Comparison of Pitch Strength With Perceptual and Other Acoustic Metric Outcome Measures Following Medialization Laryngoplasty

*,†Adam D. Rubin, *,‡Cristina Jackson-Menaldi, §Lisa M. Kopf, *Katherine Marks, *Jean Skeffington, ¶Mark D. Skowronski, **Rahul Shrivastav, and ††Eric J. Hunter, *St. Clair Shores, †Rochester, ‡Detroit, and ††East Lansing, Michigan, §Cedar Falls, Iowa, ¶Tampa, Florida, and **Athens, Georgia

Summary: Introduction. The diagnoses of voice disorders, as well as treatment outcomes, are often tracked using visual (eg, stroboscopic images), auditory (eg, perceptual ratings), objective (eg, from acoustic or aerodynamic signals), and patient report (eg, Voice Handicap Index and Voice-Related Quality of Life) measures. However, many of these measures are known to have low to moderate sensitivity and specificity for detecting changes in vocal characteristics, including vocal quality.

Objective. The objective of this study was to compare changes in estimated pitch strength (PS) with other conventionally used acoustic measures based on the cepstral peak prominence (smoothed cepstral peak prominence, cepstral spectral index of dysphonia, and acoustic voice quality index), and clinical judgments of voice quality (GRBAS [grade, roughness, breathiness, asthenia, strain] scale) following laryngeal framework surgery.

Methods. This study involved post hoc analysis of recordings from 22 patients pretreatment and post treatment (thyroplasty and behavioral therapy). Sustained vowels and connected speech were analyzed using objective measures (PS, smoothed cepstral peak prominence, cepstral spectral index of dysphonia, and acoustic voice quality index), and these results were compared with mean auditory-perceptual ratings by expert clinicians using the GRBAS scale.

Results. All four acoustic measures changed significantly in the direction that usually indicates improved voice quality following treatment ($P < 0.005$). Grade and breathiness correlated the strongest with the acoustic measures ($|r| \sim 0.7$) with strain being the least correlated.

Conclusions. Acoustic analysis on running speech highly correlates with judged ratings. PS is a robust, easily obtained acoustic measure of voice quality that could be useful in the clinical environment to follow treatment of voice disorders.

Key Words: Cepstral peak−Pitch strength−Thyroplasty−Objective voice measures−Acoustic analysis.

INTRODUCTION

In today's health-care environment, emphasis is placed on evidenced-based medicine and meaningful quality measures.[1,2] Dysphonia is a common indication for referral to an otolaryngologist and speech pathologist.[3] Evaluation of patients with voice disorders involves a combination of some or all of the following: visualization of the vocal tract and folds, perceptual analysis of the voice, quality of life measures, and acoustic and aerodynamic analyses of the voice signal.[4−6] Each of these evaluation techniques provides unique information that is valuable to diagnosis and assessment of treatment outcomes.

Visualization of the vocal folds with flexible laryngoscopy and videostroboscopy is critical for diagnosis.[7] However, interpretation and reporting of the various measures (eg, amplitude and mucosal wave) are subjective and may vary based on the experience of the clinician[8] and may be biased because of the clinician's intervention.[9,10] Although many diagnoses of voice disorders are made or confirmed through laryngeal imaging, measurement of voice quality adds additional information that can be particularly useful for comparing pre- and post-treatment results.

With dysphonia as a common concern among the population, developing objective measures that can reliably predict listener perception of the voice is important,[11] as clinician-perceived measures yield variable results, even between experienced clinicians.[12,13] Although some attempts have been made to reduce the variability across clinician ratings (eg, consensus training[14,15]), some degree of variance is likely to remain. Well-designed objective measures will better support voice care, professional recommendations, and reported patient outcomes.

Studies involving most traditional acoustic measures suggest inconsistency in their effectiveness to measure treatment outcomes.[16] This variation can be attributed to multiple factors, including the type of system performing the measurement and the acoustic environment in which the recordings are conducted.[17] Furthermore, the majority of these measures can only be used for sustained vowel assessment, which may not be

Accepted for publication March 27, 2018.

This work was partially supported by the National Institutes of Health (Grant Nos. R01DC004224 and R01DC012315) from the National Institute on Deafness and Other Communication Disorders.

Research presentation: This research was presented at the 2016 Voice Foundation Symposium in Philadelphia, PA.

From the *Lakeshore Ear, Nose, and Throat Center, Lakeshore Professional Voice Center, St. Clair Shores, Michigan; †Department of Surgery, Oakland University William Beaumont School of Medicine, Rochester, Michigan; ‡Department of Otolaryngology, School of Medicine, Wayne State University, Detroit, Michigan; §Department of Communication Sciences and Disorders, University of Northern Iowa, Cedar Falls, Iowa; ¶Department of Communication Sciences and Disorders, University of South Florida, Tampa, Florida; **Office of the Vice President for Instruction, University of Georgia, Athens, Georgia; and the ††Department of Communicative Sciences and Disorders, Michigan State University, East Lansing, Michigan.

Address correspondence and reprint requests to Adam D. Rubin, Lakeshore Ear, Nose & Throat Center, 21000 E. 12 Mile Rd. #111, St. Clair Shores, MI 48081.
E-mail: arubin@lakeshoreent.com

Journal of Voice, Vol. 33, No. 5, pp. 795−800
0892-1997

https://doi.org/10.1016/j.jvoice.2018.03.019

an accurate representation of changes in voice quality.[18] Moreover, most traditional measures rely on an accurate fundamental frequency estimation and, thus, are only appropriately used for some mildly dysphonic voices. In contrast, voices of patients with severe dysphonia often cannot be quantified with traditional approaches that rely on accurate estimation of the fundamental period.[19,20]

The inability to use many traditional measures beyond the most mildly dysphonic voices has led to a push to find other measures that can be used for more severe dysphonic voice. For example, Little et al. presented a stochastic self-similarity measure (detrended fluctuation analysis) on steady vowels that was more robust to noisy environments and was not dependent on fundamental frequency measures.[21] Another promising development is the relative fundamental frequency measure that has been shown to be sensitive to hyperfunction-related voice disorders[22]; however, estimating the relative fundamental frequency is laborious. Although these two examples seem to be effective in the contexts of application, they can only be applied to particular portions of a speech signal and are not suitable for application to running speech as a whole. In contrast, some measures can be applied to connected speech. These include three based on cepstral peak prominence: smoothed cepstral peak prominence (CPPS),[23] the acoustic voice quality index (AVQI),[24] and the cepstral spectral index of dysphonia (CSID),[25] and more recently, the estimated pitch strength (PS).[26]

CPPS measures the amount of noise or energy in a voice sample, either sustained vowel or connected speech. The general analysis for CPPS is as follows. First, the sample is converted from the spectrum to a cepstrum domain. Then, the difference between the cepstral peak (first rahmonic) and the point of equal frequency on a regression line is computed.[19] The regression line takes into account the energy and the window size of the cepstral analysis to avoid variability in the amplitude of phonation, as well as background noise in the testing environment. CPPS has been shown to have sensitivity and specificity at or above 0.77 for detecting dysphonia in both sustained vowels (CPPS-/a/) and running speech (CPPS-s),[19] with greater values of CPPS indicating higher (better) voice quality. In addition, both CPPS-/a/ and CPPS-s seem to be strongly correlated with listener perception of grade (−0.8 and −0.86, respectively) and breathiness (−0.7 and −0.71, respectively) on the GRBAS (grade, roughness, breathiness, asthenia, strain) scale.[23]

CPPS serves as a basis for calculating other indices for voice quality, including the CSID and the AVQI. Both of these measures are multivariate estimates of dysphonia severity and include variables other than CPPS alone to compute the dysphonia severity score. For example, the AVQI is calculated using a weighted combination of six time-frequency and frequency-domain metrics: CPPS, shimmer local, shimmer local dB, long-term average spectrum slope, and long-term average spectrum-tilt, and harmonics-to-noise ratio of the concatenated acoustic signal.[26,27] To obtain the AVQI, both sustained vowel and continuous speech samples are required. The CSID seems to be better associated with listener ratings of severity than the CPPS alone.[28]

In contrast, PS comes from psychoacoustic studies examining how listeners perceive sound. Although this approach has not been used widely in voice, it is a well-established aspect of sound perception.[29] From a psychoacoustic perspective, pitch is composed of three elements—the "pitch height," which allows a listener to scale sounds from low to high; the "PS," which is an estimate of the saliency of a pitch or the degree of tonality; and the "pitch chroma," which is a cyclical property that make sounds that are an octave apart sound similar. To further illustrate PS, consider the same note played on a flute and a piano. The pitch height for the two sounds would be equal, as the notes are the same, but the PS is greater in the note played by the piano.

This concept is applicable to a broad range of vocal acoustic signals, including voices that lack a clear fundamental frequency. As a result, PS is able to describe a wide range of dysphonic voices and has been successfully used in modeling the perception of dysphonic voice quality.[26,30] Previous studies have shown that PS has an inverse relationship with the perceptual construct of breathiness; the higher the value of PS, the lower the score of listener-perceived breathiness ($r = 0.989$).[28] A recent study demonstrated the effectiveness of PS in comparing pre- and post-treatment voices in patients with presbyphonia, vocal fold masses, and muscle tension dysphonia, supporting its potential as a clinical outcome measure.[31]

The present study builds on an earlier report that compared estimated PS values pretreatment and post treatment.[31] In this current retrospective cohort study, sustained /ɑ/ vowels and running speech samples were compared from 22 patients treated with type I thyroplasty for glottic insufficiency at Lakeshore Professional Voice Center, Lakeshore Ear, Nose & Throat Center (St. Clair Shores, MI). The purpose of the present study was to compare changes in estimated PS with other conventionally used acoustic measures based on the cepstral peak prominence (CPPS, CSID, and AVQI) and clinical judgments of voice quality (GRBAS scale) following laryngeal framework surgery. It was hypothesized that PS would differentiate changes in voice quality before and after thyroplasty at levels similar to, or better than, the CPPS-based measure. We also hypothesized that PS would show correlations to clinician judgments of voice quality using the GRBAS scale that would be similar to, or better than, the CPPS-based measures.

METHODS

Patients

The present study was approved by both the St. John Hospital and Medical Center's Institutional Review Board and the Michigan State University Human Research Protection Program. Patients were selected randomly from a database preserved at the Lakeshore ENT and Professional Voice Center. Twenty-two patient samples, which included both steady vowel productions and running speech, were selected randomly from a database. Of these 22 patients, 19 had unilateral vocal fold paralysis, 2 had vocal fold paresis, and 1 had an arytenoid dislocation. There were 7 men (ages 23−80 years old) and 15 women (ages 16−72 years old).

Recordings

Two sets of recordings were obtained from each patient. The first recording occurred at the patients' first visit to the clinic as a part of their routine voice evaluation. The second recording was obtained following phonosurgical intervention (type I thyroplasty) and 6 weeks of voice therapy. Each recording set included both sustained /a/ phonation and a reading of the first paragraph of the Rainbow Passage. Recordings were made with a condenser professional microphone (Perception 120; AKG, Vienna, Austria) connected to a digital recorder (Marantz PMD 671; Oade Brothers Audio, Thomasville, GA) in an IAC sound suite (IAC Acoustics, North Aurora, IL). The recording microphone-to-mouth distance was held constant at 10 cm. Digital recordings were made at a sampling rate of 48 kHz with 16 bits per sample quantization. For analysis, all recordings were resampled to 44.1 kHz. For analysis, the second sentence of the Rainbow Passage was extracted, which is consistent with prior research studies.[32,33]

Acoustic analysis

Analysis of all samples was conducted using custom *MATLAB* scripts (MathWorks, Natick, MA) and *PRAAT*.[34] Custom writing *MATLAB* scripts calculated pitch and PS using the *Aud-SWIPE* (*Auditory Sawtooth Waveform Inspired Pitch Estimator*[35,36]). Simultaneously, the *MATLAB* code also called in *PRAAT* calculated CPPS using the published routines, which are a part of the AVQI.[26] Finally, the CSID was estimated from using the *Analysis of Dysphonia in Speech and Voice* (ADSV model 5109; KayPENTAX, Montvale, NJ). It should be noted that for the CPPS, PS, and CSID, the second sentence of the Rainbow Passage was used; for the AVQI, steady vowel recordings were also used and used the entire production of the steady vowel available in the recording.

Auditory-perceptual analysis

The pre- and post-treatment recording of the second sentence of the Rainbow Passage were randomly presented to six fellowship-trained voice pathologists in a quiet room using headphones. The type of headphones was not held standard because listeners completed ratings remotely. All raters reported normal hearing, had 1−10 years of postfellowship clinical experience, and used GRBAS regularly in clinical practice. Twenty-five percent of the stimuli were randomly selected and repeated to allow estimation of the intrarater reliability. Each stimulus was rated using the GRBAS method (scale from 0−3).

Statistical analysis

Intra- and inter-rater reliabilities were calculated using intraclass correlation (ICC[37]) for each rating category on the GRBAS scale. Because factors such as age, diagnosis, and severity of disorder were not the focus of the present study, statistical analysis was primarily comparative in nature. Therefore, parameter change (pretreatment to post treatment) was analyzed for each metric individually using a simple paired Student *t* test (one-tailed, assuming that each voice should improve, thus parameters should be moving in only one direction). Type I errors due to multiple comparisons were controlled for using Bonferroni correction (five GRBAS measures and four acoustic measures). Additionally, simple cross correlations across all metrics were conducted to compare similarities.

RESULTS

All four acoustic metrics showed a significant change from pretreatment to post treatment (Table 1). Both PS and CPPS increased significantly (33.19%−37.40%, $P = 0.0039$, and 10.20−13.95 dB, $P < 0.0001$, respectively), whereas the CSID and the AVQI decreased significantly (58.47−40.75, $P < 0.0001$, and 4.92−3.30, $P < 0.0001$). Higher values for PS and CPPS indicate more periodic signals and are usually indicative of a better voice quality. Similarly, lower values for the CSID and the AVQI indicate a reduction in dysphonia severity. The skewness of the metrics (although using a small sample size) indicates that the distributions are between *fairly symmetrical* (between −0.5 and 0.5) and *moderately skewed* (between 0.5 and 1.0, both negative and positive). The kurtosis (eg, combined weight of tails relative to the rest of the distribution) of the parameters was calculated; a kurtosis less than 0 indicated light tails of the distribution, and values greater than 0 indicated a heavy tailed distribution[38]; the

TABLE 1.
Acoustic Metrics From the Second Sentence of the Rainbow Passage

	Description	PS	CPPS (dB)	CSID	AVQI
Pretreatment	Mean	33.19	10.20	58.47	4.92
	SE	1.56	−0.73	5.06	0.37
	Skewness	−0.89	−0.52	0.11	0.82
	Kurtosis	0.12	0.08	−1.18	2.30
Post-treatment	Mean	37.40	13.95	40.75	3.30
	SE	0.99	0.40	3.36	0.24
	Skewness	0.18	−0.49	−0.09	−0.15
	Kurtosis	−0.83	0.19	−0.68	−0.13
t Test	*P*	0.0039	<0.0001	<0.0001	<0.0001
Cohen's *d*		0.70	0.83	0.90	1.11

Note: Pre- and post-treatment and *P* values (paired *t* test, one-tailed) are shown.
Abbreviation: SE, standard error.

TABLE 2.
Average GRBAS (Six Reviewers) From the Second Sentence of the Rainbow Passage

	Description	G	R	B	A	S
Pretreatment	Mean	1.58	1.10	1.11	0.88	0.97
	SE	0.19	0.16	0.20	0.19	0.12
	Skewness	0.31	0.72	0.71	0.91	0.64
	Kurtosis	−1.19	−0.19	−0.80	−0.46	−0.45
Post-treatment	Mean	0.81	0.71	0.39	0.25	0.62
	SE	0.09	0.08	0.09	0.08	0.08
	Skewness	0.41	0.80	1.07	1.88	0.75
	Kurtosis	−0.96	0.19	0.05	2.84	−0.44
t Test	*P*	0.0002	0.0021	0.0002	0.0012	0.0015
Cohen's *d*		1.13	0.67	1.04	0.93	0.75

Note: Pre- and post-treatment and *P* values (paired Student *t* test, one-tailed) are shown.
Abbreviation: SE, standard error.

pretreatment AVQI had a heavy-tailed distribution representing the wide range of AVQI values before treatment. Note that the paired *t* tests compared the pre- and post *differences* in acoustic measures, which were more normally distributed and appropriate for parametric testing.

Average GRBAS ratings from the pre- and post-treatment recordings across all subjects are shown in Table 2. Results from *t* tests (with Bonferroni correction) on each rating indicate that all ratings were significantly lower (improved) post treatment.

The mean intrarater and inter-rater reliabilities for the GRBAS ratings ranged from 0.76 to 0.89 and from 0.70 to 0.90, respectively. These values are consistent with prior literature indicating a high level of reliability.[16,39] Table 3 shows intra-ICC(2, *k*) for the GRBAS ratings on the repeated files. One rater (VI) produced an intra-ICC(2, *k*) of 0.0 because of floor effects in ratings of asthenic and was excluded from mean and standard deviation calculations for asthenia. Table 4 shows inter-ICC(2, *k*) for each GRBAS rating category across all raters.

TABLE 3.
Intrarater Reliability for Each Rater

Rater	G	R	B	A	S
I	0.90	0.49	0.92	0.96	0.61
II	0.81	0.89	0.95	0.85	0.77
III	0.90	0.65	0.95	0.92	0.80
IV	0.81	0.86	0.91	0.78	0.81
V	0.92	0.74	0.91	0.88	0.83
VI	0.96	0.95	0.70	0.00*	0.83
Mean	0.88	0.76	0.89	0.88	0.78
SD	0.06	0.17	0.09	0.07	0.09

Note: All values are ICC(2, *k*) measured over 2 replicates and 12 stimuli.
* Excluded from mean and SD.
Abbreviation: SD, standard deviation.

TABLE 4.
Inter-rater Reliability Over All Raters

	G	R	B	A	S
ICC(2, *k*)	0.85	0.81	0.90	0.89	0.70

Note: All values are ICC(2, *k*) measured over 6 raters and 46 stimuli. Ratings were averaged over replicates (12 stimuli).

Cross correlation of the GRBAS ratings and the three acoustic measures is shown in Table 5. Grade and breathiness correlated strongly with the acoustic measures ($|r| \sim 0.7$). All four metrics were highly correlated ($|r| > 0.85$), and CPPS and AVQI were very highly correlated (−0.95) because of the presence of CPPS as a factor in the calculation of AVQI.

DISCUSSION

All measures (PS, AVQI, CSID, CPPS, and GRBAS ratings) showed that the patients had significant voice quality improvement following clinical intervention. This finding illustrates that each of these four objective measures are potentially of value in the clinic. All four objective measures correlated well with GRBAS dimensions except for strain. The weak correlation with strain is consistent with prior literature comparing subjective ratings and objective measures.[40] Furthermore, two factors may have impacted the weak correlation between the four voice outcome measures evaluated here and subjective judgments of strain. First, the primary deficit in patients with unilateral vocal fold paralysis is poor glottic closure, which leads to significant breathiness and sometimes roughness, but often has limited change in strain. The weak correlation may simply reflect the lack of variability in strain seen in these patients. Second, many of the outcome measures evaluated here were developed primarily to capture changes in breathiness or roughness. PS, in particular, was designed as a measure of breathiness and correlated very well with grade and breathiness of the GRBAS scale.

PS is a robust measure, based on psychoacoustic analysis. It is grounded between two values: white noise, which has a PS of zero, and a perfectly periodic sawtooth wave, which is assigned a value of one. Similar to CPPS, CSID, and AVQI, PS can easily be put into simple clinical software and is applicable to a wide range of voice signals. PS can be estimated from connected speech, and the acoustic signal needs minimal

TABLE 5.
Cross Correlation for All Ratings and Measures

	G	R	B	A	S	PS	CPPS	CSID	AVQI
G	1.00								
R	0.69	1.00							
B	0.84	0.29	1.00						
A	0.76	0.18	0.91	1.00					
S	0.58	0.59	0.33	0.44	1.00				
PS	−0.80	−0.65	−0.68	−0.50	−0.31	1.00			
CPPS	−0.73	−0.43	−0.74	−0.62	−0.26	0.69	1.00		
CSID	0.79	0.46	0.78	0.66	0.30	−0.85	−0.79	1.00	
AVQI	0.69	0.41	0.69	0.54	0.18	−0.69	−0.93	0.73	1.00

conditioning or manual checking to provide quality output. Although CPPS and PS are both clinically friendly and relatively easy to use, cepstral peak measures on voice have been shown to be dependent on several factors unrelated to the talker.[41] Estimated PS, on the other hand, is based on a perceptual construct at its core. Although PS did not show any benefit over the cepstral-based measurements in the present study, the psychoacoustic construct of PS has the advantage of attributing it directly to how a listener perceives sound quality. This attribute may not be feasible with cepstrum-based analyses.

There are several limitations to the present study. First of all, it is retrospective and acoustic analyses were performed on recorded samples at one voice center. Real-time analysis in a prospective study involving multiple institutions would yield stronger evidence for the use of measures to track treatment outcomes. Also, all recordings were made in a sound-treated booth. The majority of speech pathologists likely do not have as controlled an environment to obtain recordings. This limitation could have potential implications on the clinical utility of these measures. To this end, current research efforts are looking to quantify which measures are least sensitive to background noise and could prove useful for determining the ideal acoustic measure for everyday clinical use.

CONCLUSIONS

PS is a practical, easily obtained acoustic measure of voice quality that has shown promise for use in the clinical environment to assess treatment-related changes in voice disorders. PS performs at least as well as CPPS and AVQI and is a true perceptual construct. Reliable objective outcome measures are important to develop to continue to evaluate our diagnostic and treatment capabilities and to create evidence-based protocols for the management of voice disorders.

Acknowledgments

The authors thank Collin Potter for the hours spent sorting, trimming, cleaning, and preparing files; Mark Berardi for the initial acoustic analysis; and our nonauthor raters, Juliana Codino, Jonelyn Langenstein, James Curtis, and Melissa Cooke.

REFERENCES

1. American Speech-Language-Hearing Association. Evidence-based practice in communication disorders: An introduction [Technical Report]. 2004.
2. Sackett DL. Evidence-based medicine. *Semin Perinatol*. 1997;21:3–5.
3. Cohen SM, Kim J, Roy N, et al. Factors influencing referral of patients with voice disorders from primary care to otolaryngology. *Laryngoscope*. 2014;124:214–220. https://doi.org/10.1002/lary.24280.
4. Heman-Ackah YD, Sataloff RT, Hawkshaw MJ, et al. What to expect during a visit with a voice doctor. Part II: the examination. *J Singing*. 2009;65:321.
5. Roy N, Barkmeier-Kraemer J, Eadie T, et al. Evidence-based clinical voice assessment: a systematic review. *Am J Speech Lang Pathol*. 2013;22:212–226.
6. Mehta DD, Hillman RE. Voice assessment: updates on perceptual, acoustic, aerodynamic, and endoscopic imaging methods. *Curr Opin Otolaryngol Head Neck Surg*. 2008;16:211–215.
7. Paul BC, Chen S, Sridharan S, et al. Diagnostic accuracy of history, laryngoscopy, and stroboscopy. *Laryngoscope*. 2013;123:215–219.
8. Krausert CR, Olszewski AE, Taylor LN, et al. Mucosal wave measurement and visualization techniques. *J Voice*. 2011;25:395–405.
9. Teitler N. Examiner bias: influence of patient history on perceptual ratings of videostroboscopy. *J Voice*. 1995;9:95–105.
10. Hapner ER, Johns MM. Recognizing and understanding the limitations of laryngeal videostroboscopy. *Perspect Voice Voice Disord*. 2007;17:3–7.
11. Shrivastav R. The use of an auditory model in predicting perceptual ratings of breathy voice quality. *J Voice*. 2003;17:502–512.
12. Kreiman J, Gerratt BR, Kempster GB, et al. Perceptual evaluation of voice quality: review, tutorial, and a framework for future research. *J Speech Hear Res*. 1993;36:21–40.
13. Marks KL. *Improving Agreement Between Graduate Students And Experts For CAPE-V Measures [M.S.]*. Ann Arbor, MI: MGH Institute of Health Professions; 2015.
14. Iwarsson J, Reinholt Petersen N. Effects of consensus training on the reliability of auditory perceptual ratings of voice quality. *J Voice*. 2012;26:304–312.
15. Iwarsson J, Bingen-Jakobsen A, Johansen DS, et al. Auditory-perceptual evaluation of dysphonia: a comparison between narrow and broad terminology systems. *J Voice*. 2017. https://doi.org/10.1016/j.jvoice.2017.07.006 pii: S0892-1997(17)30196-0.
16. Maryn Y, Roy N, De Bodt M, et al. Acoustic measurement of overall voice quality: a meta-analysis. *J Acoust Soc Am*. 2009;126:2619–2634.
17. Bielamowicz S, Kreiman J, Gerratt BR, et al. Comparison of voice analysis systems for perturbation measurement. *J Speech Hear Res*. 1996;39:126–134.
18. Zhang Y, Jiang JJ. Acoustic analyses of sustained and running voices from patients with laryngeal pathologies. *J Voice*. 2008;22:1–9.
19. Heman-Ackah YD, Heuer RJ, Michael DD, et al. Cepstral peak prominence: a more reliable measure of dysphonia. *Ann Otol Rhinol Laryngol*. 2003;112:324–333.

20. Titze IR. *Workshop on Acoustic Voice Analysis: Summary Statement*. Denver, CO. National Center for Voice and Speech; 1995.
21. Little MA, McSharry PE, Hunter EJ, et al. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Trans Biomed Eng*. 2009;56:1015.
22. Stepp CE, Hillman RE, Heaton JT. The impact of vocal hyperfunction on relative fundamental frequency during voicing offset and onset. *J Speech Lang Hear Res*. 2010;53:1220–1226.
23. Heman-Ackah YD, Michael DD, Goding Jr GS. The relationship between cepstral peak prominence and selected parameters of dysphonia. *J Voice*. 2002;16:20–27.
24. Maryn Y, De Bodt M, Roy N. The acoustic voice quality index: toward improved treatment outcomes assessment in voice disorders. *J Commun Disord*. 2010;43:161–174.
25. Awan SN, Roy N, Zhang D, et al. Validation of the cepstral spectral index of dysphonia (CSID) as a screening tool for voice disorders: development of clinical cutoff scores. *J Voice*. 2016;30:130–144.
26. Shrivastav R, Eddins DA, Anand S. Pitch strength of normal and dysphonic voices. *J Acoust Soc Am*. 2012;131:2261–2269.
27. Maryn Y, Corthals P, Van Cauwenberge P, et al. Toward improved ecological validity in the acoustic measurement of overall voice quality: combining continuous speech and sustained vowels. *J Voice*. 2010;24:540–555.
28. Peterson EA, Roy N, Awan SN, et al. Toward validation of the cepstral spectral index of dysphonia (CSID) as an objective treatment outcomes measure. *J Voice*. 2013;27:401–410.
29. Fastl H, Zwicker E. *Psychoacoustics: Facts and Models*. New York; Berlin: Springer; 2007.
30. Eddins DA, Anand S, Camacho A, et al. Modeling of breathy voice quality using pitch-strength estimates. *J Voice*. 2016;30:774 e1-774.e7.
31. Kopf LM, Jackson-Menaldi C, Rubin AD, et al. Pitch strength as an outcome measure for treatment of dysphonia. *J Voice*. 2017;31: 691–696.
32. Watts CR, Awan SN. Use of spectral/cepstral analyses for differentiating normal from hypofunctional voices in sustained vowel and continuous speech contexts. *J Speech Lang Hear Res*. 2011;54:1525–1537.
33. Rothman HB, Brown Jr WS, Sapienza CM, et al. Acoustic analyses of trained singers perceptually identified from speaking samples. *J Voice*. 2001;15:25–35.
34. Boersma P, Weenink D. PRAAT: A system for doing phonetics by computer. Report of the Institute of Phonetic Sciences of the University of Amsterdam 132. 1996.
35. Camacho A, Harris JG. A sawtooth waveform inspired pitch estimator for speech and music. *J Acoust Soc Am*. 2008;124:1638–1652.
36. Camacho A. On the use of auditory models' elements to enhance a sawtooth waveform inspired pitch estimator on telephone-quality signals. 2012 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA); 2012 2−5 July. 2012.
37. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86:420.
38. Westfall PH. Kurtosis as peakedness, 1905 − 2014. R.I.P. *Am Stat*. 2014;68:191–195.
39. Shrivastav R, Sapienza CM, Nandur V. Application of psychometric theory to the measurement of voice quality using rating scales. *J Speech Lang Hear Res*. 2005;48:323–335.
40. Bhuta T, Patrick L, Garnett JD. Perceptual evaluation of voice quality and its correlation with acoustic measurements. *J Voice*. 2004;18:299–304.
41. Skowronski MD, Shrivastav R, Hunter EJ. Cepstral peak sensitivity: a theoretic analysis and comparison of several implementations. *J Voice*. 2015;29:670–681.