Introduction

Voice is the sound that the listener perceives when the adducted vocal folds are driven into vibration by the pulmonary air stream. The four most common approaches for clinically assessing the various aspects of voice function include auditory-perceptual assessment of voice quality, acoustic assessment of voiced sound production, aerodynamic assessment of subglottal air pressures and glottal air flow rates during voicing, and endoscopic imaging of vocal fold tissue vibration. This paper provides a review of recent advances in each of these four voice assessment modalities based on literature that has been published in the last 2 years.

Perceptual assessment

Speech-language pathologists are increasingly being encouraged to use the new consensus auditory-perceptual evaluation of voice inventory for auditory-perceptual assessment of voice quality, and recent studies have provided new insights into listener reliability issues that have plagued subjective perceptual judgments of voice quality.

Acoustic assessment: Progress is being made on the development of algorithms that are more robust for analyzing disordered voices, including the capability to extract voice quality-related measures from running speech segments.

Aerodynamic assessment: New devices for measuring phonation threshold air pressures and air flows have the potential to serve as sensitive indices of glottal phonatory conditions, and recent developments in aeroacoustic theory may provide new insights into laryngeal sound production mechanisms.

Endoscopic imaging: The increased light sensitivity of new ultra high-speed color digital video processors is enabling high-quality endoscopic imaging of vocal fold tissue motion at unprecedented image capture rates, which promises to provide new insights into the mechanisms of normal and disordered voice production.

Summary

Some of the recent research advances in voice function assessment could be more readily adopted into clinical practice, whereas others will require further development.

Keywords

acoustic voice analysis, aerodynamics of voice production, high-speed endoscopic imaging, perception of voice, voice quality assessment
straining, pitch, and loudness [1]. Although CAPE-V represents an important step toward more uniform procedures for the clinical assessment of voice quality, it was not designed to resolve all of the persistent reliability problems that become even more apparent when the exact agreement of subjective listener ratings is accurately examined (eg, use of statistical procedures that examine the exact agreement between judgments of the same voice sample) [2].

Exploring sources of listener disagreement using synthesized speech
There are ongoing efforts to better understand and account for the unreliability of auditory-perceptual voice quality ratings. In one approach aimed at modeling sources of listener variability, Kreiman et al. [3**] recently used copy-synthesized vowel samples of disordered voices in an attempt to identify and quantify sources of listener disagreement. The authors found that 84% of the variance in the extent to which listeners do or do not agree could be accounted for by four factors that can be controlled for by the experimental design: instability of internal memory standards for levels of a perceptual dimension, ability to isolate single dimensions in a complex context, scale resolution, and absolute magnitude of the attribute being measured. Although this work is providing extremely valuable insights into sources of listener disagreement, direct clinical application is currently limited by the substantial technical and practical challenges associated with acoustically analyzing (see Acoustic assessment section) and synthesizing the disordered voices being assessed.

Application of psychometric principles to improve listener agreement
In an alternative approach, Shrivastav et al. [4] have recently shown that listener variability can be minimized by applying psychometric principles when designing the listening task. Such principles include averaging the ratings from multiple presentations of the same stimulus, with the investigators showing that a minimum of five repetitions provide the best results for judging voice quality in sustained vowels. In addition, to account for scale resolution and edge effects related to the absolute magnitude of a perceptual attribute, a standardized value for each rating is computed. The averaging and standardization procedures attempt to minimize variability and response biases of individual listeners. Although this approach allows for the experimenter to obtain more reliable perceptual ratings of natural (versus synthesized) vowel stimuli, the approach has limited clinical application due to the impracticality of taking the time to have several clinicians independently rate multiple repetitions of the same voice samples.

Acoustic assessment
The validity and reliability of acoustic measures currently used in the clinic to objectively assess voice quality (e.g., jitter, shimmer, and noise-to-harmonics ratio) are inherently limited by a reliance on the accurate determination of fundamental frequency ($F_0$), and these measures have been further restricted to the analysis of sustained vowels. $F_0$ can be difficult or impossible to extract in disordered voices, and sustained vowels may not be representative of vocal function or voice quality during continuous speech [5**]. In response to these limitations, there have been recent attempts to develop acoustic measures of voice quality that do not rely on accurate $F_0$ estimation and can be extended to the analysis of continuous speech. Two basic directions are prominent in recent attempts to develop acoustic measures that overcome the limitations of those currently being used in clinical voice assessment.

Cepstral-based methods
One set of approaches is based on cepstral analysis (a spectral-type method), which is inherently attractive since the cepstrum can be computed for any segment of speech and not just for steady vowel-like sounds. Recent work has demonstrated that the cepstral peak prominence (CPP) correlates well with perceptual judgments of overall severity of dysphonia [6], and research continues to understand and improve upon technical challenges of cepstral analysis methods [7,8*].

On the less positive side, the interpretation of cepstral-based measures relative to the underlying physiology of vocal fold vibration is not as intuitive as more traditional perturbation (e.g., jitter and shimmer) and noise (eg, noise-to-harmonics ratio) measures, which points to the need for studies that can better delineate relationships between cepstral measures and vocal fold function. There is also a need for more robust studies of how well cepstral-based measures correlate with perceived voice quality attributes to better assess the clinical potential of such measures (see Perceptual assessment section).

Nonlinear dynamics-based methods
The second set of recently described acoustic voice measures is based on nonlinear dynamics or chaos analysis [9*], which are much more robust with respect to analyzing atypical signals (e.g., aperiodic signals from pathological voices) than the measures currently being used for clinical voice assessment. In a proof-of-concept study, Zhang and Jiang [9*] recently demonstrated that nonlinear measures could better distinguish between normal and pathological voices than could the commonly-used measures of jitter and shimmer. Much more work needs to be done to understand how nonlinear measures of the acoustic signal relate to the underlying...
physiology of voice production and whether it will be possible to further develop such measures to differentially (and meaningfully) delineate varying levels of dysphonia severity.

**Aerodynamic assessment**

Since the early 1980s, clinical assessment of aerodynamic voice parameters has typically involved extracting estimates of average subglottal air pressures and glottal air flow rates from noninvasive measures of intraoral air pressures and oral air flow rates during the controlled (constant pitch and loudness) repetition of simple syllable strings. It was subsequently shown that important additional information about glottal phonatory status (including the presence of pathology) could be obtained from estimates of the phonation threshold pressure, the minimum air pressures required to initiate the softest possible voice production. Work using mathematical and physical laryngeal models has demonstrated that phonation threshold pressure is sensitive to vocal fold thinning, viscous shear properties of the tissue, and vocal tract inertance [10].

**Mechanical devices for measuring phonation threshold air pressure and air flow**

In the standard method, subglottal air pressure estimates are based on intraoral air pressure measurements that are obtained during lip closure of the /p/ sound. This method works because air pressure equilibrates throughout the airway (subglottal pressure equals intraoral air pressure) when the vocal folds are abducted for /p/-sounds produced in strings of /p/-plus-vowel syllables (eg, /pa-pa-pa-pa/). Jiang and his colleagues [11] have raised concerns that the accuracy of subglottal air pressure estimates may be influenced by undesirable adjustments to respiratory forces and vocal tract shapes that untrained subjects can display during the procedure. For several years, Jiang’s group has reported on the development of approaches to overcome these behavioral influences using mechanical devices that interrupt the oral air flow at unpredictable times during sustained vowel production to permit air pressure estimates. More recently, technological modifications to the airflow interruption technique have resulted in systems that adopt partial flow interruption to minimize effects related to abrupt cessation and an airflow redirection tube to minimize the impact of laryngeal reflexes [12*]. Jiang and Tao [13*] have also recently demonstrated in a mathematical model that the air flow rate at phonation threshold varies systematically with changes in simulated vocal fold tissue properties, vocal tract loading, and glottal area configurations. These initial results provide evidence that such a simple air flow-based measure may have some clinical utility, particularly given the relative ease of measuring oral air flow versus air pressure.

**Application of aeroacoustics to voice production**

There has been an apparent resurgence of interest in applying more sophisticated aeroacoustic theories and approaches to the study of laryngeal sound production. These studies indicate that other aerodynamic phenomena (e.g., vortical flow and vortex shedding), beyond what is portrayed in classic source-filter descriptions of voice production, could be contributing significantly to the sound that is produced during phonation [14*,15]. This information may ultimately have clinical significance because these higher order aeroacoustic phenomena could play an even greater role in mechanisms of disordered voice production than in normal phonatory functions.

**Endoscopic imaging**

There is an ongoing interest in exploring the use of high-speed imaging to supplement or replace stroboscopy in the endoscopic assessment of vocal fold vibration. This is because stroboscopy only provides a highly averaged view of the vibration pattern and is not capable of resolving detailed tissue motion within individual vibratory cycles. Digital video camera systems have recently become available with adequate light sensitivity and recording speeds to capture 4000 to 10000 high-resolution color images per second through a transoral endoscope [16**], a substantial improvement over previous high-speed systems that produced lower quality grayscale images at slower capture rates. The new higher speed systems provide adequate imaging for examining higher pitched phonation (i.e., sampling a sufficient number of images per vibratory cycle) and facilitate direct correlations with recordings from other voice measurement devices that capture signals at comparable sampling rates. For example, the capability to accurately synchronize and compare vocal fold vibration captured at 10000 images per second with the simultaneously recorded acoustic (microphone) signal that is also sampled at 10000 Hz promises to provide new insights into relationships between vocal fold tissue motion and sound production (e.g., relationships between asymmetries in tissue motion and perturbations in the acoustic signal). High-speed digital imaging can also facilitate and further enhance videokymographic assessment of vocal fold vibration, allowing detailed imaging of a glottal axis perpendicular to the midline to better examine the symmetry of vibrations between the left and right vocal folds at a chosen location. Judgments of asymmetry and disordered classification schemes based on kymographic images have been recently advocated by Švec and his colleagues [17*].

The recent surge of interest in developing high-speed endoscopic imaging of the vocal folds has spawned two major approaches for analyzing the massive amount of imaging data that is generated by these systems. One
approach focuses more on facilitating visual inspection of important parameters in the native images (mucosal wave, symmetry of vibration, etc.). The other approach utilizes higher order quantitative methods (e.g., Nyquist plots) to characterize features, such as glottal area variation, that are extracted from the images but not directly related back to image visualizations.

**Methods to facilitate visualization of high-speed images**

Deliyski and his colleagues [16] have focused their efforts on developing software to facilitate the efficient identification and visual examination of important phonatory events (voice onset, phonatory instabilities, etc.) with additional capabilities to quantify selected vocal parameters and features (open quotient, symmetry of vocal fold vibration, etc.). Examples of two such displays that facilitate visualization of the mucosal wave — mucosal wave playback and mucosal wave kymography — are shown in Fig. 1. These new image analysis tools have recently been studied, and investigations have begun to reveal that even some degree of asymmetry during vocal fold vibration can be tolerated in normal-sounding voices [18].

**Quantitative approaches to assessing high-speed vocal fold sequences**

Yan and colleagues [19,20] have used Nyquist plots, nonlinear processing of both the glottal area extracted from high-speed imaging and from the acoustic signal, to discriminate normal versus pathological voices. The quantification of glottal variations at a high temporal resolution has also laid the ground work for more sophisticated computer models that can simulate asymmetric vocal fold tissue motion observed in excised larynx preparations [21] and live human subjects [22,23]. Developing more accurate and robust models of vocal fold function will not only provide additional insights into the biomechanics of normal and disordered voice production, but also the potential to assist in planning phonosurgery by predicting the impact of a planned procedure on vocal function.

**Conclusion**

Within the last 2 years, there have been published reports of new developments in auditory-perceptual, acoustic, aerodynamic and endoscopic imaging approaches for assessing voice quality and function. Some of the advances in perceptual (CAPE-V), acoustic (cepstral-based methods), and aerodynamic (phonation threshold pressure) assessment seem to have a better chance of being more rapidly adopted in the clinic, while others will require further development and testing.

**References and recommended reading**

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest

Additional references related to this topic can also be found in the Current World Literature section in this issue (p. 295).


Zhang Y, Jiang JJ. Acoustic analyses of sustained and running voices from patients with laryngeal pathologies. J Voice 2008; 22:1–9. This article presents initial results on the use of nonlinear dynamics-based algorithms on running speech to better differentiate normal from disordered voices versus traditional perturbation analysis.


Baggott CD, Yuen AK, Hoffman MR, et al. Estimating subglottal pressure via airflow redirection. Laryngoscope 2007; 117:1491–1495. This study introduces a system for estimating subglottal pressure that may be easily transferred to a clinical setting.

Jiang JJ, Tao C. The minimum glottal airflow to initiate vocal fold oscillation. J Acoust Soc Am 2007; 121:2873–2881. This article introduces the phonation threshold flow parameter and demonstrates the parameter’s sensitivity to simulated vocal fold properties using a mathematical model.

Howe MS, McGowan RS. Sound generated by aerodynamic sources near a deformable body, with application to voiced speech. J Fluid Mech 2007; 592:397–402. This article summarizes the aerodynamic theory of sound production and presents a detailed analysis of different source mechanisms that contribute to voice.


Deliyski DD, Petrushev PP, Bonilha HS, et al. Clinical implementation of laryngeal high-speed videostroscopy: challenges and evolution. Folia Phoniat Logop 2008; 60:33–44. This article comprehensively presents the latest results of the utility of high-speed imaging of the vocal folds, as well as new facilitative visualizations that promise to offer clinically-salient information.


Shaw HS, Deliyski DD. Mucosal wave: a normophonic study across visualization techniques. J Voice 2008; 22:23–33. The goal of this study is to determine the variations in image-based judgments of the mucosal wave using stroboscopic and high-speed methods.


