch and pitch strength. *Psychoacoustics: Facts and Models*. New York: Springer-Verlag; 1990:102–132.
35. von Bismarck G. Sharpness as an attribute of the timbre of steady sounds. *Acta Acust United Acust*. 1974;30:159–172.
36. Shrivastav R, Eddins DA, Anand S. Pitch strength of normal and dysphonic voices. *J Acoust Soc Am*. 2012;131:2261–2269.
37. Camacho A. On the use of auditory models' elements to enhance a sawtooth waveform inspired pitch estimator on telephone-quality signals. *Paper presented at the 2012 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA); 2012*.
38. Patel S, Shrivastav R, Eddins DA. Identifying a comparison for matching rough voice quality. *J Speech Lang Hear Res*. 2012;55:1407–1422.
39. Shrivastav R, Sapienza CM, Nandur V. Application of psychometric theory to the measurement of voice quality using rating scales. *J Speech Lang Hearing Res*. 2005;48:323–335.
40. Hillenbrand J, Cleveland RA, Erickson RL. Acoustic correlates of breathy vocal quality. *J Speech Lang Hear Res*. 1994;37:769–778.