ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

Development and Validation of a Single-Variable Comparison Stimulus for Matching Strained Voice Quality Using a Psychoacoustic Framework

Yeonggwang Park,[a] Supraja Anand,[a] Sophia M. Gifford,[a] Rahul Shrivastav,[b] and David A. Eddins[a]

[a] Department of Communication Sciences & Disorders, University of South Florida, Tampa [b] Office of the Provost & Executive Vice President, Indiana University, Bloomington

ARTICLE INFO

Article History:
Received May 16, 2022
Revision received August 17, 2022
Accepted September 1, 2022

Editor-in-Chief: Peggy B. Nelson
Editor: Nancy Pearl Solomon

https://doi.org/10.1044/2022_JSLHR-22-00280

ABSTRACT

Purpose: Acoustic and perceptual quantification of vocal strain has been a vexing problem for years. To increase measurement rigor, a suitable single-variable matching stimulus for strain was developed and validated, based on the matching stimulus used previously for breathy and rough voice qualities.
Method: A set of 21 comparison stimuli for a single-variable matching task (SVMT) was synthesized based on a speech-shaped sawtooth waveform mixed with speech-shaped noise. Variable bandpass filter gain in mid-to-high frequencies achieved a wide range of computed sharpness (in constant sharpness steps) and served as the independent variable for the SVMT. Ten natural /ɑ/ stimuli with a wide range of the primary voice quality of strain and a minimum of breathiness or roughness were selected and assessed using the SVMT. Natural voice samples and synthetic comparison stimuli were also assessed using a perceptual magnitude estimation (ME) task.
Results: ME data validated the correspondence of the set of comparison stimuli to varying perceived strain. Perceived strain magnitudes of the comparison stimuli increased significantly and linearly with computed sharpness (r^2 = .99). A linear regression revealed that strain matching values were significantly predicted by computed sharpness (r^2 = .96) and perceived strain magnitudes (r^2 = .95) of the natural voice stimuli.
Conclusion: The perception of vocal strain is strongly associated with computed sharpness and is captured accurately and precisely using an SVMT, in which the independent variable is the bandpass filter gain (in steps of equal sharpness) applied to the comparison stimuli.

Vocal strain is an auditory-perceptual attribute of voice, defined as the "perception of excessive vocal effort," which is also termed *vocal hyperfunction* (Kempster et al., 2009). Excessive vocal effort is a common symptom associated with several voice disorders, such as vocal nodules and muscle tension dysphonia (Hillman et al., 2020; Morrison, 1997; Ramig & Verdolini, 1998). Because of the prevalence of increased strained voice quality in individuals with voice disorders, an accurate evaluation of vocal strain is crucial in diagnosis, assessment, and treatment. Among the voice evaluation methods, auditory-perceptual evaluation plays a key role in clinical assessment because it is easy to complete and can guide the diagnostic and treatment processes (Carding et al., 2009; Oates, 2009). However, current methods of the auditory-perceptual evaluation of strain may be problematic because of their low reliability (Dahl et al., 2021; Kelchner et al., 2010; Webb et al., 2004; Zraick et al., 2011). Indeed, the reliability of the auditory-perceptual evaluation of voice is a common topic in the literature (e.g., Chan & Yiu, 2002; dos Santos et al., 2019; Kapsner-Smith et al., 2021; Kreiman et al., 1992).

Conventional auditory-perceptual evaluation methods, such as *N*-point rating scales or visual analog scales, require the arbitrary assignment of a number to perception, which reduces the reliability of the judgments due to internal and

Correspondence to Yeonggwang Park: park21@usf.edu. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

context biases (Gerratt et al., 1993; Kreiman & Gerratt, 1998; Kreiman et al., 1992). Because of the random assignment of numbers and possible effects of biases, the auditory-perceptual ratings used in clinics are recommended to be treated as ordinal data that cannot accurately represent the magnitude of changes (Shrivastav et al., 2005). The accurate representation of the magnitude of change in voice quality is necessary to serve as an important outcome measure in voice treatment and therapy.

Matching Tasks

To address these problems of conventional methods, matching tasks have been recommended (Kreiman & Gerratt, 1996; Patel et al., 2012a, 2012b). In matching tasks, listeners compare and match a test stimulus (i.e., voice of interest) to a reference stimulus (i.e., synthetic comparison sound). The matching process involves a systematic change of one or more acoustic features (termed *independent variables* in psychology and the study of sensory perception) of the comparison sound until a suitable match to the test stimulus is identified. The physical property of the comparison sound (i.e., the independent variable) at the point of subjective equality between the two stimuli is used to quantify the magnitude of the perception. Matching tasks have advantages over rating or visual analog scales in that they do not involve the arbitrary assignment of numbers; rather, the physical property that is used to produce a perceptual match is associated with specific physical units. This reduces internal and context biases and results in improved reliability (Kreiman & Gerratt, 2005; Kreiman et al., 2007; Patel et al., 2010). Moreover, values of the independent variable, corresponding to parameters of the comparison sound, exist on both an interval and a ratio-level scale, allowing mathematical comparisons of voice quality within individuals over time, across individuals, and among judges.

Matching tasks require a comparison stimulus that can be varied using either a single independent variable or multiple independent variables (i.e., parameters) to achieve a perceptual match. Although multiparametric matching tasks have been performed previously (Kreiman & Gerratt, 2005), those tasks may result in multiple different solutions in the match to a single stimulus. Single-variable matching tasks (SVMTs) are often used because of their simplicity in performance and interpretation of the data. Adjusting a single variable reduces task time compared with adjusting multiple variables to achieve a perceptual match. The physical units of the single variable allow for the estimation and quantification of perceptual distances between stimuli, contributing to the development of objective correlates and computational models of voice quality.

One of the comparison stimuli successfully used in SVMTs to quantify voice quality combines a sawtooth waveform with speech-shaped noise. The combination was low-pass filtered to match the general spectral slope of disordered voices. For quantifying breathy voice quality, the signal-to-noise ratio (SNR; dB SNR) of the sawtooth waveform mixed with noise was used as the independent variable in the SVMT (Patel et al., 2012a). For quantifying rough voice quality, the depth of amplitude modulation (dB) superimposed on the sawtooth waveform mixed with noise has been used as the independent variable in the SVMT (Patel et al., 2012b). A suitable independent variable for matching strained voice quality has yet to be identified, yet development of a matching stimulus for strained voice quality would be most helpful as strain is the least reliable of the voice quality dimensions (Dahl et al., 2021; Kelchner et al., 2010; Webb et al., 2004; Zraick et al., 2011).

Sharpness

Sharpness is an auditory percept that contributes to the umbrella concept of timbre, related to the density of high-frequency spectral content over the full spectral content of a sound (Fastl & Zwicker, 2007). By convention, the perceived sharpness of an arbitrary sound can be expressed on a scale of sharpness that has, as an acoustic reference, narrowband noise that is one auditory critical band wide, centered at 1000 Hz, and presented at a level of 60 dB SPL (Fastl & Zwicker, 2007). That reference sound has a sharpness of 1 acum (the Latin word meaning "sharp"). The sharpness scale is analogous to the sone scale of loudness, which can be applied to quantify the loudness of any arbitrary sound in relation to a 1000-Hz tone with a level of 40 dB SPL. The sharpness of a sound can also be estimated in units of acum by calculation through a model developed by Fastl and Zwicker (2007). This model equation of general sound sharpness has been extended to voice signals and applied to estimate sharpness of natural voice samples (Anand et al., 2019).

Kopf et al. (2013) hypothesized that the perception of sharpness would be associated with strained voice quality due to the link between changes in the spectral envelope and increased perceived strain in voices (i.e., reduced spectral tilt). Anand et al. (2019) investigated the relationship between sharpness computed from the model of Fastl and Zwicker (2007) and vocal strain and observed that sharpness computed from this model was strongly correlated with listener perception of strain based on perceptual judgments using a Likert rating scale from 1 (*least strain*) to 7 (*most strain*). The hypothesized relationship between sharpness and vocal strain is consistent with more recent analyses of strained voice quality, which observed increased mid- to high-frequency spectral energy in strained voices (Lowell, Kelley, Awan, et al., 2012; McKenna & Stepp, 2018). The physiological mechanisms behind this relationship

may be increases in the degree of laryngeal muscle activity and supraglottal compression when vocal effort is increased (Hillman et al., 2020; McKenna et al., 2016, 2019; Stager et al., 2000), both of which may increase the high-frequency content of the voice (Fant et al., 1985; Zhang, 2016b).

Purpose

Figure 1 summarizes the hypothesized relationship between sharpness and vocal strain in the physiological, acoustic, and perceptual domains. Recognizing the strong association of sharpness computed from the model of Fastl and Zwicker (2007) and vocal strain, the current investigation evaluates the use of an SVMT in which the sharpness of the comparison stimulus is manipulated to achieve a perceptual match to the perceived strain of a target voice sample. Specifically, the gain of a filter applied to the spectrum of the sawtooth-plus-noise comparison sound is adjusted to obtain that perceptual match. By adjusting the filter gain, a set of comparison stimuli with a wide range of computed sharpness values was developed. We hypothesized that the range of the developed comparison stimuli would correspond to a wide range of perceived vocal strain obtained from the magnitude estimation (ME) task. We also evaluated a hypothesis that the computed sharpness of natural voice samples would be strongly and significantly correlated with perceived strain magnitudes.

Using the developed comparison stimulus, an SVMT is used to quantify perceived strain in the same set of natural voice samples and to test additional hypotheses. First, if the chosen independent variable maps onto the percept of vocal strain, then there should be strong inter- and intrarater reliability based on the obtained matching values. Second, because the set of independent variables of the comparison stimuli was chosen to be linearly associated with computed sharpness values, sharpness values computed from the actual voice samples should be strongly and significantly correlated with strain matching values. Third, if sharpness is a binding concept for vocal strain, then strain magnitudes from the ME task should be strongly and significantly correlated with strain matching values, which use a sharpness-based comparison sound as a reference. If these hypotheses are supported, then the SVMT applied to the strained voice quality will allow for a more rigorous evaluation of strained voice than prior psychophysical methods, and the accompanying model of strain, computed spectral sharpness, may lead to improved objective indices of strained voice quality.

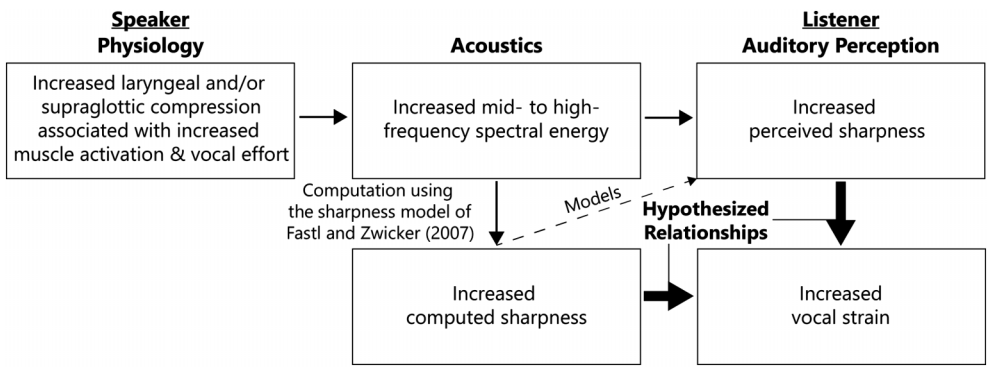
Method

Comparison Stimuli

Similar to the previous SVMTs for measuring breathiness and roughness, the comparison sound is based on a base stimulus consisting of a sawtooth wave with f_o = 151.8 Hz mixed with noise with the same spectral slope as the sawtooth waveform and an SNR of 20 dB (Patel et al., 2012a, 2012b). The mixture was filtered with a second-order Butterworth low-pass filter with a cutoff frequency of 151.8 Hz at a rate of −12 dB to better approximate the average spectral slope observed in disordered voice samples.

To vary the sharpness of the base stimulus used in the prior comparison sounds (sawtooth wave plus noise), the gain of an irregularly shaped bandpass filter was varied systematically. The gain was always 0 dB or greater, meaning that when greater than 0 dB, the filter increased the magnitude of the stimulus within its passband. The bandpass filter applied to the sawtooth-plus-noise stimulus also created substantial increases in perceived loudness, even when the root-mean-square level of the sound was normalized to a single level. To better control for loudness, the intensity of each of the filtered stimuli spanning

Figure 1. A schematic diagram summarizing the hypothesized relationship between sharpness and vocal strain in the physiological, acoustic, and perceptual domains.



the entire range of possible independent variable values (filter gain) was normalized to a single loudness level, or phon level, which was computed using the loudness model of Moore et al. (1997). By systematically varying the gain of the filter that was applied to the sawtooth-plus-noise complex, a series of stimuli was created that varied in perceived sharpness. The shape of the filter included a nominal passband that extended from a low cutoff frequency of 3000 Hz to a high cutoff frequency of 6000 Hz. Within this passband, a tilt was applied such that the gain at 6000 Hz was 1.23 times the gain at 3000 Hz (e.g., for a 20-dB gain at 3000 Hz, the tilt over that range was from 20 to 24.67 dB at 6000 Hz). A shallow roll-off was implemented from 3000 Hz down to 900 Hz, which was 0.44 times the gain at 3000 Hz (e.g., for a 20-dB gain at 3000 Hz, the gain at 900 Hz was 8.89 dB). Below 900 Hz, there was a steeper roll-off such that the gain at 1 Hz was 0 dB. Between 6000 and 9000 Hz, the roll-off was 0.45 times the gain at 6000 Hz (e.g., 24.67 dB at 6000 Hz down to 11.22 dB at 9000 Hz). The gain decreased from the value at 9000 Hz to 0 dB at the Nyquist rate (12207 Hz).

The stimuli were then analyzed in terms of the objective measure of computed sharpness using the model put forth by Fastl and Zwicker (2007) to compute sharpness in units of acum. A goal was for the comparison stimuli to have computed sharpness that spans a wide range of sharpness values and for those computed sharpness values to be systematically spread throughout the range. Fastl and Zwicker described a model of sharpness, S, based on the spectral envelope of the stimulus. Although specific spectral components, including harmonics, may influence the spectral envelope, the model was designed to be insensitive to the fine spectral details of the stimulus. In addition, the model is designed to capture features of the auditory system and auditory processing, and thus, the spectrum is computed on the basis of the presumed excitation pattern, which reflects the sensitivity and filtering properties of the cochlea, and early neural excitation, transformed on the basis of loudness perception. Thus, spectral analysis is based on estimates of auditory filtering, represented by a series of cascaded filters separated by auditory "critical bands" expressed on a Bark scale (Zwicker, 1961) and the perception of loudness. The Bark scale is analogous to mel (Stevens et al., 1937) and equivalent rectangular bandwidth (Moore & Glasberg, 1983) scales. Each of these scales transforms audio frequency into critical-band rate (i.e., discrete filter number along the continuous Bark axis) spanning the range of center frequencies from 50 to 13500 Hz. In terms of sharpness, S, the model below maps a high-frequency weighted transform of specific loudness, N', onto the Bark scale, where specific loudness is a logarithmic transform of band-specific excitation. The output is somewhat similar to the more recent specific loudness model of Moore et al. (1997) used in estimates of the breathiness of voiced stimuli (e.g., Shrivastav, 2003). In the model description below, the high-frequency weighted first moment of specific loudness is normalized by total loudness when expressed. This is expressed mathematically in Equation 1 (Fastl & Zwicker, 2007):

$$S\,[\text{acum}] = 0.11\frac{\int_0^{24\ \text{Bark}} N'g(z)z\,dz}{\int_0^{24\ \text{Bark}} N'\,dz}. \quad (1)$$

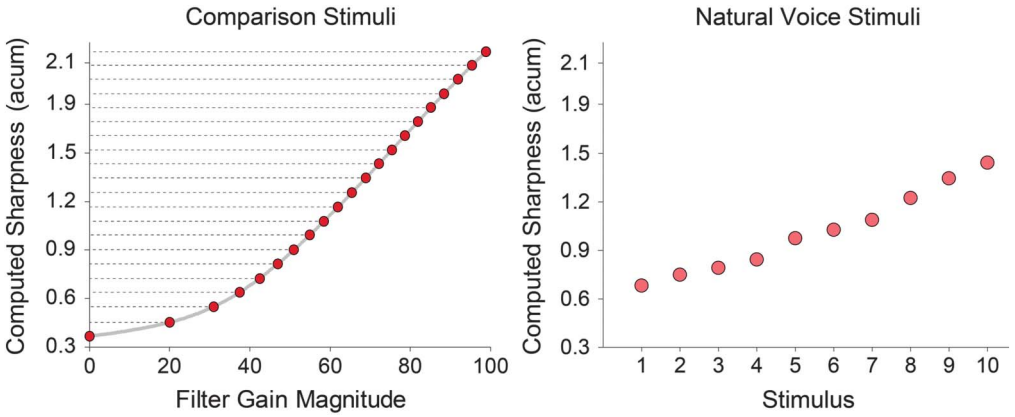
In this equation, the stimulus spectrum is represented on a Bark scale, from 0 to 24 Bark units. Sharpness, S, is defined in terms of total loudness, specified in the denominator and computed as the integral of specific loudness, N', over the frequency range from 0 to 24 Bark units. Fastl and Zwicker (2007) describe the numerator as a quantity similar to the first moment (analogous to the arithmetic mean) of specific loudness as a function of critical-band rate, z, on that same Bark scale. That quantity includes the weighting factor $g(z)$, which is dependent on the critical-band rate, z, and effectively produces a weighting function with a high-frequency emphasis. That weighting factor departs from a value of 1.0 beginning at a center frequency of approximately 16 Bark units and grows exponentially to a value of 4 as Bark units increase (Fastl & Zwicker, 2007).

As shown in the left panel of Figure 2, the final set of comparison sounds consisted of 21 stimuli (red markers) that varied in computed sharpness, S, from 0.37 to 2.12 acum in steps of between 0.8 and 0.9 acum. An informal perceptual evaluation of this set of comparison stimuli indicated that this range of computed sharpness exceeded the range of computed sharpness in the set of strained voices used by Anand et al. (2019). Based on this comparison, the set of 21 stimuli was considered to be wide enough to be compared with voices with the least and most possible strain percept. Figure 3 illustrates the magnitude spectra of the two comparison stimuli with the lowest (0.37 acum) and highest (2.12 acum) computed sharpness. The spectral shape of the comparison stimulus with the highest computed sharpness (right panel in Figure 3) would be rarely observed in natural speech, but the higher ends of the comparison stimulus were included to ensure matching of the extremely strained voices.

Test Stimuli

Ten samples of /ɑ/ phonations were selected from the University of Florida Disordered Voice Database and The University of Sydney Database. Sample selection was based on an interactive approach. From among a large set of possible samples with strained voice quality, candidate samples were chosen to have a wide range of strain, ranging from least strain to most strain, and to have as little

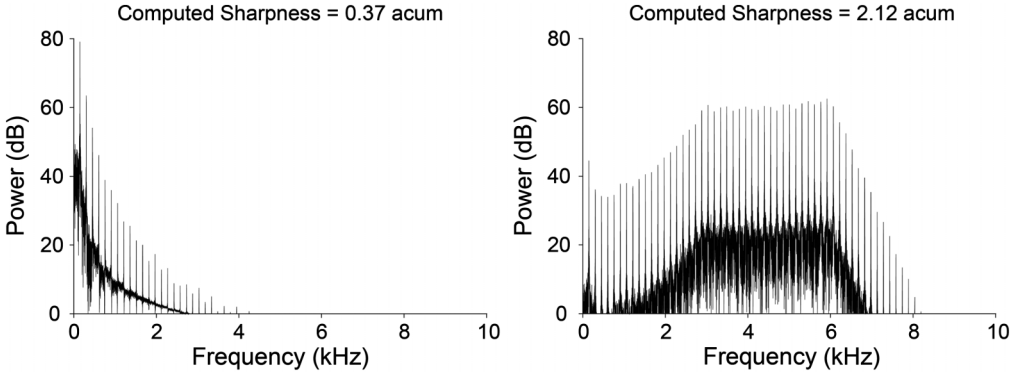
Figure 2. Left panel: Scatter plot of computed sharpness on the ordinate in acum units versus filter gain magnitude on the abscissa. Closely spaced gray symbols represent all stimuli created, with filter gain magnitude ranging from 0 to 99. Red markers represent the 21 filter gain magnitudes chosen as the independent variable, selected because they vary in sharpness systematically from low to high. Right panel: Scatter plot of computed sharpness over the 10 natural voice stimuli on the abscissa. Natural voice stimuli are ordered from lowest to highest computed sharpness.



breathy or rough voice quality as possible, using the stratified random sampling method described by Shrivastav (2006). This criterion was chosen to increase the probability that perceptual judgments could be limited to the strain percept. Stimulus selection was carried out by the first, second, and fifth authors. Candidate samples were then analyzed in terms of computed sharpness using the model described in Equation 1. The goal was for the chosen samples to have computed sharpness that spans a relatively large range of computed sharpness values and for those sharpness values to be spread throughout the range in a somewhat uniform manner. As shown in the right panel of Figure 2, the final 10 stimuli chosen for the study ranged in computed sharpness from 0.68 to 1.44 acum, as computed from Equation 1. The range of computed sharpness from natural voice stimuli with extensive vocal strain is much smaller than the range of computed sharpness from the comparison stimuli (see Figure 2), indicating that these newly developed comparison stimuli are sufficient to evaluate extremely strained voices.

All samples were cropped to 500 ms in duration and were down-sampled to 24414 Hz to match the available sampling rate of the Tucker-Davis Technologies hardware used to deliver stimuli to the listeners. Stimuli were shaped with cosine-squared onset and offset ramps 10 ms in duration to minimize audible clicks associated with the artificially imposed onset and offset amplitude changes. To control for loudness, the intensity of each stimulus was normalized to a single phon level computed using the loudness model of Moore et al. (1997). Eight of the samples were disordered voices, including bilateral nodules, glottis insufficiency, abductor vocal paralysis, and adductor laryngeal dystonia. There were two samples of healthy voices from The University of Sydney. The speakers of the two samples were instructed to constrict their false vocal folds and thicken their true vocal folds in habitual

Figure 3. Magnitude spectra of the two comparison stimuli with the lowest (left) and highest (right) computed sharpness.



larynx positions to ultimately increase their vocal effort (Madill & Nguyen, 2020).

Participants (Listeners)

Participants (nine women, one man) were fluent speakers of American English, had normal hearing in both ears (pure-tone thresholds < 25 dB HL at audiometric frequencies from 250 to 8000 Hz; American National Standards Institute, 2010), and ranged in age from 20 to 24 years (M = 21.8). Participants had no previous training in voice quality evaluation. Each participant provided informed consent in accordance with the procedures approved by the University of South Florida Institutional Review Board (IRB Pro0012381).

Familiarization Using a Visual Sort and Rank Task

Prior to familiarization with ME and the SVMT, each participant watched a short slide presentation designed to familiarize them with the terms used in the experiment. After the presentation, participants completed a visual sort and rank (VSR) task with the set of 10 natural voice samples. The VSR task ensured that participants would be familiar with a wide range of natural samples with strained voice quality. The VSR task was completed using a custom-designed graphical user interface (GUI) developed in MATLAB (The MathWorks, Inc.). Stimuli were presented via Etymotic ER-2 insert earphones, and participants were seated in a sound-attenuating chamber in front of a video monitor that displayed the VSR interface. On the right side of the interface were buttons, each representing a different strained voice sample. Participants were instructed to select one sound button at a time using the mouse. A mouse click on a button triggered the presentation of the associated sound file to the participants via the earphones. After listening to the sound, the participants were instructed to rank-order the sounds in terms of least to most strain by dragging the buttons associated with each sound to the left side of the interface and sorting them by strain magnitude.

ME Task

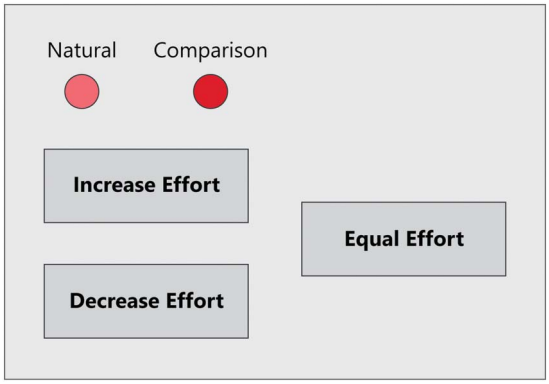
The ME experiment was completed with two sets of stimuli: the same 10 natural voice stimuli described above and the set of 21 comparison stimuli described above in the context of the SVMT. On each trial of the ME task, a single stimulus was presented, and participants were instructed to estimate the magnitude of perceived vocal strain on a ratio scale from 1 to 1,000 on a ratio-level scale, with 1 being the least amount of strain and 1,000 being the most amount of strain. They were instructed that if a sample was perceived to have twice as much strained quality as another sample, the assigned magnitude should be doubled. The GUI for this task included an edit box for typing in the numeric value using a computer keyboard. The test stimuli were presented 10 times in a block in random order, and the three blocks were performed, for a total of 30 presentations per stimulus. Each block took about 10–15 min to complete, and participants took a short break between blocks.

Prior to the ME experiment, participants were familiarized with the ME task by evaluating the perceived loudness of a set of nine-tone (1000 Hz) stimuli that varied in level from 60 to 92 dB. First, the stimuli were presented as an ascending level series. Next, the participants completed the ME procedure in which they judged loudness on a scale from 1 to 1,000, where 1 corresponds to the low end of the loudness scale (soft) and 1,000 corresponds to the high end of the scale (loud). Participants were also asked to estimate loudness on a ratio scale; if a sample sounded twice as loud as another sample, they were asked to assign double the magnitude to that sound. Participants completed the loudness estimation task with 10 repetitions of each sound. This task was repeated 3 times to familiarize them with the magnitude assessment task, and participants received feedback on how they used the scale each time. After the loudness ME, familiarization included the VSR task described above with the set of 10 natural voices. The session concluded with the full ME.

SVMT

This task follows the same procedures used in the investigation of the SVMT for perceived breathiness (e.g., Patel et al., 2012a) and perceived roughness (e.g., Eddins et al., 2015; Patel et al., 2012b). At the start of a trial, the natural stimulus, 500 ms in duration, is presented, followed by 500 ms of silence and then by the comparison sound, which is also 500 ms in duration. The subject interface for the SVMT (see Figure 4) consists of a GUI with two labels at the top, namely, "Natural" and "Comparison." Those correspond to buttons just below the labels, and those buttons turn red when either the natural voice sample is presented or the comparison sound is presented. There are three response buttons, labeled "Increase Effort," "Decrease Effort," and "Equal Effort." On the current trial, if the comparison sound is perceived as being produced with less vocal effort (lower vocal strain) than the natural sound, then "Increase Effort" would be selected. That results in an increase in the independent variable value for the next trial, resulting in a comparison sound with greater sharpness. On the contrary, if they perceived the comparison sound as being produced with more vocal effort (higher vocal strain) than the natural sound, then "Decrease Effort" would be selected, which results in

Figure 4. Graphical user interface used in the single-variable matching task.



a decrease in the independent variable value. When the two sounds are judged to be perceptually consistent in perceived vocal effort (vocal strain), the "Equal Effort" button would be selected. Participants were instructed on the use of the SVMT GUI as described above. Participants were also instructed to ignore other voice qualities (i.e., roughness and breathiness) as well as pitch, loudness, and vowel types.

Individual stimulus trials were completed in blocks of trials that began with a very high initial independent variable value (descending blocks) or a very low independent variable value (ascending blocks). In a descending block, the high independent variable value would lead to a "Decrease Effort" response and a reduced independent variable value on the next trial in the block. The independent variable value was adjusted down and up until a perceptual match was obtained. The opposite was true for ascending blocks of trials.

Before the actual experiment, participants were familiarized with the SVMT using the comparison sounds described above and two unique natural voice samples. One voice sample was selected to have high perceived strain, and the other was selected to have low perceived strain. Feedback was provided if the participants did not appropriately match the comparison stimulus to the training stimulus. The task was completed for both stimuli prior to the start of the main experiment.

For the main experimental task, the 10 natural voice samples described above were used. Matching for each stimulus included initial comparisons with low or high independent variable values, such that there were 20 conditions from which to select. The 20 conditions were pseudorandomly divided among participants into five blocks of four sets, and each block contained four different stimuli and equal numbers of initial conditions (e.g., two sets of high initial independent variable values and two sets of low initial independent variable values). Each block took approximately 15–20 min to complete. After one block was finished, a short break was given to the participants before starting with the next block. Participants completed five blocks in one to two scheduled test sessions (on different days). If participants did not complete all five blocks in the first session, then they were completed in the second session. No session exceeded 2 hr in duration. Each session began with the VSR task, followed by SVMT practice and then by SVMT testing. Participant activities across sessions are summarized in Table 1.

Perceptual Data Analysis

From strain ME tasks, strain magnitudes were obtained from 30 repetitions of each stimulus and were averaged within a listener. The strain magnitudes among 10 listeners were averaged to represent the mean strain magnitude of the stimulus for both test and comparison stimuli. Strain matching values correspond to the independent variable in the matching task, which is the numeric computation of sharpness (acum) of the comparison stimuli to which listeners matched the perceived strain of the natural stimuli. A final strain matching value was based on 10 repetitions of each stimulus, including both low and high initial sharpness conditions, which were averaged within a listener. The strain matching values among 10 listeners were averaged to represent the mean strain matching value for each natural voice stimulus.

Statistical Analysis

Statistical analyses were performed in SPSS (Version 27.0; IBM Corp.). Intralistener reliability was calculated

Table 1. Summary of listening tasks by participant visit.

Task	Visit 1	Visit 2	Visit 3	Visit 4
Task 1	VSR: 10 test stimuli	VSR: 10 test stimuli	VSR: 10 test stimuli	VSR: 10 test stimuli
Task 2	SVMT practice: 2 voice samples	SVMT practice: 2 voice samples	Loudness estimation: 9 tones	Loudness estimation: 9 tones
Task 3	SVMT: 10 test stimuli	SVMT: remainders of 10 test stimuli	Magnitude estimation: 10 test stimuli	Magnitude estimation: 21 comparison stimuli

Note. VSR = visual sort and rank task; SVMT = single-variable matching task.

via a two-way, mixed-effects intraclass correlation coefficient (ICC) for absolute agreement and averaged measures. Interlistener reliability was calculated via a two-way, mixed-effects ICC for consistency and averaged measures (ICC[2, k]). Simple linear regressions were performed to test if mean strain magnitudes were predicted by sharpness computed from the comparison stimuli and if the strain matching values of the 10 natural voice stimuli were predicted by computed sharpness and mean perceived strain magnitudes of the voice stimuli. To account for multiple (four) tests, p values were adjusted using the Bonferroni correction (p = .05/4 = .015).

Results

ME

The comparison sounds developed for the SVMT experiment reported below were selected to vary systematically, from low to high model-computed sharpness values, consistent with the hypothesis that perceived sharpness will map onto the strained voice quality. Informal listening by the investigative team indicated that those stimuli varied substantially in a manner analogous to perceived vocal effort or strain. To formally evaluate the perception of strain associated with those stimuli, ME was completed for all 21 stimuli in the set by each of the 10 participants. The intrarater reliability (ICC[2, k], absolute agreement) for the 10 listeners ranged from .98 to .99.

Figure 5 shows the mean perceived strain magnitude data plotted for all 21 stimuli with computed sharpness on the abscissa. As hypothesized, perceived strain magnitudes increased linearly as the computed sharpness of the comparison stimuli (the basis for stimulus selection) increased. Moreover, all stimuli at the higher end of the computed sharpness continuum had greater perceived strain than those at the lower end of the computed sharpness continuum. In Figure 5, prediction of the strain MEs (ordinate) by the sharpness model (abscissa) using linear regression resulted in r^2 = .99, $F(1, 19)$ = 1,229.3, p < .001.

Computation of sharpness using the Fastl and Zwicker (2007) model confirmed that the set of 10 natural stimuli selected for study varied over a wide range of computed sharpness values (i.e., the right panel in Figure 2). To establish the perceptual relevance of that relationship, the perceived strain magnitude of the natural voices was measured using the ME procedure. The intrarater reliability (ICC[2, k], absolute agreement) for the 10 listeners was .99. The interrater reliability (ICC[2, k], consistency) among the 10 listeners was .95. Figure 6 shows the perceived strain magnitude on the ordinate, with computed sharpness on the abscissa. Stimulus numbers inside the symbols are assigned in order from lowest to highest computed sharpness (i.e., the right panel in Figure 3). The perceived magnitudes indicate a wide range of strain magnitudes associated with the set of natural voice samples, and the hierarchical order of strain magnitudes closely matches the hierarchical order of computed sharpness, except for Stimuli 2, 4, and 6. A linear regression indicated that computed sharpness values were a significant predictor of strain magnitude, with r^2 = .93, $F(1, 8)$ = 110.6, p < .001.

SVMT

The same 10 participants completed the SVMT for strain using the novel comparison sound. For the strain matching task, the intrarater reliability (ICC[2, k], absolute agreement) of the 10 listeners was .99. The interrater reliability (ICC[2, k], consistency) among the 10 listeners was .98. Thus, the SVMT resulted in very high reliability, consistent with the ME task above and prior use of the

Figure 5. Mean strain magnitudes of the comparison stimuli in a continuum of computed sharpness from lowest to highest. Error bars indicate 95% confidence intervals.

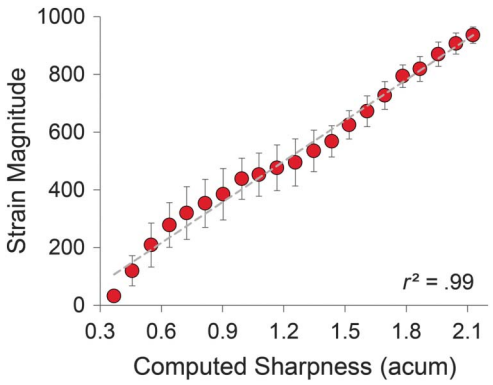
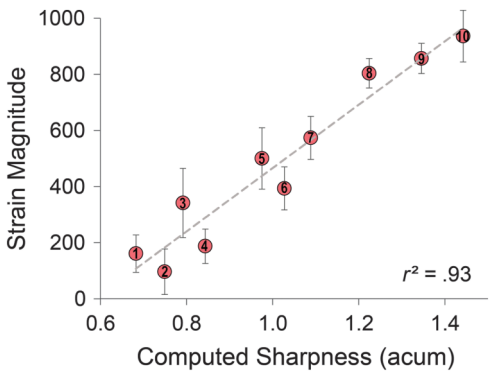


Figure 6. Mean strain magnitudes of the natural voice stimuli over computed sharpness on the abscissa. The numbers inside the symbols indicate the stimulus order from lowest to highest computed sharpness. Error bars indicate 95% confidence intervals.

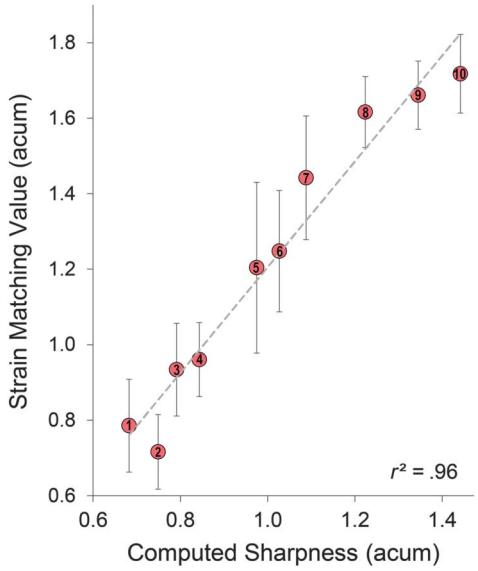


SVMT to quantify perceived breathiness (Patel et al., 2012a) and roughness (Patel et al., 2012b).

Perceived vocal strain estimated from the SVMT is shown in Figure 7 with strain matching values on the ordinate, quantified in acum units, and computed sharpness on the abscissa. Symbols reflect mean matching values for the 10 listeners, with error bars indicating 95% confidence intervals. The strain matching values for the natural stimuli ranged from 0.76 to 1.72 acum. The perceptual data in Figure 7 are similar in form to the computed sharpness data in the right panel of Figure 2. The hierarchical order of strain matching values (see Figure 7) and the sharpness computed for each of the natural stimuli corresponded exactly to the eight stimuli with the highest matching values, whereas the matching value for the stimulus with the lowest computed sharpness (Stimulus 1) was slightly greater than the matching value for the stimulus with the second-lowest computed sharpness (Stimulus 2). The overall range of computed sharpness values was slightly compressed in the model data (0.68–1.44 acum) relative to the perceptual matching judgments (0.76–1.72 acum). As hypothesized, the mean strain matching values increased linearly with computed sharpness values. Taking computed sharpness as the reference data set, the validity of the strain matching values for the natural voice stimuli can be evaluated by computing a simple linear regression to test whether the model calculations of sharpness were a significant predictor of strain matching values. Shown as the dashed line in Figure 7, the linear regression model indicates that the model calculations of sharpness were indeed a significant predictor of the perceptual strain matching values, $r^2 = .96$, $F(1, 8) = 168.5$, $p < .001$.

The perception of vocal strain measured using the ME procedure can be compared with perception measured with the SVMT procedure. Figure 8 shows these data in the form of a scatter plot, with strain matching values in acum units on the ordinate and strain ME (unitless) on the abscissa. Each symbol represents the mean strain matching values and magnitude estimate for one of the natural voice samples in the set of 10. Only the matching values for two stimuli (Stimuli 4 and 6) deviated from the hierarchical order of strain magnitude estimates. There was a strong and significant linear relation between the data sets, $r^2 = .95$, $F(1, 8) = 154.9$, $p < .001$.

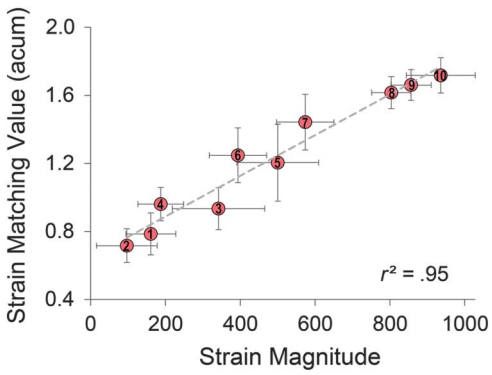
Figure 7. Mean strain matching values of the natural voice stimuli on the ordinate over model estimates of sharpness on the abscissa. The numbers inside the symbols indicate the stimulus order from lowest to highest computed sharpness. Error bars indicate 95% confidence intervals.



Discussion

The goal of this study was to evaluate the utility of novel comparison sounds for use in an SVMT to measure the perception of vocal strain. The development of the comparison sounds reported here was guided by an overarching psychoacoustic framework (Fastl & Zwicker, 2007) within which tonality, related to periodicity, is associated with vocal breathiness (Eddins et al., 2016; Patel et al., 2012a); roughness, related to amplitude fluctuation in the temporal envelope, is associated with vocal roughness (Eddins et al., 2015; Park, Anand, Ozmeral, et al., 2022; Patel et al., 2012b); and sharpness, related to the shape of the spectral envelope, is associated with vocal strain (Anand et al., 2019; Kopf et al., 2013). Accordingly, the same base stimulus used as the comparison sound in SVMTs, which successfully quantifies the perception of breathiness (Patel et al., 2012a) and roughness (Patel et al.,

Figure 8. Mean strain matching values of the natural voice stimuli on the ordinate over mean perceived strain magnitude on the abscissa. The numbers inside the symbols indicate the stimulus order from lowest to highest computed sharpness. Error bars indicate 95% confidence intervals.



2012b), was used in the comparison stimulus for matching perceived strain. The spectral envelope was modified systematically by varying the gain of a bandpass filter from low to high. The independent variable in the matching task was the filter gain value, which changed the sharpness of the comparison stimuli computed using the model of Fastl and Zwicker (2007), as expressed in Equation 1.

To validate the comparison sound, the ME of the perceived strain of the comparison stimuli was compared with sharpness calculated from the model. The resulting data supported the hypotheses that measures of the perception of sharpness are reliable within and between listeners and are related to the perceived strain of the comparison stimuli. The hypothesis that the perceived strain of natural voice samples is also strongly correlated with sharpness computed from the natural voice samples was confirmed by computing the sharpness for each voice in the set of natural voice stimuli and then using ME to index the perceived strain for the same stimuli.

Subsequently, the newly developed synthetic comparison sounds, grounded in physical values associated with filter gain and scaled in terms of equal steps of computed sharpness, were used in an SVMT to quantify perceived strain in relation to those physical values. The results from the SVMT supported the hypothesis that computed sharpness and perceived strain magnitudes of the natural stimuli would be strongly correlated with the strain matching values. A simple linear regression indicated that sharpness computations are a significant predictor of perceived strain matching values, resulting in a highly significant r^2 of .96. As the perceived strain of the voice stimuli increased from low to high matching values, computed sharpness increased linearly, indicating that listeners were able to match perceived strain in the natural voice samples with the newly developed synthetic comparison stimuli with great precision. Furthermore, sharpness computed from the comparison sounds (0.68–1.44 acum) resulted in a similar range of computed sharpness based on the strain matching values (0.76–1.72 acum). These results indicate that listeners associate perceived strain in voice to the sharpness of the comparison stimuli. Moreover, the results demonstrated that the strain matching task can be used to measure a wide range of perceived strain in natural voices. With this result, an accurate and reliable SVMT has been added to a battery of voice quality matching techniques.

Prior measures of perceived strain have been shown to be the least reliable among measures of the different voice qualities (Webb et al., 2004; Zraick et al., 2011). With this novel SVMT, the accuracy and reliability of the measurement of perceived strain have been improved dramatically. Indeed, the reliability of the SVMT for strain observed in this study was as high as the reliability of the SVMT for breathiness (Eddins et al., 2016; Patel et al., 2012a) and roughness (Eddins et al., 2015; Patel et al., 2012b). Furthermore, using the comparison stimuli as a reference sound during the perceptual task can reduce the individual and context biases and can improve reliability over other methods, such as rating scales and visual analog scales (Chan & Yiu, 2002; dos Santos et al., 2019; Kapsner-Smith et al., 2021; Kreiman & Gerratt, 2011).

Although the ME data also demonstrate a strong and consistent relationship between perceived strain magnitudes and computed sharpness, the unitless data from the ME task make it difficult to relate changes in the strain percept between stimuli, between listeners, and over time. It is anticipated that the physical units and measurement precision associated with the SVMT for strain will overcome these limitations.

The Relationship Between Perception of Strain and Sharpness

Anand et al. (2019) demonstrated that perceptual ratings of strain in voice samples are strongly correlated with computations of sharpness, along with other spectral moment measures derived from the voice stimuli. This study extends that work using both the ME task and the SVMT. Taken together, these data are consistent with the previously observed relationship between the low-to-high (LH) ratio, computed as the ratio of power over a range of low audio frequencies to high audio frequencies, and perceived strain in disordered voices and normophonic individuals who intentionally modulated their vocal effort (Lowell, Kelley, Awan, et al., 2012; McKenna & Stepp, 2018). Conceptually, the LH ratio and the computation of spectral sharpness are similar. They differ in that the LH ratio is based solely on the acoustic stimulus calculated directly from the Fourier transform of the stimulus waveform, whereas the sharpness model is based on known transformations of the acoustic stimulus by the auditory periphery (Fastl & Zwicker, 2007). For this reason, computed sharpness may better represent the perception of strain than the LH ratio.

The increase in power at high frequencies in strained voices, relative to normophonic voices, could be related to increased vocal fold adduction and laryngeal adductor activation during voice production associated with increased vocal effort. Increased adduction is known to result in decreased spectral tilt, which corresponds to increased power at high relative to low audio frequencies (Klatt & Klatt, 1990). With greater adduction, thickness, and stiffness of the vocal folds from increased laryngeal adductor activation, the closing velocity of vocal fold vibration is faster (Stepp et al., 2010; Zhang, 2016a), resulting in an increase in power at high frequencies (Fant et al., 1985; Zhang, 2016b). The faster closing velocity is related to a higher maximum flow declination rate (MFDR) during

the vibratory cycle, and a high MFDR has been observed in individuals with voice disorders associated with increased vocal effort such as vocal hyperfunction (Espinoza et al., 2017; Hillman et al., 1989).

Similarly, a second physiological basis for reduced spectral tilt, decreased LH ratio, and increased sharpness associated with strained voices may be increased compression of the supraglottis and the vocal tract. Individuals with vocal hyperfunction often present with increased laryngeal height (Lowell, Kelley, Colton, et al., 2012; Roy & Ferguson, 2001) and compressed supraglottal cavity (Stager et al., 2000). Individuals with healthy voices have also shown increased compression of the supraglottal cavity when they voluntarily increased vocal effort (Madill & Nguyen, 2020; McKenna et al., 2019). Compression of the vocal tract can result in increases in resonant frequencies and resulting formants due to decreases in the size of the vocal tract (resonating chamber). Because of the possible relationship between the physiology of vocal effort and sharpness, the perceived strain obtained in the newly developed SVMT and the calculated sharpness measured from speech samples may be possible indicators of the speaker's vocal effort during voice production.

Limitations

The listeners in this study were inexperienced with no prior training in voice quality perception, whereas in clinical voice evaluation, the perceptual assessment is performed by a trained clinician. It is possible that the perceptual strategies used by experts and inexperienced listeners to evaluate voice quality may differ. However, there is evidence that ratings of voice quality by expert listeners may even demonstrate greater variability than ratings by inexperienced listeners (e.g., Kreiman et al., 1990). Moreover, overcoming differences between classes of listeners is the ultimate goal of developing a matching task. The use of an explicit comparison sound rather than reliance on an internal standard or a domain-specific template is designed to reduce the influence of the internal standard, regardless of prior listener experience.

The natural voice stimuli chosen for this investigation were selected based on their perceived strain and general lack of perceived breathiness and roughness. The rationale for selecting stimuli with a primary voice quality of strain and minimal presence of other voice qualities was to minimize the potential influence of those other voice qualities in this initial investigation of strain using the SVMT. Although the goals of the study were achieved, having demonstrated strong relationships between perceptual data and computational estimates of spectral sharpness, the extent to which the presence of covarying voice qualities may have influenced strain matching values remains unclear. A recent study examining the interaction between systematically altered perceived breathiness and roughness in a series of stimuli showed that matching values of one quality were not significantly affected by the other (Park, Anand, Kopf, et al., 2022). However, it is not yet guaranteed that strain matching values will also behave independently like breathiness and roughness matching values until future studies investigate the covariance of all three voice qualities using a matching task.

Clinical Implications and Future Directions

It is important to note that the SVMT method used in this study has been developed for laboratory measurements of voice quality and requires several minutes to achieve a perceptual match for a single target voice. The assessment of voice quality in clinics typically requires a real-time or rapid assessment that can aid in the diagnosis and treatment of voice disorders. For this psychoacoustic framework to be applied successfully in clinical settings, the methods will need to be adapted to be successful over a much shorter time. In addition, individuals with voice disorders who present with strained voice quality frequently have elevated breathiness, roughness, or both, due to factors such as glottal insufficiency and organic lesions (Holmberg et al., 2001; Lowell, Kelley, Awan, et al., 2012). Because other voice qualities of people with voice disorders may affect strain evaluation by the matching task, future studies may combine three SVMT measures, targeting separately the breathy, rough, and strained percepts, into one three-dimensional matching task. This would allow the evaluation of possible interactions between voice quality dimensions as well as comparisons between the results from SVMTs and simultaneous judgments of all three dimensions.

Conclusions

This investigation sought to establish a suitable comparison stimulus for use in a matching task to quantify the perception of vocal strain in stimuli that varied widely in their perceived strain. In doing so, the overarching goal was to develop a laboratory procedure that would provide accurate and reliable estimates of strain for a wide variety of voice samples. A psychoacoustic framework was used to guide the design of the comparison stimulus, and that stimulus was based on the base stimulus used previously for matching the perceived breathiness and roughness of voice samples. Two different psychophysical methods were used to evaluate the natural stimulus set, namely, ME and an SVMT. The results indicate that matching the perceived strain of natural voices with a synthetic comparison sound, consisting of a sawtooth waveform mixed with noise and bandpass-filtered with

variable filter gain, produces highly accurate matching values with strong intra- and interrater reliability. Furthermore, the psychoacoustic approach adopted a well-established model of spectral sharpness to quantify both the natural stimuli and the comparison sound. Model predictions of sharpness perception indicate that the percept of strain in natural voices is strongly related to the percept of sharpness, a concept that was developed entirely on the basis of synthetic sounds (e.g., Fastl & Zwicker, 2007).

Data Availability Statement

The published data are available from the corresponding author upon reasonable request.

Acknowledgments

This work was supported by National Institute on Deafness and Other Communication Disorders Grant R01 DC009029, awarded to David A. Eddins and Rahul Shrivastav. The authors would like to thank Mark D. Skowronski for the help with developing comparison stimuli.

References

American National Standards Institute. (2010). *Methods for manual pure-tone threshold audiometry* (ANSI S3.21-2010).

Anand, S., Kopf, L. M., Shrivastav, R., & Eddins, D. A. (2019). Objective indices of perceived vocal strain. *Journal of Voice, 33*(6), 838–845. https://doi.org/10.1016/j.jvoice.2018.06.005

Carding, P. N., Wilson, J. A., MacKenzie, K., & Deary, I. J. (2009). Measuring voice outcomes: State of the science review. *The Journal of Laryngology & Otology, 123*(8), 823–829. https://doi.org/10.1017/S0022215109005398

Chan, K. M. K., & Yiu, E. M.-L. (2002). The effect of anchors and training on the reliability of perceptual voice evaluation. *Journal of Speech, Language, and Hearing Research, 45*(1), 111–126. https://doi.org/10.1044/1092-4388(2002/009)

Dahl, K. L., Weerathunge, H. R., Buckley, D. P., Dolling, A. S., Diaz-Cadiz, M., Tracy, L. F., & Stepp, C. E. (2021). Reliability and accuracy of expert auditory-perceptual evaluation of voice via telepractice platforms. *American Journal of Speech-Language Pathology, 30*(6), 2446–2455. https://doi.org/10.1044/2021_AJSLP-21-00091

dos Santos, P. C. M., Vieira, M. N., Sansão, J. P. H., & Gama, A. C. C. (2019). Effect of auditory-perceptual training with natural voice anchors on vocal quality evaluation. *Journal of Voice, 33*(2), 220–225. https://doi.org/10.1016/j.jvoice.2017.10.020

Eddins, D. A., Anand, S., Camacho, A., & Shrivastav, R. (2016). Modeling of breathy voice quality using pitch-strength estimates. *Journal of Voice, 30*(6), 774.e1–774.e7. https://doi.org/10.1016/j.jvoice.2015.11.016

Eddins, D. A., Kopf, L. M., & Shrivastav, R. (2015). The psychophysics of roughness applied to dysphonic voice. *The Journal of the Acoustical Society of America, 138*(6), 3820–3825. https://doi.org/10.1121/1.4937753

Espinoza, V. M., Zañartu, M., Van Stan, J. H., Mehta, D. D., & Hillman, R. E. (2017). Glottal aerodynamic measures in women with phonotraumatic and nonphonotraumatic vocal hyperfunction. *Journal of Speech, Language, and Hearing Research, 60*(8), 2159–2169. https://doi.org/10.1044/2017_JSLHR-S-16-0337

Fant, G., Liljencrants, J., & Lin, Q. (1985). A four-parameter model of glottal flow. *STL-QPSR, 26*(4), 001–013.

Fastl, H., & Zwicker, E. (2007). *Psychoacoustics: Facts and models* (3rd ed.). Springer. https://doi.org/10.1007/978-3-540-68888-4

Gerratt, B. R., Kreiman, J., Antonanzas-Barroso, N., & Berke, G. S. (1993). Comparing internal and external standards in voice quality judgments. *Journal of Speech and Hearing Research, 36*(1), 14–20. https://doi.org/10.1044/jshr.3601.14

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*(2), 373–392. https://doi.org/10.1044/jshr.3202.373

Hillman, R. E., Stepp, C. E., Van Stan, J. H., Zañartu, M., & Mehta, D. D. (2020). An updated theoretical framework for vocal hyperfunction. *American Journal of Speech-Language Pathology, 29*(4), 2254–2260. https://doi.org/10.1044/2020_AJSLP-20-00104

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*(3), 395–412. https://doi.org/10.1016/S0892-1997(01)00041-8

Kapsner-Smith, M. R., Opuszynski, A., Stepp, C. E., & Eadie, T. L. (2021). The effect of visual sort and rate versus visual analog scales on the reliability of judgments of dysphonia. *Journal of Speech, Language, and Hearing Research, 64*(5), 1571–1580. https://doi.org/10.1044/2021_JSLHR-20-00623

Kelchner, L. N., Brehm, S. B., Weinrich, B., Middendorf, J., deAlarcon, A., Levin, L., & Elluru, R. (2010). Perceptual evaluation of severe pediatric voice disorders: Rater reliability using the Consensus Auditory-Perceptual Evaluation of Voice. *Journal of Voice, 24*(4), 441–449. https://doi.org/10.1016/j.jvoice.2008.09.004

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*(2), 124–132. https://doi.org/10.1044/1058-0360(2008/08-0017)

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustical Society of America, 87*(2), 820–857. https://doi.org/10.1121/1.398894

Kopf, L. M., Shrivastav, R., & Eddins, D. A. (2013). *Isolating the effects of strain on voice quality* [Poster presentation]. Pan-European Voice Conference, Prague, Czech Republic.

Kreiman, J., & Gerratt, B. R. (1996). The perceptual structure of pathologic voice quality. *The Journal of the Acoustical Society of America, 100*(3), 1787–1795. https://doi.org/10.1121/1.416074

Kreiman, J., & Gerratt, B. R. (1998). Validity of rating scale measures of voice quality. *The Journal of the Acoustical Society of America, 104*(3), 1598–1608. https://doi.org/10.1121/1.424372

Kreiman, J., & Gerratt, B. R. (2005). Perception of aperiodicity in pathological voice. *The Journal of the Acoustical Society of America, 117*(4), 2201–2211. https://doi.org/10.1121/1.1858351

Kreiman, J., & Gerratt, B. R. (2011). Comparing two methods for reducing variability in voice quality measurements. *Journal of Speech, Language, and Hearing Research, 54*(3), 803–812. https://doi.org/10.1044/1092-4388(2010/10-0083)

Kreiman, J., Gerratt, B. R., & Ito, M. (2007). When and why listeners disagree in voice quality assessment tasks. *The Journal of the Acoustical Society of America, 122*(4), 2354–2364. https://doi.org/10.1121/1.2770547

Kreiman, J., Gerratt, B. R., & Precoda, K. (1990). Listener experience and perception of voice quality. *Journal of Speech and Hearing Research, 33*(1), 103–115. https://doi.org/10.1044/jshr.3301.103

Kreiman, J., Gerratt, B. R., Precoda, K., & Berke, G. S. (1992). Individual differences in voice quality perception. *Journal of Speech and Hearing Research, 35*(3), 512–520. https://doi.org/10.1044/jshr.3503.512

Lowell, S. Y., Kelley, R. T., Awan, S. N., Colton, R. H., & Chan, N. H. (2012). Spectral- and cepstral-based acoustic features of dysphonic, strained voice quality. *Annals of Otology, Rhinology & Laryngology, 121*(8), 539–548. https://doi.org/10.1177/000348941212100808

Lowell, S. Y., Kelley, R. T., Colton, R. H., Smith, P. B., & Portnoy, J. E. (2012). Position of the hyoid and larynx in people with muscle tension dysphonia. *The Laryngoscope, 122*(2), 370–377. https://doi.org/10.1002/lary.22482

Madill, C., & Nguyen, D. D. (2020). Impact of instructed laryngeal manipulation on acoustic measures of voice—Preliminary results. *Journal of Voice.* Advance online publication. https://doi.org/10.1016/j.jvoice.2020.11.004

McKenna, V. S., Diaz-Cadiz, M. E., Shembel, A. C., Enos, N. M., & Stepp, C. E. (2019). The relationship between physiological mechanisms and the self-perception of vocal effort. *Journal of Speech, Language, and Hearing Research, 62*(4), 815–834. https://doi.org/10.1044/2018_JSLHR-S-18-0205

McKenna, V. S., Heller Murray, E. S., Lien, Y.-A. S., & Stepp, C. E. (2016). The relationship between relative fundamental frequency and a kinematic estimate of laryngeal stiffness in healthy adults. *Journal of Speech, Language, and Hearing Research, 59*(6), 1283–1294. https://doi.org/10.1044/2016_JSLHR-S-15-0406

McKenna, V. S., & Stepp, C. E. (2018). The relationship between acoustical and perceptual measures of vocal effort. *The Journal of the Acoustical Society of America, 144*(3), 1643–1658. https://doi.org/10.1121/1.5055234

Moore, B. C. J., & Glasberg, B. R. (1983). Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *The Journal of the Acoustical Society of America, 74*(3), 750–753. https://doi.org/10.1121/1.389861

Moore, B. C. J., Glasberg, B. R., & Baer, T. (1997). A model for the prediction of thresholds, loudness, and partial loudness. *Journal of the Audio Engineering Society, 45*(4), 224–240. http://www.aes.org/e-lib/browse.cfm?elib=10272

Morrison, M. (1997). Pattern recognition in muscle misuse voice disorders: How I do it. *Journal of Voice, 11*(1), 108–114. https://doi.org/10.1016/s0892-1997(97)80031-8

Oates, J. (2009). Auditory-perceptual evaluation of disordered voice quality: Pros, cons and future directions. *Folia Phoniatrica et Logopaedica, 61*(1), 49–56. https://doi.org/10.1159/000200768

Park, Y., Anand, S., Kopf, L. M., Shrivastav, R., & Eddins, D. A. (2022). Interactions between breathy and rough voice qualities and their contributions to overall dysphonia severity. *Journal of Speech, Language, and Hearing Research, 65*(11), 4071–4084. https://doi.org/10.1044/2022_JSLHR-22-00012

Park, Y., Anand, S., Ozmeral, E. J., Shrivastav, R., & Eddins, D. A. (2022). Predicting perceived vocal roughness using a bio-inspired computational model of auditory temporal envelope processing. *Journal of Speech, Language, and Hearing Research, 65*(8), 2748–2758. https://doi.org/10.1044/2022_JSLHR-22-00101

Patel, S., Shrivastav, R., & Eddins, D. A. (2010). Perceptual distances of breathy voice quality: A comparison of psychophysical methods. *Journal of Voice, 24*(2), 168–177. https://doi.org/10.1016/j.jvoice.2008.08.002

Patel, S., Shrivastav, R., & Eddins, D. A. (2012a). Developing a single comparison stimulus for matching breathy voice quality. *Journal of Speech, Language, and Hearing Research, 55*(2), 639–647. https://doi.org/10.1044/1092-4388(2011/10-0337)

Patel, S., Shrivastav, R., & Eddins, D. A. (2012b). Identifying a comparison for matching rough voice quality. *Journal of Speech, Language, and Hearing Research, 55*(5), 1407–1422. https://doi.org/10.1044/1092-4388(2012/11-0160)

Ramig, L. O., & Verdolini, K. (1998). Treatment efficacy: Voice disorders. *Journal of Speech, Language, and Hearing Research, 41*(1), S101–S116. https://doi.org/10.1044/jslhr.4101.s101

Roy, N., & Ferguson, N. A. (2001). Formant frequency changes following manual circumlaryngeal therapy for functional dysphonia: Evidence of laryngeal lowering. *Journal of Medical Speech-Language Pathology, 9*(3), 169–176.

Shrivastav, R. (2003). The use of an auditory model in predicting perceptual ratings of breathy voice quality. *Journal of Voice, 17*(4), 502–512. https://doi.org/10.1067/s0892-1997(03)00077-8

Shrivastav, R. (2006). Multidimensional scaling of breathy voice quality: Individual differences in perception. *Journal of Voice, 20*(2), 211–222. https://doi.org/10.1016/j.jvoice.2005.04.005

Shrivastav, R., Sapienza, C. M., & Nandur, V. (2005). Application of psychometric theory to the measurement of voice quality using rating scales. *Journal of Speech, Language, and Hearing Research, 48*(2), 323–335. https://doi.org/10.1044/1092-4388(2005/022)

Stager, S. V., Bielamowicz, S. A., Regnell, J. R., Gupta, A., & Barkmeier, J. M. (2000). Supraglottic activity: Evidence of vocal hyperfunction or laryngeal articulation? *Journal of Speech, Language, and Hearing Research, 43*(1), 229–238. https://doi.org/10.1044/jslhr.4301.229

Stepp, C. E., Hillman, R. E., & Heaton, J. T. (2010). A virtual trajectory model predicts differences in vocal fold kinematics in individuals with vocal hyperfunction. *The Journal of the Acoustical Society of America, 127*(5), 3166–3176. https://doi.org/10.1121/1.3365257

Stevens, S. S., Volkmann, J., & Newman, E. B. (1937). A scale for the measurement of the psychological magnitude pitch. *The Journal of the Acoustical Society of America, 8*(3), 185–190. https://doi.org/10.1121/1.1915893

Webb, A. L., Carding, P. N., Deary, I. J., MacKenzie, K., Steen, N., & Wilson, J. A. (2004). The reliability of three perceptual evaluation scales for dysphonia. *European Archives of Oto-Rhino-Laryngology and Head & Neck, 261*(8), 429–434. https://doi.org/10.1007/s00405-003-0707-7

Zhang, Z. (2016a). Cause–effect relationship between vocal fold physiology and voice production in a three-dimensional

phonation model. *The Journal of the Acoustical Society of America, 139*(4), 1493–1507. https://doi.org/10.1121/1.4944754

Zhang, Z. (2016b). Mechanics of human voice production and control. *The Journal of the Acoustical Society of America, 140*(4), 2614–2635. https://doi.org/10.1121/1.4964509

Zraick, R. I., Kempster, G. B., Connor, N. P., Thibeault, S., Klaben, B. K., Bursac, Z., Thrush, C. R., & Glaze, L. E. (2011). Establishing validity of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V). *American Journal of Speech-Language Pathology, 20*(1), 14–22. https://doi.org/10.1044/1058-0360(2010/09-0105)

Zwicker, E. (1961). Subdivision of the audible frequency range into critical bands (Frequenzgruppen). *The Journal of the Acoustical Society of America, 33*(2), 248. https://doi.org/10.1121/1.1908630