Comparison of Post-therapy Dysphonic Voices and Normal Voices

Natalie Schaeffer and Akiko Fuse, *Brooklyn, New York*

Summary: Purpose. The purpose of the present investigation was to compare the voices of post-therapy dysphonic participants with participants who have normal voices to determine how close the corrected voices approached normal vocal levels. The present investigation is a follow-up to the authors' previous research in which dysphonic participants, with voices ranging from moderate-to-severe dysphonia, were evaluated pre- and post therapy using the Dysphonic Severity Percentage scale and the interval scale.

Methods. In the present study, five raters, three speech-language pathologists experienced in assessing dysphonia, and two trained speech-language pathology college students evaluated 20 participants with normal voices under the same two conditions as those of the corrected participants—when reading a paragraph aloud and during spontaneous speech. While listening to the recordings of the normal voices, the raters tallied any dysphonic syllables produced by the participants to obtain a Dysphonic Severity Percentage for both paragraph reading and spontaneous speech. The raters also evaluated the normal voices on the interval scale. These data were compared with those of the post-therapy participants, who were evaluated under the same conditions and methods pre- and post therapy.

Results and Conclusion. The dysphonic participants' voices improved significantly post therapy in comparison with their pretherapy result; their improvement, however, was not commensurate with the voices of the normal participants, and the data showed a significant difference between the two groups. Both evaluation scales reflected a high agreement among raters.

Key Words: Dysphonic Severity Percentage (DSP)−Interval scale−Post dysphonic participants−Participants with normal voices−Experienced raters.

INTRODUCTION—REVIEW OF THE LITERATURE

According to Ramig and Verdolini,[1] research has shown mounting evidence of the benefits of voice disorder treatment by speech-language pathologists (SLPs). These authors state that it is critical to note whether the voice disorders result from functional, organic, or psychological factors, or any combination of the three factors. Phonotraumatic behaviors include excessive throat clearing, shouting over background noise, and screaming. Furthermore, voice disorders may also be the result of stress, personality disorders, or conversion reactions. The inability to perform a certain job secondary to impaired vocal functioning can be referred to as a handicap. Dysphonia caused by misuse and hyperfunction can result in vocal fatigue or a total loss of voice. Studies have shown that voice treatment can eliminate the need for surgical intervention, which may not be effective overall.[1]

Hasanvand et al[2] stress that, when evaluating voice disorders, it is important for SLPs to assess both continuous speech and sustained vowels to obtain a full picture of the client's vocal pattern. Because continuous speech allows for evaluation of the client's everyday vocal patterns, timed assessments are not applicable in this case. According to the authors, timed assessments (eg, s/z ratio) are not appropriate for the identification of vocal characteristics that need to be targeted in treatment. Voices that are considered normal have both periodic and harmonic frequency acoustic signals, whereas disordered voices are classified as nonharmonic. Although an auditory perceptual assessment plays a key role in diagnosing voice disorders, it is a subjective measure.[2] Ghio et al[3] stated that the most common aspects assessed in voice examinations are roughness, grade (overall grade of hoarseness), strain, and asthenia. These authors argue that it is problematic to use prolonged vowels, as this method does not incorporate everyday speech.

Voice rating scales have been used to determine the severity of dysphonia. For example, the equal appearing interval scale can range from one point to five points, with one point being least severe and five points being most severe.[4,5] According to Patel et al,[6] the units on a scale must be clearly defined for inter-rater consistency, and different raters may interpret the units on a scale differently.[7] Another type of rating scale is the Consensus-Auditory-Perceptual-Evaluation of Voice (CAPE-V, described by American Speech-Language-Hearing Association) is an analog scale in which raters assess both overall severity of the dysphonia and specific dysphonic characteristics (ie, roughness, breathiness, strain, pitch, and intensity or loudness). "The rater subjectively judges the degree of perceived deviance from the norm on a continuum from 0 to 100 for sustained vowels, sentences, and running speech."[8] The Oveall grade of vocal hoarseness, roughness, asthenia, strain scale GRBAS is also a clinically based scale that is a four-point scale from 0 (normal) to 3 (extreme). The rater assigns a number that represents mild (1), moderate (2), and severe (3): in this scale, the G represents the overall grade of vocal hoarseness, as well as the specific parameters of rough (R), breathy (B), asthenic or weak (A), and strained (S).[9]

Accepted for publication December 1, 2017.

From the Department of Speech Communication Arts and Sciences, Brooklyn College, Brooklyn, New York.

Address correspondence and reprint requests to Natalie Schaeffer, Department of Speech Communication Arts and Sciences, Brooklyn College, 3310 Nostrand Ave. Apt. 501, Brooklyn, NY 11229. E-mail: natalies@brooklyn.cuny.edu

Journal of Voice, Vol. 33, No. 4, pp. 581.e1−581.e6

0892-1997

https://doi.org/10.1016/j.jvoice.2017.12.001

Perceptual rating scales are subjective and thus open to biases and variability.[7,10] In terms of voice evaluation techniques, Schaeffer[11] evaluated clients using the Dysphonic Severity Scale (DSP) scale pre- and post therapy, as this method assessed the participants' continuous speech (typed by the principal investigator), as well as a paragraph read aloud by the participants. The raters circled only the dysphonic syllables, thus making it easier to evaluate these voices perceptually; that is, the evaluators simply had to identify if dysphonia was present or not present in the syllables of the voice samples. This perceptual rating scale used with spontaneous speech and a read-aloud paragraph illustrated a high degree of inter-rater reliability and consistency among experienced raters and trained student raters, possibly because the raters focused and circled only those areas (while listening to the recordings) where dysphonia was present. This assessment was performed for voices in the initial research (pre- and post therapy of the participants), using raters experienced with voice and trained student raters, pre- and post therapy. As noted, perceptual evaluation of dysphonia, using the DSP scale, indicated high reliability among raters (0.901) for both experienced raters and trained student raters. Because raters could selectively listen to the syllables in the content to tally a percentage of dysphonia (pretherapy),[11] the DSP method reduced subjectivity. An interval scale confirmed the DSP results.

Kreiman et al[10] state that, in the process of rating vocal quality, it has been found that multiple factors affect listener perception. For example, internal standards may give rise to biases in judging vocal quality; that is, listeners have comparable internal standards for normal voice related to experiences in daily interactions; inexperienced raters tend to judge dysphonic voices with their "normal" internal standard, whereas experienced raters have a separate standard for dysphonic voices because of their exposure to the target population. Training, however, may help create a new standard for inexperienced raters. Additionally, "listener drift" can also affect ratings. For example, a moderate voice may be rated as more severe when preceded by mild voices. The units on the scale must be clearly defined to increase inter-rater consistency.

Eadie and Baylor[12] found that after the training of students in the perception of dysphonia, inter-rater reliability was least variable for overall severity and improved on vowels with regard to roughness and breathiness. These authors contend that the training program has an effect on listener reliability. Listeners must be consistent in their evaluation of voice and use the scale the same way each time. These authors contend the aspects that can contribute to variability are sample size, errors in specificity, and clarity of the rating scales. Methods of training may also contribute to the variability in raters' decisions.

Researchers have also postulated that variability may be attributed to the sample size of experienced raters in the studies, which typically range from one to three raters. Additionally, errors in specificity and clarity of the rating scales, as well as methods of training that differ between studies, may add to the variability in the findings.[6,7]

De Bodt et al[9] presented voice samples to both experienced and inexperienced ear, nose and throat (ENT) doctors and SLPs. Voice samples were rated on the GRBAS scale. Reliability was strongest for the G (grade) parameter than for the S (strained) parameter on the GRBAS scale. The results suggest that listeners, especially inexperienced listeners (ie, both ENT and SLPs), have difficulty isolating discrete aspects of voice (eg, breathiness and roughness). Kreiman and Gerratt[13] highlight the difficulty of isolating individual parameters of voice, thus undermining the validity and utility of such rating scales. Moreover, listener agreement was found to be higher for more mild and severe voices than for moderate ones, as there are few options on the end points of the rating scale. Eadie and Doyle[14] also support the use of a scale that assesses the overall severity of voice samples.

The Dysphonic Severity Percentage (DSP) is aimed at obtaining specific quantitative percentages of dysphonia by tallying the number of dysphonic syllables in a 100-syllable paragraph of spontaneous speech and in a paragraph reading of 122 syllables of written samples. The raters independently listen to the recordings of dysphonic voice samples and focus on and circle the dysphonic syllables on the written samples. Voice characteristics such as breathy, strained, and rough are circled but not specifically named, as they are components of the overall percentage of dysphonia.

As an attempt to increase reliability in voice ratings and adjust internal standards, many studies suggest the use of external reference standards, known as anchors. These studies demonstrated that anchors help increase inter-rater reliability among inexperienced raters. For example, Gerratt et al[15] presented 12 experienced raters with anchored and unanchored stimuli (rated roughness of synthesized voice stimuli). Overall, results indicated increased inter-rater reliability on anchored stimuli. Chan and Yiu[16] studied the effects of anchors, training, or both (ie, anchors and training) on the reliability of raters using a 100-mm visual analog scale. Results indicated that pairing anchors with training improved reliability in rating dysphonic voices. In a second study conducted by Chan and Yiu,[17] anchors and an alternative training method were used to evaluate breathiness in vocal stimuli. Both methods resulted in increased perceptual accuracy in inexperienced raters. Eadie et al[18] investigated the use of anchors and the effects of listener experience on the evaluation of dysphonia. These authors found that anchors increased consistency in rating, as well as inter-rater reliability. Although experience did not significantly affect the evaluation of dysphonia when using anchors, there was a slight decrease in inter-rater variability.

According to Hasanvand et al,[2] cepstral peak prominence and cepstral peak prominence—smooth are frequency measures developed to identify breathiness and roughness in vocal quality. These measures have been shown to work in continuous speech and prolonged vowel production: normal voices have a higher cepstral peak that is farther from the regression line than a disordered voice, whereas a disordered voice has a shorter peak or is close to the regression line.

A study by Aghajanzadeh et al[19] examined how the following factors influence or are related to the perceptual

assessment of different types of dysphonia: vital capacity, maximum phonation time (MPT), and phonation quotient. For example, professionals can assess how participants with dysphonia use respiration, the degree of effort exerted for phonation, and compensatory strategies. Assessing these factors can provide SLPs with information about the client's physiological context of phonation. To evaluate aerodynamic status, MPT, phonation quotient (determines average sufficiency of the glottis), and vital capacity may be incorporated in the evaluation.

Lin et al[20] found "a significant correlation between acoustic measurements and self-reported dysphonic symptoms." A study by Schaeffer and Kim[21] showed that the Phonatory Aeryodynamic System reflected measurements (eg, pitch, airflow rate, and sound pressure level) that were consistent with participants' reports of their dysphonic symptoms. Schaeffer and Kim's study included data on pre- and post stimulation. The same measurements were taken after stimulation, and an improvement in the values toward normal was reflected in the data. Furthermore, the participants reported a reduction in a strained voice and easier phonation.

NEED FOR THE PRESENT RESEARCH

Research that compares post-therapeutic phonotraumatic voices with normal voices is unavailable or scarce. There is a need to measure how close post-therapeutic voices can approach normal voice production to increase understanding regarding the components of dysphonia. The present research, therefore, compared post-therapeutic participants with participants with normal voice. The DSP scale and the interval scale were selected (as in the initial research) to determine if the DSP and the interval scores of corrected dysphonic participants[22] matched the DSP and the interval scores of participants with normal voices, using the same 122-syllable paragraph and the 100-syllable spontaneous speech sample of the normal group.

The following questions were asked:

1. How do the DSP and interval scale results of post-therapeutic participants compare with the results of participants with normal voices?
2. Was there agreement between the interval scale and the DSP scale results with regard to the normal group?
3. Was the inter-rater agreement strong?

METHODS

Experienced SLPs with 10 or more years of experience with voice disorders listened to 20 participants (primarily students) with normal voices, aged 20–46 (one participant), with a mean age of 26.25. These participants were recruited from Brooklyn College and agreed to participate in the research. Exclusion criteria were respiratory problems, neurologic difficulties, reflux, and smoking. A spontaneous speech sample was taken from the normal group to be compared with the sample from the post-therapy group. Moreover, all the participants with normal voices read the same paragraph, "Arthur the Shirker" (Appendix), into a tape recorder (SONY Corporation, Shanghai, China) as did the post-therapy group. The paragraph consisted of 122 syllables, and the spontaneous speech sample consisted of the first hundred syllables. Voice samples were obtained in a quiet room on a SONY tape recorder with a Compaq microphone (Deskpro 6000EP Series, Shanghai Shuai Yin electronics Co. LTD); the distance between the microphone and the participants was 10 cm or 5 inches from the mouth, as per McManus.[23] The content was orthographically transcribed to obtain a written sample of the participants' voices during spontaneous speech. All experienced raters agreed that the voices of these participants were normal.

In the original research, there were 10 dysphonic participants (three males and seven females) with moderate-to-severe phonotrauma. The ages of these participants with phonotrauma ranged from 21 to 48 years, with a mean age of 33.4 years. These participants all received an ENT examination before participating in the study and met the following criteria: negative medical histories (except for dysphonia), no respiratory problems, ability to read and speak English, no previous training in singing, and no history of smoking.[11]

For the investigation, each participant gave a sample of his or her spontaneous speech and read the paragraph as well. As noted, the spontaneous sample consisted of the first 100 syllables, and the paragraph consisted of 122 syllables.

All the participants significantly improved their voices. Therapy took place in semester or about 4 months. As noted under Limitations, retesting after a period in which the participants use their voices correctly, may result in normal levels. Regression into old habits is a possibility, and some may have to return to therapy for a period to obtain improved levels of motivation and vocal use.

In the original pre- and post-DSP study, the authors measured the first 100 syllables of the spontaneous speech sample and 122 syllables of the paragraph (see Appendix) pre- and post therapy to obtain a percentage of dysphonia by counting the dysphonic syllables. The "frequency" component of the stuttering severity index was adapted to obtain a percentage of dysphonia by counting the dysphonic syllables. In accordance with the protocol, the participants with normal voices also gave a spontaneous speech sample and read the same paragraph. For the normal group, the DSP and the interval scale results were calculated and compared with the data of the post-therapy group.[11]

With regard to the interval scale, the raters in the present study listened to the tapes of the participants' voices, keeping the interval scale in front of them, and circled the number on the scale that represented their (raters') perception of the voice quality. There was high agreement among the raters. The raters in the original study used the interval scale in the same manner to rate the pre- and post-therapeutic groups. The raters in the original study and in the present study were in agreement, as seen by the mean in the interval scale. In the present research, the raters also used the DSP scale to rate

TABLE 1.
Interval Scale

	Mean	Standard Deviation
Normal voices	0.285	0.237
Pretherapy voices	4.0	1.03
Post-therapy voices	1.7	0.978

the normal voices, and there was no circling of dysphonic syllables. The few instances of glottal fry on the interval scale kept the normal participants from having a perfect score. The result on the DSP for the normal group confirmed that none of the syllables were considered dysphonic, despite the instances of glottal fry on the interval scale.

RESULTS

An independent *t* test revealed a significant difference (.001 level) between the normal voices and the post-therapeutic voices on the interval scale for overall voice quality, despite the significant vocal improvement in the dysphonic participants' voices, as illustrated in Table 1 and in Figure 1. The voices of the post-therapeutic group improved significantly on the interval scale and received a mean of 1.7 (down from a mean of 4.0). The raters were told to circle only dysphonic syllables (eg, rough, strained, pressed, breathy, and hoarse) on the DSP and to rate the general voice quality of the normal group on the interval scale. The normal voices received a mean of 0.285 secondary to intermittent glottal fry because a few of the normal participants used intermittent glottal fry. The raters, therefore, did not rate the normal group with a perfect score on the interval scale with regard to general voice quality. As noted, glottal fry (especially intermittent glottal fry) is not considered dysphonic by Boone et al[24] and Colton and Casper.[25] Furthermore, none of the raters circled any dysphonic syllables on the DSP for the normal group. In the initial study,[22] the post-therapeutic

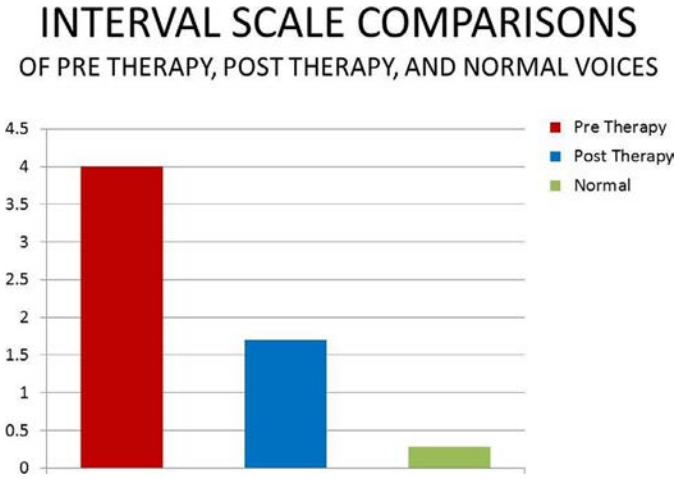

FIGURE 1. Interval scale comparisons of pretherapy, post-therapy, and normal voices.

group improved significantly regarding the reduction of dysphonic syllables on the DSP scale[22] and demonstrated no glottal fry pre- or post therapy.

DISCUSSION

There is little or no literature comparing postvoice therapy participants with participants with normal voices. The discussion will focus, therefore, on the results of the present study's findings. The purpose of the present investigation was to compare the voices of post-therapy dysphonic participants with participants who have normal voices to determine how close corrected voices approached normal vocal levels.

The results of the research indicated a significant difference between the corrected dysphonic group and the participants with normal voices on the interval scale despite the vocal improvement in the dysphonic group's voices post therapy; that is, the corrected dysphonic group did not reach the level of the normal group on the interval scale. A few of the participants in the normal group exhibited intermittent glottal fry (the lowest vocal register), which precluded a perfect score on the interval scale in terms of general vocal quality, as per the raters. For example, some of the ratings were above zero (0.05, 1.0) on the interval scale. This intermittent characteristic, however, is not considered true dysphonia.[25,26] The dysphonic group pre- and post therapy exhibited no glottal fry. It is possible that the vocal fold vibration in their dysphonic voices and in their corrected voices did not allow for the descent into the lowest register.

With regard to the DSP, there were no dysphonic syllables circled for the participants with normal voices, thus confirming normal vocal quality according to the raters. The raters listened to the tape recordings (paragraph and spontaneous speech samples) of the chosen normal voices and found no dysphonic syllables to circle on the DSP scale. There was high agreement among the raters for both the interval and DSP ratings for the normal group. The post-dysphonic group did improve on the DSP scale in that the number of circled dysphonic syllables in the previous study was significantly reduced (see Schaeffer and Sidavi[22]).

That the voices of the dysphonic participants improved significantly is consistent with the findings of Ramig and Verdolini[1] regarding the correction of dysphonia by SLPs. Additionally, the literature stresses the importance of assessing continuous speech,[2] which took place in the present study, but the DSP method made the perceptual evaluation more direct. Furthermore, there was a high rater agreement in terms of the results for both the interval scale and the DSP scale regarding the participants with normal voices. For example, **none** of the raters found dysphonic syllables on the DSP scale, but they recognized glottal fry on a few of the normal participants, which negated a perfect score on the interval scale for those participants.

The advantages of the DSP scale are that raters can focus on actual areas of dysphonia (eg, strained, breathy, and rough) and calculate an overall percentage of dysphonia rather than a single judgment. In the study, the raters had to

make 100 and 122 separate judgments instead of a single judgment, and the DSP resulted in a specific percentage dysphonia; the results present increased specificity and greater quantification. The normal group had no dysphonic syllables circled on the DSP, indicating no dysphonia. Their general quality was slightly reduced on the interval scale; that is, the mean was not 0 because of glottal fry (which is not considered dysphonia).

The advantages of the DSP is that raters can focus into actual dysphonia, be it strained, breathy, rough, and so on. In sum, both the DSP and the interval scales were used for the perceptual evaluation.

As noted by De Bodt et al,[9] it is extremely important for the evaluator to notice the aerodynamic aspects of vocal use observed during continuous speech and to communicate these areas to the SLP who is treating the client. Should there be concern, these areas can be tested separately (eg, MPT) to determine a connection to spontaneous speech.

Another aspect to confirm dysphonia, beside subjective analysis, would be to use cepstrum analysis[2] as this technique allows the SLP to compare normal and dysphonic voice productions in that normal voices have a high cepstral peak and a well-defined harmonic structure, whereas a hoarse or breathy voice results in poor harmonic structure and low cepstral peak. Furthermore, the Multidimensional Voice Program can also define the acoustics of normal and dysphonic voice.[24] Therefore, the combination of evaluating continuous speech, the acoustics, and the aerodynamic aspects of voice production provides information to the SLP to determine how close the corrected voice reach normal production and which aspects need to be addressed.

Lin et al[20] found "a significant correlation between acoustic measurements and self-reported dysphonic symptoms." A study by Schaeffer and Kim[21] showed that the Phonatory Aerodynamic System reflected measurements (eg, pitch, airflow rate, and sound pressure level) that were consistent with participants' reports of their dysphonic symptoms. Schaeffer and Kim's study included data on pre- and post stimulation. The same measurements were obtained after stimulation, and an improvement in the values toward normal was reflected in the data. Furthermore, the participants reported a reduction in strained voice and easier phonation after stimulation.

Although the posttherapeutic participants did not reach those levels of the participants with normal voices in the present study, the previously mentioned delineated techniques may provide insight into aspects that may be compared with the data of participants with normal voices both perceptually and acoustically, and with regard to aerodynamic measures to improve any remaining dysphonic symptoms.

In the final outcome, all the participants significantly improved their voices. Therapy took place in a semester or about 4 months. As noted under Limitations, retesting after a period in which the participants correctly use their voices without therapy, their voices may reach normal levels. Regression into old habits is a possibility, and some may have to return to therapy for a period to obtain improved levels of motivation and vocal use.

LIMITATIONS

A limitation of the study was the small number of dysphonic participants (10 dysphonic participants) and 20 normal participants. A larger N for both groups may have revealed other aspects of the participants' voices (normal and corrected voices).

A follow-up, in terms of retesting the corrected voices, is necessary to determine if the participants maintained or increased their improvements by comparing the follow-up data to results at the conclusion of therapy. Some of the participants may have regressed and will be invited for a period of therapy to again improve their voices and motivation.

CONCLUSION

In sum, a significant difference between the 2 groups on the interval scale was found despite the large vocal improvement in the dysphonic group's voices. Furthermore, the raters found no dysphonic syllables to circle on the DSP scale for the normal group. The dysphonic group, which exhibited significant dysphonic syllables pretherapy, improved significantly post therapy (see Schaeffer and Sidavi[22]). Although the post-therapeutic participants did not reach levels similar to those of the participants with normal voices, the above research may provide insight into which aspects may be compared and addressed in a study that examines post therapeutic participants and participants with normal voices on acoustic and aerodynamic instruments.

Re-measure the corrected voices after a period of carryover without therapy with the following tests to determine if the data have reached normal limits: (1) the DSP, (2) Interval scale, (3) Acoustic and Aerodynamic.

Acknowledgments

The authors thank their colleagues Barbara Bennet, Lucy Girlando, Sunyoung Kim, and their students Pamela Nahmias and Ximena Morocho for their conscientious contributions in rating the participants and their help in organizing the data. I am also grateful to my statistician, Dr. Howard Spivak, for his superior advice regarding this research.

APPENDIX

"Arthur, the Shirker"

Once there was a young rat named Arthur, who could never make up his mind. Whenever the other rats asked him if he would like to go out with them, he would answer, "I don't know." And when they said, "Would you like to stop at home?" he wouldn't say yes or no either. He would always shirk at making a choice. One day his aunt said to him, "Now look here. No one will ever care for you if you carry on like this; you have no more mind than a blade of grass!" The young rat coughed and looked wise as usual, but said nothing.[26]

REFERENCES

1. Ramig LO, Verdolini K. Treatment efficacy: voice disorders. *J Speech Lang Hear Res*. 1998;41:S101–S116.

2. Hasanvand A, Salehi A, Ebrahaimipour M. A cepstral analysis of normal and pathologic voice qualities in Iranian adults: a comparative study. *J Voice*. 2016;31:508.e17–508.e23. https://doi.org/10.1016/j.jvoice.2016.10.017.
3. Ghio A, Dufour S, Wengler A, et al. Perceptual evaluation of dysphonic voices: can a training protocol lead to the development of perceptual categories? *J Voice*. 2015;29:304–311.
4. Sansone FE, Emanual FW. Spectral noise levels and roughness severity ratings for normal and simulated rough vowels produced by adult males. *J Speech Hear Res*. 1970;13:489–502.
5. Yiu EM-L, Ng C-Y. Equal appearing interval and visual analogue scaling of perceptual roughness and breathiness. *Clin Linguist Phon*. 2004;18:211–229.
6. Patel S, Shrivastav R, Eddins DA. Perceptual distances of breathy voice quality: a comparison of psychophysical methods. *J Voice*. 2010;24:168–177.
7. Shrivastav R, Sapienza CM, Nadur V. Application of psychometric therapy to the measurement of voice quality using rating scales. *J Speech Lang Hear Res*. 2005;48:323–335.
8. ASHA. Consensus-auditory-perceptual-evaluation of voice (CAPE-V). The American Speech-Language-Hearing Association, Special Interest, Division 3: Voice and Voice Disorders 2002−2006:1–3.
9. De Bodt MS, Wuyts FL, Van de Heyning PH, et al. Test-retest study of the GRBAS scale: influence of experience and professional background on perceptual rating of voice quality. *J Voice*. 1997;11:74–80.
10. 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.
11. Schaeffer N. Toward a more quantitative measure to assess severity of dysphonia: preliminary observations. *J Voice*. 2008;12:004.
12. Eadie TL, Baylor CR. The effect of perceptual training on inexperienced listeners' judgments of dysphonic voice. *J Voice*. 2006;20:527–544.
13. Kreiman J, Gerratt BR. Sources of listener disagreement in voice quality assessment. *J Acoust Soc Am*. 2000;108:1867–1876.
14. Eadie TL, Doyle PC. Direct magnitude estimation and interval scaling of naturalness and severity in tracheoesophageal (TE) speakers. *J Speech Lang Hear Res*. 2002;45:1088–1096.
15. Gerratt BR, Kreiman J, Antonanzas-Barroso N, et al. Comparing internal and external standards in voice quality judgments. *J Speech Hear Res*. 1993;36:14–20.
16. Chan KMK, Yiu EML. The effects of anchors and training on the reliability of perceptual voice evaluation. *J Speech Lang Hear Res*. 2002;45:111–126.
17. Chan KMK, Yiu EML. A comparison of two perceptual voice evaluation training programs for naïve listeners. *J Voice*. 2006;20:229–241.
18. Eadie TL, Kapsner-Smith M, Rosenzweig J, et al. The role of experience on judgments of dysphonia. *J Voice*. 2010;24:564–573.
19. Aghajanzadeh M, Darouie A, Dabirmoghaddam P, et al. The relationship between the aerodynamic parameters of voice and perceptual evaluation in the Iranian population with or without voice disorders. *J Voice*. 2017;31 250−e9.
20. Lin F, Chen SH, Chen S, et al. Correlation between acoustic measurements and self-reported voice disorders among female teachers. *J Voice*. 2016;30:460–465. https://doi.org/10.1016/j.jvoice.2015.05.013.
21. Schaeffer N, Kim S. Pre- and poststimulation study on the phonatory aerodynamic system on participants with dysphonia. *J Voice*. 2017;31:254.e1–254.e9.
22. Schaeffer N, Sidavi A. Toward a more quantitative measure to assess severity of dysphonia posttherapy: preliminary observations. *J Voice*. 2011;25:159–165.
23. McManus P. *Effective Microphone Techniques for Speakers*. San Diego: Paul R. McManus; 2004.
24. Boone DR, McFarlane SC, Von Berg SL, et al. *The Voice and Voice Therapy*. 9th ed. New York: Allyn & Bacon; 2014:52–53.
25. Colton RH, Casper JK. *Understanding Voice Problems: A Physiological Perspective for Diagnosis and Treatment*. 2nd ed. New York: Lippincott Williams and Wilkins; 1996:84.
26. Anderson V. *In Training the Speaking Voice*. 2nd ed. New York: Oxford University Press; 1977:401.