

Auditory-Perceptual Evaluation of Disordered Voice Quality

Pros, Cons and Future Directions

Jennifer Oates

School of Human Communication Sciences, La Trobe University, Melbourne, Vic., Australia

Key Words

Auditory-perceptual evaluation · Disordered voice quality · Rating scales

Abstract

Auditory-perceptual evaluation is the most commonly used clinical voice assessment method, and is often considered a gold standard for documentation of voice disorders. This view has arisen for many reasons, including the fact that voice quality is perceptual in nature and that the perceptual characteristics of voice have greater intuitive meaning and shared reality among listeners than do many instrumental measures. Other factors include limitations in the validity and reliability of instrumental methods and lack of agreement as to the most sensitive and specific instrumental measures of voice quality. Perceptual evaluation has, however, been heavily criticised because it is subjective. As a result, listener reliability is not always adequate and auditory-perceptual ratings can be confounded by factors such as the listener's shifting internal standards, listener experience, type of rating scale used and the voice sample being evaluated. This paper discusses these pros and cons of perceptual evaluation, and outlines clinical strategies and research approaches that may lead to improvements in the assessment of voice quality. In particular, clinicians are advised to use multiple methods of voice quality evaluation, and to include both subjective and objective evaluation tools.

Copyright © 2009 S. Karger AG, Basel

Introduction

Even among experienced voice clinicians who work in facilities with access to sophisticated acoustic, aerodynamic and vocal fold imaging instrumentation, auditory-perceptual evaluation is the most commonly used clinical assessment method for disordered voice quality [1, 2]. Some clinicians and voice scientists go a step further and consider perceptual assessment as a gold standard for documentation of voice disorders [3–6]. Why is auditory-perceptual evaluation of voice so popular in clinical practice and why is this type of voice assessment considered by some as a gold standard? Does auditory-perceptual evaluation warrant this level of popularity and can we justify considering this type of assessment as a gold standard? This paper endeavours to consider these questions by discussing the pros and cons of auditory-perceptual evaluation on the basis of pragmatic clinical perspectives as well as research evidence. No attempt is made, however, to systematically review the enormous literature on perceptual assessment of voice quality. Rather, reference is made to key papers that exemplify relevant issues. In addition, this paper outlines some clinical strategies and research approaches that may lead to improvements in auditory-perceptual assessment and voice evaluation in general.

This paper is an updated version of a lecture presented at the special symposium of the IALP Voice Committee at the IALP World Congress in Copenhagen, Denmark, 2007. Guest Editor: Jan Svec, Groningen.

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2009 S. Karger AG, Basel
1021-7762/09/0611-0049\$26.00/0

Accessible online at:
www.karger.com/fpl

Dr. Jennifer Oates
School of Human Communication Sciences
La Trobe University
Melbourne, Vic. 3086 (Australia)
Tel. +61 3 9479 1810, Fax +61 3 9479 1874, E-Mail j.oates@latrobe.edu.au

The Pros of Auditory-Perceptual Evaluation of Disordered Voice Quality

The popularity of auditory-perceptual evaluation arises from multiple factors, many of which are admittedly pragmatic, but many of which derive from the fundamental nature of voice quality itself and the significant limitations of alternative instrumental approaches to the measurement of voice.

Most importantly, voice is fundamentally a perceptual phenomenon in response to an acoustic stimulus [7–9]. Perceptual evaluation is therefore a logical candidate for a gold standard assessment of the phenomenon. Because voice is perceptual in nature, perceptual features of voice quality are also likely to have greater shared reality among a wide range of listeners including clinicians, clients, employers and other associates of those clients [10]. A consequence of this shared reality is that perceptual descriptions of voice are potentially intuitively meaningful to anyone [3]. The way in which the results of voice evaluations are communicated to our clients and their associates, their referrers and related health professionals must be meaningful and interpretable. Perceptual descriptions of voice quality can fulfil this requirement. The same cannot be said for many instrumental measures of voice quality. Describing a particular voice as breathy and rough, for example, is likely to be more easily interpreted by a wide range of people than a description that specifies the noise-to-harmonic ratio associated with that voice.

We are now closer than ever before to a widely agreed set of perceptual terms for describing vocal dysfunction. The most commonly used perceptual evaluation systems have many similarities in terms of the voice features evaluated and definitions of those features. The GRBAS (Grade, Roughness, Breathiness, Asthenia and Strain) [11], CAPE-V (Consensus Auditory-Perceptual Evaluation of Voice) [12], Stockholm Voice Evaluation Approach [13] and the Perceptual Voice Profile [14], for example, all incorporate the perceptual features of breathy, rough and strained. This is a very different situation from the 1960s and 1970s when clinicians were faced with a baffling range of perceptual evaluation approaches with inconsistent and undefined terminology [15]. Although consensus on the terms used for describing dysphonia and their definitions does not imply that these voice evaluation systems are associated with adequate validity and reliability, this consensus may contribute to the popularity of auditory-perceptual analysis of disordered voice.

Several research teams have also demonstrated since the 1980s that listeners can make judgements of the over-

all severity of disordered voices and discriminate some voice qualities such as hoarseness, breathiness and roughness consistently and with good intra-rater and inter-rater reliability and agreement, as long as variables known to affect perceptual judgements are controlled [2, 16–20]. Such variables include the type and complexity of the rating task, listener background and training, the type of voice sample and other factors that are outlined later in this paper. Although the variables that affect perceptual judgements of voice can limit the reliability of listener judgements of voice quality, there is some evidence that with adequate control of such potential confounding variables, perceptual evaluation could be reasonably robust. Robust perceptual measures could in turn provide baseline information about the degree and nature of clients' voice problems, and allow clinicians to measure their clients' progress throughout intervention.

The reliability of perceptual judgements of disordered voice quality is not the only determinant of the adequacy of this type of voice evaluation method. The validity of the method (the extent to which the evaluation method measures what it is intended to measure) should also be considered because it is possible for a reliable assessment tool to be invalid [21]. It is unclear whether anything apart from the obvious face validity of auditory-perceptual evaluation (that is, auditory-perceptual evaluation appears to measure listener perceptions and seems a plausible method for making such measures) has had any bearing on the popularity of this type of assessment. There is, however, some evidence from studies reporting rigorous factor analysis approaches and those reporting on the relationships between acoustic and perceptual measures that the validity of perceptual evaluation may be adequate in some circumstances [17, 22–24].

Because auditory-perceptual evaluation of dysphonia has been heavily criticised on the basis of its reliability, instrumental measures have sometimes been promoted as a better alternative. However, despite decades of intensive research, an ideal set of objective instrumental measures of disordered voice quality remains elusive [1, 5, 25, 26]. While some acoustic and aerodynamic measures of voice – such as parameters of the glottal flow waveform, spectral measures such as spectral tilt and cepstral peak prominence smoothed, voice range profiles and high-speed digital imaging of vocal fold vibration – show promise [25, 26], there is still uncertainty as to the most sensitive, specific and valid set of instrumental measures [5]. There is a high degree of inter- and intra-individual variability on some instrumental measures, and therefore uncertainty as to normative values [1, 25, 27]. Objec-

tive measures of voice quality are also susceptible to multiple confounding influences that can be difficult to control in a clinical environment. Such confounding factors include environmental recording conditions, the specific characteristics of hardware and software systems, the testing and analysis protocols used, individual variability in acoustic and aerodynamic parameters, and the degree of severity of the dysphonia [1, 5, 25–31]. Further, many instrumental measures can only be made from voice samples, such as sustained vowels and syllable strings (e.g., transglottal air pressure measurements, time-based perturbation measures). It is not clear whether measures made from such samples can be generalised to a client's everyday speech [32, 33]. In addition, several studies have demonstrated that there are only low to moderate predictive relationships between instrumental measures and perceptual ratings of impaired voice quality [e.g., 5, 34, 35]. Acknowledging that part of the problem here lies in the reliability limitations associated with some perceptual measures, the lack of a strong association between so-called objective features and the sound of a client's voice quality is surely problematic from a validity perspective. While this high degree of uncertainty around instrumental voice assessments remains, it is not surprising, then, that many clinicians hold tight to perceptual approaches to voice evaluation.

Finally, the popularity of auditory-perceptual evaluation of disordered voice is also likely to derive, at least in part, from pragmatic advantages including cost, time required for the voice evaluation, patient comfort, and requirements for technical knowledge and skill. Perceptual assessment is associated with minimal cost. Apart from a good quality microphone and audio-recording device, no specialised instrumentation is required [2, 6]. Perceptual evaluation is also quick, unobtrusive and comfortable for the client [2, 36]. Further, although training in perceptual voice evaluation has been shown to improve the reliability of listener judgements [e.g., 7, 19], extensive technical training and sophisticated technical knowledge and skill are not required. By no means do all speech-language pathology or vocology programs in universities around the world provide strong technical education, and not all provide students with access to state of the art instrumentation [37]. Similarly, many clinical facilities do not have such equipment for instrumental voice evaluation or the technical support to ensure that the technology is used optimally [1]. It is not surprising, then, that auditory-perceptual methods of voice evaluation are more popular among practising clinicians than instrumental approaches.

The Cons of Auditory-Perceptual Evaluation of Disordered Voice Quality

Despite the many pros of perceptual evaluation and the concerning limitations of instrumental assessment, the question remains: can auditory-perceptual assessment really be considered a gold standard for evaluation of voice quality? Perceptual measures can only be accepted as the gold standard if they can be shown to be valid and reliable. Auditory-perceptual evaluation has been criticised for decades on the basis of its inherent subjectivity. There is no doubt that perceptual ratings of voice are subjective, but subjectivity alone is not a sufficient reason for rejecting this approach to voice assessment. Subjectivity does not necessarily mean that reliability and validity are inadequate. If we could be confident that intra- and inter-rater reliability and agreement were high and if perceptual judgements showed clear relationships with underlying vocal tract physiology and acoustic parameters (i.e., if validity could be demonstrated), then clinicians' preference for auditory-perceptual evaluation could be justified.

As outlined in the previous section of this paper, there is a considerable amount of research evidence on the reliability of perceptual voice evaluation. While several studies that have controlled for a range of influences on listener judgement have demonstrated good to high reliability for overall severity of dysphonia and for a number of specific perceptual voice features such as breathiness and roughness [2, 16–20], some perceptual voice features such as vocal strain/hyperfunction are notorious for their association with poor listener reliability and agreement and several studies that incorporated little or no control over the key influences on perceptual judgement demonstrated only low to fair intra- and inter-judge reliability on most individual voice features [3, 6]. Similarly, the degree of listener agreement and reliability appears to be reduced when judges are asked to isolate specific perceptual features in voices with a complex mix of perceptual characteristics, particularly if those voices are mildly-moderately impaired [38].

The key proposition of Kreiman et al. [3] that listeners make perceptual judgements according to stored mental representations that serve as internal standards is fundamental to our understanding of reliability limitations of auditory-perceptual evaluation of disordered voice quality. These internal standards vary between listeners, and even within one listener over time and under different circumstances. If we accept this proposition, and there is good evidence that we should [3, 6, 38], then it is impor-

tant to understand the influences on those internal standards.

A valuable evidence base has developed to demonstrate some of the important influences on listener judgements that in turn act as sources of measurement error for perceptual evaluation. It is beyond the scope of this paper to critically review the large amount of research evidence available, but the most common conclusions are that these sources of measurement error include listener experience with clients with voice disorders, listener experience of perceptual voice evaluation, type and amount of training in auditory-perceptual assessment (e.g., using natural or synthetic anchors), type of ratings task (e.g., matching tasks vs. direct magnitude estimation), type of rating scale (e.g., equal-appearing interval vs. visual analogue scales), type of voice sample (e.g., sustained vowels vs. connected speech, loud vs. comfortable voice), listener fatigue, perceptual sensitivity of the listener and the particular perceptual voice features to be judged [6, 7, 9, 10, 19, 32, 38]. Will a clinician whose experience is mainly with voice disorders of neurological origin, for example, have the same internal standards as a clinician who works mainly with children with hyperfunctional voice disorders? Probably not. Will a listener who has been trained to use the complex Vocal Profile Analysis System [39] have the same internal standards as a clinician who has been trained to use the more straightforward GRBAS system [11]? Possibly not. Similar examples can be given for each of these sources of measurement error. It is not surprising then, that auditory-perceptual evaluation of impaired voices has been criticised in terms of its reliability.

The validity of auditory-perceptual evaluation is just as important as its reliability. How valid are perceptual ratings of dysphonia? How well do perceptual judgements correlate with instrumental measures of voice quality and how well do perceptual ratings inform the clinician about underlying vocal tract functioning? Again, a complete review of the evidence base associated with these questions is beyond the scope of this paper, but many papers that address the validity of auditory-perceptual evaluation are available [e.g., 3, 13, 17, 35, 40]. Most of these papers address the relationships between perceptual ratings and acoustic features. As discussed above, most previous studies have demonstrated that there are only low to moderate associations between instrumental measures and perceptual ratings of dysphonia [e.g., 5, 34, 35]. Further, the results of these studies have been inconsistent [8]. The extent to which these findings are due to the methodological complexities involved in the study of acoustic-perceptual relationships and/or to the limita-

tions inherent in acoustic and/or auditory-perceptual evaluations is not clear.

Unless a voice evaluation method provides information about vocal tract function, its value will be limited to monitoring and evaluating voice change over time and following intervention. While monitoring progress in itself is important for clinical practice, such information will not inform clinical decisions as to which intervention methods will be appropriate for particular patients with voice problems. It would therefore be of great value if perceptual assessment could inform the clinician about the abnormal patterns of vocal functioning that require modification in vocal rehabilitation. Unfortunately, there is far less research evidence about this aspect of the validity of auditory-perceptual evaluation than there is about perceptual-acoustic relationships in dysphonia [41, 42]. The paper of Hammarberg and Gauffin [13] from the early 1990s is one of the few studies that examined vocal tract physiology in specific abnormal voice qualities. More recently, Dejonckere and Lebacqz [43] have also investigated acoustic, perceptual, aerodynamic and anatomical relationships in dysphonia. Clinicians live with many inadequately tested assumptions about the physiology that gives rise to the sound of a client's voice. Ask almost any new graduate of a speech pathology program and he or she is likely to confidently assert that a breathy voice arises from incomplete glottal closure and/or the presence of a posterior glottal chink, that a rough voice arises from irregular vocal fold vibration patterns, and that a strained or pressed voice is due to excess laryngeal muscle tension. The face validity of these assumptions may be reasonable, well-founded in theory and supported to a degree by the findings of Hammarberg and Gauffin [13], but far more research is required before clinicians can be confident that all of their assertions about the relationships between what is perceived and what is actually occurring in the vocal tract are accurate [41]. Even if the reliability deficiencies of current approaches to auditory-perceptual evaluation can be redressed, perceptual evaluation will remain limited in its clinical application until we know more about its relationships to vocal tract function.

Unresolved Research Questions and Promising Directions

As might be anticipated given the foregoing discussion of the pros and cons of auditory-perceptual evaluation of disordered voice quality, many unresolved research ques-

tions remain about this type of voice assessment and its relationships with instrumental measures of voice. Until these questions are addressed, guidelines for clinical voice evaluation will remain incomplete. Nevertheless, there are several promising research directions that are likely to lead to greater certainty in the not too distant future.

Some critical unresolved questions concern the way in which listeners process the complex acoustic voice signal, whether that signal is normal or disordered. The fields of psychoacoustics and psychophysics have much to offer here, although most of this research has focussed on models for perception of normal voice and on pitch and loudness perception rather than on the way in which listeners process voice quality [see 44–47 for valuable reviews of perceptual scaling and measurement as well as current perspectives on human perception of sound and voice]. While there is considerable research evidence around the processing of pitch and loudness components of speech and voice, far more research is required in the voice quality domain and particularly disordered voice quality [but see 48 for expositions on the psychoacoustic measurement of normal voice quality, as well as 3, 6, 49 for seminal publications on the perception of pathological voice]. Despite the reasonably extensive literature around the psychoacoustics and psychophysics of voice, it remains unclear how models of perception of normal voice apply to disordered voice quality, whether or not a universal model for perception of abnormal voice quality can be found, which auditory-perceptual features of voice (or combinations of features) are most salient and important for clinical evaluation of disordered voice, what factors determine listener sensitivity to variations in perceptual features of disordered voice, and which perceptual voice features are the most sensitive to change during vocal rehabilitation.

The current research programs of Kreiman, Gerratt, Shrivastav, Sapienza and their colleagues are addressing these key questions, and have considerable potential to lead to improvements in clinical voice assessment [e.g., 9, 50, 51, 52]. These research teams use approaches that seek to integrate perceptual and acoustic features of disordered voice, rather than simply correlating traditional listener ratings with acoustic measures. For example, Kreiman and Gerratt [51] have undertaken experiments where listeners adjust levels of acoustic parameters in a speech synthesiser to match natural voice samples to synthesised samples. These experiments have demonstrated that listeners, both expert and naïve, respond consistently to changes in noise-to-signal ratio and that noise-to-signal

ratio is a reliable and significant determinant of voice quality. Shrivastav and Sapienza [52] have examined difference limens in adaptive listening tasks using synthesised voice stimuli. The focus of their research is on how listeners judge breathiness and on deriving algorithms that could eventually allow the prediction of the degree of perceived breathiness from the acoustic signal alone. This research is far from complete, but is encouraging and may mean that the reliability deficiencies of auditory-perceptual evaluation will no longer be a limitation of clinical voice assessment.

To return to more clinical aspects of auditory-perceptual assessment of dysphonia, some other unresolved research questions include the following:

- Which rating tasks, scale types and response formats are associated with optimum listener reliability and agreement?
- Which voice samples (i.e., vowel segments, connected speech etc.) maximise listener reliability and agreement and which result in the most representative judgements of our clients' everyday vocal functioning?
- Which listener training approaches are most effective and long-lasting? What is the relative effectiveness of training that uses analysis-by-synthesis methods, visual representations of the acoustic signal, and natural and synthetic referent samples or anchors? Which training methods have the best duration of effects and how often is recalibration of listeners required?
- How do auditory-perceptual judgements of voice quality interact with acoustic measures of voice? Which acoustic measures are the best and most robust predictors of listeners' perceptions? Do multiparametric acoustic indices or single acoustic parameters best reflect particular perceptual characteristics of disordered voice?
- To what extent can auditory-perceptual features of voice quality reflect underlying vocal tract functioning?

All but the last question listed above have received considerable research attention over the last decade [2, 5, 7, 10, 16, 19, 20, 32, 34, 37, 38], although further investigation is warranted in each area [5, 7]. The final question is far from resolved and requires further investigation. Admittedly, however, this is difficult to address, particularly because the field is still some way from being able to measure vocal tract physiology objectively and accurately. The advent of various methods for high-speed digital imaging of vocal fold vibration with quantitative analysis as well as the development of computer and biomechanical

cal modelling of vocal fold characteristics and behaviour, however, have good potential to shed new light on important physiological-perceptual relationships [53–58]. It should be remembered, however, that this final question may always be difficult to answer because perceptions of a particular voice quality can be associated with multiple acoustic patterns and multiple underlying physiological settings of the vocal tract [50, 59, 60]. Further, the likely non-linear relationships between human auditory perception, acoustic measures of voice quality and vocal tract configurations [60, 61] will also limit the extent to which perceptions of voice quality can inform us as to the underlying vocal tract physiology.

Some Recommendations for Clinical Voice Evaluation

While the clinical community awaits the findings of ongoing research, clinicians must continue to adopt best practice principles for the evaluation of their clients' voices. Although the preceding discussion has shown that auditory-perceptual analysis of disordered voice cannot be assigned the status of the gold standard for assessment of dysphonia, it is clear that perceptual evaluation will retain its important role in voice assessment of the foreseeable future. However, clinicians would be wise to adopt a principled approach to perceptual evaluation using the extensive knowledge we now have about the factors influencing listener perceptions and the reliability, consistency and validity of listener ratings. Selection of a perceptual evaluation scheme requires careful consideration of the underlying theoretical framework, the voice quality features included and the operational definitions provided for those features. Consideration of the type of rating scale, the protocol for voice sampling and recording, the availability of training and training resources, and associated reliability and validity data is also advisable. It is not the role of this author to recommend a specific auditory-perceptual evaluation system. None of the available perceptual assessment systems is perfect, but selection of a scheme that incorporates perceptual voice features that are commonly used and widely understood is a good start. Each of the perceptual evaluation systems listed earlier in this paper is a reasonable candidate (i.e., GRBAS [11], CAPE-V [12], Stockholm Voice Evaluation Approach [13], Perceptual Voice Profile [14]), but reviews of other perceptual evaluation schemes provide further information that can assist clinicians in making this important decision [e.g., 2, 20].

With the exception of the CAPE-V [12], most perceptual voice evaluation systems do not specify detailed protocols for voice sampling and recording, yet these are important considerations because of their potential influence on perceptual ratings. Again, it is not the role of this author to recommend a particular voice sampling and recording protocol, but use of a consistent sampling protocol for all clients incorporating a range of voice tasks including sustained vowels, production of standard sentences and free monologue is recommended. Recording of those voice samples in an environment with low ambient noise using a good quality microphone and digital recording device should be standard practice. Clinicians are also advised to ensure that their own fatigue and attention lapses do not interfere with their perceptual evaluations.

Because training in auditory-perceptual evaluation of disordered voice has been demonstrated to positively influence the reliability and consistency of perceptual judgements [7, 19, 62], voice practitioners should, wherever possible, participate in a well-designed training program and/or avail themselves of self-study training resources that are available [e.g., 63]. Although it is not yet clear which type of training is most effective and how much training is required to achieve and maintain consistency in perceptual judgement, there is sufficient evidence that training incorporating referent voice samples and/or anchors is likely to be valuable [7, 19]. It may also be that maintenance of auditory-perceptual rating skill and consistency requires periodic recalibration. Participation in regular perceptual rating sessions with peers and experts using a wide range of voice types and severities of dysphonia may reduce the extent to which drifts in internal standards occur, although this contention has yet to be the focus of systematic research.

Although the focus of this paper has been on auditory-perceptual evaluation of disordered voice quality, the preceding discussion of the limitations of perceptual evaluation suggests that dependence on clinician ratings of voice alone is far from ideal. Although auditory-perceptual assessment will continue to be a fundamental clinical tool, multiple methods of clinical voice evaluation incorporating both subjective and properly implemented instrumental measures are recommended [5, 25, 64]. With the advent of downloadable, inexpensive and relatively robust software for voice analysis, clinicians have little reason to depend entirely on auditory-perceptual assessment of their clients with voice problems. Finally, because several research groups have demonstrated the added value of having clients evaluate their own

vocal impairment and voice-related activity, participation and well-being through tools such as the Vocal Handicap Index [65], the Voice-Related Quality of Life measure [66], the Voice Symptom Scale [67] and the

Voice Activity and Participation Profile [68], clinicians are also advised to include such tools to complement measures of vocal impairment in their voice evaluation protocols [20, 69].

References

- 1 Behrman A: Common practices of voice therapists in the evaluation of patients. *J Voice* 2005;19:454–469.
- 2 Webb AL, Carding PN, Deary IJ, Mackenzie K, Steen N, Wilson JA: The reliability of three perceptual evaluation scales for dysphonia. *Eur Arch Otorhinolaryngol* 2004;261:429–434.
- 3 Kreiman J, Gerratt BR, Kempster GB, Erma A, Berke GS: Perceptual evaluation of voice quality: review, tutorial and a framework for future research. *J Speech Hear Res* 1993;36:21–40.
- 4 De Bodt M, Van de Heyning PH, Wuyts F, Lambrechts L: The perceptual evaluation of voice disorders. *Acta Otorhinolaryngol Belg* 1996;50:283–291.
- 5 Ma EP, Yiu EM: Mutiparametric evaluation of dysphonic severity. *J Voice* 2006;20:380–390.
- 6 Kent RD: Hearing and believing: some limits to the auditory-perceptual assessment of speech and voice disorders. *Am J Speech Lang Pathol* 1996;5:7–23.
- 7 Eadie TL, Baylor CR: The effect of perceptual training on inexperienced listeners' judgements of dysphonic voice. *J Voice* 2006;20:527–544.
- 8 Shrivastav R: The use of an auditory model in predicting perceptual ratings of breathy voice quality. *J Voice* 2003;17:502–512.
- 9 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.
- 10 Wuyts FL, De Bodt MS, Van de Heyning PH: Is the reliability of a visual analog scale higher than an ordinal scale? An experiment with the GRBAS scale for the perceptual evaluation of dysphonia. *J Voice* 1999;13:508–517.
- 11 Hirano M: *Clinical Examination of Voice*. New York, Springer, 1981, pp 83–84.
- 12 American Speech-Language-Hearing Association: *Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V)*. Rockville Pike, American Speech-Language-Hearing Association, 2002.
- 13 Hammarberg B, Gauffin J: Perceptual and acoustic correlates of quality differences in pathological voices as related to physiological aspects; in Fujimura O, Hirano M: *Vocal Fold Physiology*. Voice Quality Control. San Diego, Singular, 1995, pp 283–303.
- 14 Oates J, Russell A: Learning voice analysis using an interactive multi-media package: development and preliminary evaluation. *J Voice* 1998;12:500–512.
- 15 Fex S: Perceptual evaluation. *J Voice* 1992;6:155–158.
- 16 De Bodt MS, Wuyts FL, Van de Heyning PH, Croux C: Test-retest study of the GRBAS scale: influence of experience and professional background on perceptual ratings of voice quality. *J Voice* 1997;11:74–80.
- 17 Hammarberg B, Fritzell B, Gauffin J, Sundberg J, Wedin L: Perceptual and acoustic correlates of abnormal voice qualities. *Acta Otorhinolaryngol* 1980;90:441–451.
- 18 Dejonckere PH, Obbens C, de Moor GM, Wieneke GH: Perceptual evaluation of dysphonia: reliability and relevance. *Folia Phoniatri* 1993;45:76–83.
- 19 Chan MK, Yiu EM: A comparison of two perceptual voice evaluation training programs for naïve listeners. *J Voice* 2006;20:229–241.
- 20 Karnell MP, Melton SD, Childes JM, Coleman TC, Dailey SA, Hoffman HT: Reliability of clinician-based (GRBAS and CAPE-V) and patient-based (V-RQOL and IPVI) documentation of voice disorders. *J Voice* 2007;21:576–590.
- 21 Portney LG, Watkins MP: *Foundations of Clinical Research*, ed 2. Upper Saddle River, Prentice Hall Health, 2000.
- 22 Hammarberg B: *Perceptual and acoustic analysis of dysphonia*; thesis. Karolinska Institute, Stockholm, 1986.
- 23 Hammarberg B, Fritzell B, Gauffin J, Sundberg J: Acoustic and perceptual analysis of vocal dysfunction. *J Phonet* 1986;14:533–547.
- 24 Hillenbrand J, Houde RA: Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. *J Speech Hear Res* 1996;39:311–321.
- 25 Hillman RE, Montgomery WW, Zeitels SM: Appropriate use of objective measures of vocal function in the multidisciplinary management of voice disorders. *Curr Opin Otorhinolaryngol Head Neck Surg* 1997;5:172–175.
- 26 Deliyiski DD, Shaw HS, Evans MK, Vesselinov R: Regression tree approach to studying factors influencing acoustic voice analysis. *Folia Phoniatri Logop* 2006;58:274–288.
- 27 Holmberg EB, Hillman RE, Perkell JS, Gress C: Relationships between intra-speaker variation in aerodynamic measures of voice production and variation in SPL across repeated recordings. *J Speech Hear Res* 1994;37:484–495.
- 28 Kreiman J, Gerratt BR, Antoñanzas-Barroso N: Measures of the glottal source spectrum. *J Speech Lang Hear Res* 2007;50:595–610.
- 29 Carson CP, Ingrisano DR, Eggleston KD: The effect of noise on computer-aided measures of voice: a comparison of CSpeechSP and the Multi-Dimensional Voice Program software using the CSL 4300B module and Multi-Speech for Windows. *J Voice* 2003;17:12–20.
- 30 Tize I: *Workshop on acoustic voice analysis*. Denver, National Center for Voice and Speech, 1995.
- 31 Holmberg EB, Doyle P, Perkell JS, Hammarberg B, Hillman RE: Aerodynamic and acoustic voice measurements of patients with vocal nodules: variation in baseline and changes across voice therapy. *J Voice* 2003;17:269–282.
- 32 Revis J, Giovanni A, Wuyts F, Triglia JM: Comparison of different voice samples for perceptual analysis. *Folia Phoniatri Logop* 1999;51:108–116.
- 33 Wolfe V, Cornell R, Fitch J: Sentence/vowel correlation in the evaluation of dysphonia. *J Voice* 1995;9:297–303.
- 34 Yiu EM: Limitations of perturbation measures in clinical acoustic voice analysis. *Asia Pacific J Speech Lang Hear* 1999;4:155–166.
- 35 Kempster GB, Kistler DJ, Hillenbrand J: Multidimensional scaling analysis of dysphonia in two speaker groups. *J Speech Hear Res* 1991;34:534–543.
- 36 Yamaguchi H, Shrivastav R, Andrews ML, Niimi S: A comparison of voice quality ratings made by Japanese and American listeners using the GRBAS scale. *Folia Phoniatri Logop* 2003;55:147–157.
- 37 Oates J, Bain B, Davis P, Chapman J, Kenny D: Development of an auditory-perceptual rating instrument for the operatic singing voice. *J Voice* 2006;20:71–81.
- 38 Kreiman J, Gerratt BR: Sources of listener disagreement in voice quality assessment. *J Acoust Soc Am* 2000;108:1867–1876.
- 39 Laver J, Wirz S, Mackenzie J, Hiller S: *A Perceptual Protocol for the Analysis of Vocal Profiles (work in progress)*. Edinburgh, University of Edinburgh, 1981, pp 265–280.

- 40 Wolfe V, Martin D: Acoustic correlates of dysphonia: type and severity. *J Commun Dis* 1997;30:403–416.
- 41 Oates J: The evidence base for the management of individuals with voice disorders; in Reilly S, Douglas J, Oates J (eds): *Evidence-Based Practice in Speech Pathology*. London, Whurr, 2004, pp 110–139.
- 42 Colton RH, Casper JK, Leonard R: *Understanding Voice Problems: A Physiological Basis for Diagnosis and Treatment*, ed 3. Baltimore, Lippincott Williams and Wilkins, 2006, pp 16–17.
- 43 Dejonckere PH, Lebacqz J: Acoustic, perceptual, aerodynamic and anatomical correlations in voice pathology. *ORL J Otorhinolaryngol Relat Spec* 1996;58:326–332.
- 44 Marks LE, Gescheider GA: Psychophysical scaling; in Wixted J (ed): *Steven's Handbook of Experimental Psychology*, ed 3. New York, John Wiley, 2002, pp 91–138.
- 45 Marks LE, Algorn D: Psychophysical scaling; in Birnbaum MH (ed): *Measurement, Judgment, and Decision Making*. San Diego, Academic Press, 1998, pp 81–178.
- 46 Houtsma AJM: Pitch perception; in Moore BCJ (ed): *Hearing*. San Diego, Academic Press, 1995, pp 267–295.
- 47 Zwicker E, Fastl H: *Psychoacoustics: Facts and Models*. Berlin, Springer, 1999.
- 48 Kent RD, Ball MJ: *Voice Quality Measurement*. San Diego, Singular, 2000.
- 49 Laver J: *The Phonetic Description of Voice Quality*. Cambridge, Cambridge University Press, 1980.
- 50 Gerratt BR, Kreiman J: Measuring vocal quality with speech synthesis. *J Acoust Soc Am* 2001;110:2560–2566.
- 51 Kreiman J, Gerratt BR: Perception of aperiodicity in pathological voice. *J Acoust Soc Am* 2005;117:2201–2211.
- 52 Shrivastav R, Sapienza CM: Some difference limens for the perception of breathiness. *J Acoust Soc Am* 2006;120:416–422.
- 53 Eysholdt U, Tigges M, Wittenberg T, Proschel U: Direct evaluation of high-speed recordings of vocal fold vibrations. *Folia Phoniatri Logoped* 1996;48:163–170.
- 54 Zhang Y, Jiang JJ: Chaotic vibrations of a vocal fold model with a unilateral polyp. *J Acoust Soc Am* 2004;115:1266–1269.
- 55 Yan Y, Damrose E, Bless D: Functional analysis of voice using simultaneous high-speed imaging and acoustic recordings. *J Voice* 2007;21:604–616.
- 56 Bonilha HS, Deliyski DD: Period and glottal width irregularities in vocally normal speakers. *J Voice* 2008;22:699–708.
- 57 Deliyski DD, Petrushev PP, Bonilha HS, Gerlach TT, Martin-Harris B, Hillman RE: Clinical implementation of laryngeal high-speed videoendoscopy: challenges and evolution. *Folia Phoniatri Logop* 2008;60:33–44.
- 58 Švec JG, Šram F, Schutte HK: Videokymography in voice disorders: what to look for? *Ann Otol Rhinol Laryngol* 2006;116:172–180.
- 59 Gorham-Rowan MM, Laures-Gore J: Acoustic-perceptual correlates of voice quality in elderly men and women. *J Commun Disord* 2006;39:171–184.
- 60 Shrivastav R, Sapienza CM: Objective measures of breathy voice quality obtained using an auditory model. *J Acoust Soc Am* 2003;113:2217–2224.
- 61 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.
- 62 Martin DP, Wolfe VI: Effects of perceptual training based upon synthesized voice signals. *Percept Motor Skills* 1996;83:1291–1298.
- 63 Oates J, Russell A: *A Sound Judgement*. Melbourne, La Trobe University, 2003.
- 64 Orlikoff RF, Dejonckere PH, Dembowski J, Fitch J, Gelfer MP, Gerratt BR, et al: The perceived role of voice perception in clinical practice. *Phonoscope* 1999;2:89–106.
- 65 Jacobson BH, Johnson A, Grywalski C, Silbergleit A, Jacobson G, Benninger MS, Newman CW: The Voice Handicap Index (VHI): development and validation. *Am J Speech Lang Pathol* 1997;6:66–70.
- 66 Hogikyan ND, Sethuraman G: Validation of an instrument to measure Voice-Related Quality of Life (V-RQOL). *J Voice* 1999;13:557–569.
- 67 Deary IJ, Wilson JA, Carding PN, Mackenzie K: VoiSS: a patient-derived voice symptom scale. *J Psychosom Res* 2003;54:483–489.
- 68 Ma EP, Yiu EM: Voice activity and participation profile: assessing the impact of voice disorders on daily activities. *J Speech Lang Hear Res* 2001;44:511–524.
- 69 Murry T, Medrado R, Hogikyan ND, Aviv JE: The relationship between ratings of voice quality and quality of life measures. *J Voice* 2004;18:183–192.