AJSLP

Tutorial

Recommended Protocols for Instrumental Assessment of Voice: American Speech-Language-Hearing Association Expert Panel to Develop a Protocol for Instrumental Assessment of Vocal Function

Rita R. Patel,[a] Shaheen N. Awan,[b] Julie Barkmeier-Kraemer,[c] Mark Courey,[d] Dimitar Deliyski,[e] Tanya Eadie,[f] Diane Paul,[g] Jan G. Švec,[h] and Robert Hillman[i]

Purpose: The aim of this study was to recommend protocols for instrumental assessment of voice production in the areas of laryngeal endoscopic imaging, acoustic analyses, and aerodynamic procedures, which will (a) improve the evidence for voice assessment measures, (b) enable valid comparisons of assessment results within and across clients and facilities, and (c) facilitate the evaluation of treatment efficacy.
Method: Existing evidence was combined with expert consensus in areas with a lack of evidence. In addition, a survey of clinicians and a peer review of an initial version of the protocol via VoiceServe and the American Speech-Language-Hearing Association's Special Interest Group 3 (Voice and Voice Disorders) Community were used to create the recommendations for the final protocols.
Results: The protocols include recommendations regarding technical specifications for data acquisition, voice and speech tasks, analysis methods, and reporting of results for instrumental evaluation of voice production in the areas of laryngeal endoscopic imaging, acoustics, and aerodynamics.
Conclusion: The recommended protocols for instrumental assessment of voice using laryngeal endoscopic imaging, acoustic, and aerodynamic methods will enable clinicians and researchers to collect a uniform set of valid and reliable measures that can be compared across assessments, clients, and facilities.

[a]Department of Speech and Hearing Sciences, Indiana University, Bloomington
[b]Department of Audiology and Speech-Language Pathology, Bloomsburg University of Pennsylvania
[c]Division of Otolaryngology–Head and Neck Surgery, University of Utah, Salt Lake City
[d]Otolaryngology, The Mount Sinai Hospital, New York Eye and Ear Infirmary of Mount Sinai
[e]Department of Communicative Sciences and Disorders, Michigan State University, East Lansing
[f]Department of Speech and Hearing Sciences, University of Washington, Seattle
[g]Director, Clinical Issues in Speech-Language Pathology, American Speech-Language-Hearing Association, Rockville, MD
[h]Department of Biophysics, Faculty of Science, Palacký University, Olomouc, Czech Republic
[i]Massachusetts General Hospital, Harvard Medical School, MGH Institute of Health Professions, Boston
Correspondence to Rita R. Patel: patelrir@indiana.edu

Editor-in-Chief: Krista Wilkinson
Editor: Jeannette Hoit
Received January 20, 2017
Revision received July 17, 2017
Accepted February 17, 2018
https://doi.org/10.1044/2018_AJSLP-17-0009

Voice plays a crucial role in human communication and function. Voice production is multidimensional, involving physiologic, biomechanical, and aerodynamic mechanisms that produce an acoustic output that is perceived by the auditory system. When evaluating clients with voice disorders, it is preferable, whenever possible, to characterize the impact of the disorder(s) on all of the pertinent mechanisms/dimensions by obtaining complete case histories and performing the following battery of assessments: auditory–perceptual, laryngeal endoscopic imaging, acoustic, aerodynamic, and clients' self-perception of the impact of the voice disorders on their daily function (Behrman, 2005; Hillman, Montgomery, & Zeitels, 1997; Hirano, 1989; Roy et al., 2013). Although these types of assessments

Disclosure: *Shaheen N. Awan is a consultant to KayPENTAX/Pentax Medical (Montvale, NJ) for the development of commercial acoustic analysis computer software and is the licensee of computer algorithms (including cepstral analysis of continuous speech algorithms) used in the Analysis of Dysphonia in Speech and Voice (ADSV) program. Diane Paul is an employee of the American Speech-Language-Hearing Association. Jan G. Švec is the inventor of videokymography. All other co-authors have declared that no competing interests existed at the time of publication.*

This work is licensed under a Creative Commons Attribution 4.0 International License.
Downloaded from: https://pubs.asha.org Katherine Hoener on 10/07/2021, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

are performed on a regular basis at many research and clinical facilities in the United States, a lack of standardized procedures/protocols currently limits the extent to which the results can be used to facilitate comparisons across clinics and research studies to improve the evidence base for the management of voice disorders. Although it is true that practice guidelines by both the American Speech-Language-Hearing Association (ASHA) and the American Academy of Otolaryngology-Head and Neck Surgery recommend general approaches for evaluation of hoarseness (ASHA, 2004a; Schwartz et al., 2009), there continues to be a large variability in specific protocols used for evaluation of dysphonia including differences in data collection, measures, client tasks, and so forth. Such differences in evaluation procedures also are reflected in the research literature, making it difficult to compare outcomes and interpret results across studies, thus contributing to the difficulties in recommending evidence-based guidelines for voice assessment (Roy et al., 2013). A previous effort to provide a basic protocol for functional evaluation of individuals with voice disorders by the European Laryngological Society was specifically designed to address the aforementioned issues and allow relevant comparisons with the literature when presenting or publishing the results of voice treatment (Dejonckere et al., 2001). However, the European Laryngological Society's basic protocol does not provide sufficient technical and procedural details to ensure measurement consistency/repeatability.

For more than a decade, ASHA's Special Interest Group (SIG) 3 for Voice and Voice Disorders (originally Special Interest Division 3) has pursued the development of guidance for voice assessment. This effort began by focusing on the development of a standardized approach for the most universally used method, auditory–perceptual assessment, and produced the widely used "Consensus Auditory–Perceptual Evaluation of Voice" (CAPE-V), which was first rolled out in 2002 and subsequently revised in 2009 (Kempster, Gerratt, Verdolini Abbott, Barkmeier-Kraemer, & Hillman, 2009). Subsequently, a Working Group on Clinical Voice Assessment (composed of members of SIG 3 and ASHA Speech-Language Pathology Clinical Issues staff), in conjunction with the National Center for Evidence-Based Practice in Communication Disorders, conducted an evidence-based systematic review (EBSR) of the literature for clinical voice assessment procedures to develop guidelines on instrumental clinical voice assessment. The main conclusion of the EBSR was that the review "…did not produce sufficient evidence on which to recommend a comprehensive set of methods for a standard clinical voice evaluation" (Roy et al., 2013, p. 220). This was largely because methodological inconsistencies and a lack of unified standards restricted and/or precluded the ability to make valid comparisons of vocal function between facilities, clients, and repeated assessments of the same client. The authors also concluded and recommended that further efforts to improve the evidence base for voice assessment measures "…would be greatly assisted by first establishing a minimal set of recommended guidelines (perhaps via expert consensus)…" (Roy et al., 2013, p. 220), which would include basic technical specifications and protocols for instrumental assessment methods.

To follow up on the recommendations stemming from the EBSR, in 2012, ASHA approved creation of the Expert Panel to Develop a Protocol for Instrumental Assessment of Vocal Function (IVAP) in collaboration with SIG 3 to develop a core set of recommended protocols for the most commonly used instrumental voice assessment methods. It was assumed that this should optimally include, in order of importance, laryngeal endoscopic imaging, acoustic analysis, and aerodynamic assessment, in addition to other noninstrumental parts of the evaluation (e.g., perceptual assessment using the CAPE-V and self-report instruments). Use of all three instrumental approaches is deemed preferable because, together, they more fully characterize the fundamental components of voice production (Hillman et al., 1997; Mehta & Hillman, 2008). The proposed protocol is designed to assist both clinicians and researchers. This article presents the recommendations that were developed by the ASHA IVAP expert panel for instrumental evaluation of voice. The recommendations include not only specifications for voice/speech tasks and data analysis/measures but also specifications for data acquisition (technical instrumental specifications and examination procedures). Adoption of these basic recommendations is expected to improve the evidence base of instrumental voice measures for evaluation and treatment; enable valid comparisons of assessment results within and across clients, studies, and facilities; and facilitate the evaluation of treatment efficacy and effectiveness. Such uniform assessment protocols are expected to greatly facilitate valid meta-analyses of future treatment studies and the eventual development of evidence-based clinical practice guidelines (ASHA, 2004a) for the assessment and treatment of voice disorders. The end result would be improved quality care for individuals with these disorders (Schwartz et al., 2009).

The recommended protocols are meant to produce a core set of well-defined measures using instrumental approaches that can be universally interpreted and compared. It is not the intent of these recommendations to preclude the use of additional measures or protocols that individual clinics/clinicians or researchers deem useful in assessing vocal function.

Method

During 2012, on the basis of nominations from ASHA's SIG 3, ASHA established an expert panel consisting of recognized experts in voice disorders (speech-language pathology and otolaryngology/laryngology) and voice/speech science to develop basic recommended protocols for instrumental assessment of vocal function, including laryngeal endoscopic imaging, acoustics, and aerodynamics. The initial charge of the expert panel was for 3 years but needed to be extended for an additional year. In developing recommendations, the expert panel reviewed protocols from various sources: textbooks, peer-reviewed

Downloaded from: https://pubs.asha.org Katherine Hoener on 10/07/2021, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

publications, non–peer-reviewed publications, and materials received from requests for protocols posted on ASHA SIG 3 Community and VoiceServe 2012–2013. In areas where evidence was not available, expert consensus among the expert panel members was reached through multiple discussions carried out via phone conferences and face-to-face meetings. It is important to note that this type of expert consensus is commonly used in medical specialties to establish a starting point for developing standards of care when there is insufficient scientific evidence (Fink, Kosecoff, Chassin, & Brook, 1984). The initial draft of the protocol was first presented at the ASHA 2014 annual convention (Awan et al., 2014). Feedback received from this session was used to revise the protocol. A second revised draft of the protocol was circulated on the ASHA SIG 3 Community and VoiceServe for additional feedback on February 5, 2015. After 3 months of posting the document, the expert panel revised the protocols based on the received feedback. Fourteen voice experts provided detailed written feedback on the second draft of the recommendations. The revised protocol was subsequently circulated as hard copies at the ASHA 2015 annual convention at the SIG 3 Affiliates' meeting and Speech-Language Pathology Practice lounge, exhibit hall. Overall, the IVAP panel had seven consensus conference calls and three consensus face-to-face meetings from 2013 to 2015. The final goal was to disseminate the information via a publication in a peer-reviewed journal.

Results

The ASHA IVAP expert panel's recommended protocols include a core set of recommendations for laryngeal endoscopic imaging (Appendix A), acoustics (Appendix B), and aerodynamic (Appendix C) assessments covering (a) data acquisition, (b) voice and speech tasks, and (c) data analysis (standardized signal measurement methodology) to achieve uniformity in clinical and research reporting of voice evaluation outcomes. An effort was made to maintain consistency regarding the level of detail for laryngeal endoscopic imaging, acoustic analysis, and aerodynamic assessments. However, because of the inherent differences between approaches and variations in the level of evidence for each, there are some unavoidable variations in the degree of detail across the three instrumental assessment modalities.

Recommendations for Laryngeal Endoscopic Imaging

Office-based (outpatient) laryngeal endoscopic imaging should be used to obtain measures of structure and gross function as well as measures of vocal fold vibration in individuals with voice complaints. Various terms have been used to describe endoscopic imaging/examination of the larynx and vocal folds such as videolaryngoscopy (Ward, Hanson, Gerratt, & Berke, 1989), videostroboscopy (Woo, 1996), and strobovideolaryngoscopy (Sataloff, Spiegel, & Hawkshaw, 1991). In this publication, we use the terms videoendoscopy and videostroboscopy (Deliyski, Hillman, & Mehta, 2015). Videoendoscopy uses a constant light source to assess the structures and nonvibratory function of the larynx. Gross level visual–perceptual assessment is recommended to include inspection of the vocal fold medial edges, vocal fold mobility (e.g., abduction/adduction), supraglottic activity during phonation, and laryngeal maneuvers during transitional behaviors. Videostroboscopy relies on the strobe effect (usually by using a strobe light source) to assess the vibratory function of the vocal folds during phonation, considered critical for determining the nature, or causes of dysphonia (Eller et al., 2008; Mehta & Hillman, 2012; Mendelsohn, Remacle, Courey, Gerhard, & Postma, 2013; Paul et al., 2013). Stroboscopic visual–perceptual assessment is recommended to address the parameters of regularity, vocal fold vibratory amplitude, mucosal wave, vocal fold phase symmetry, vertical level, and glottal closure pattern (Bless, Hirano, & Feder, 1987). Nonvibratory function (e.g., vocal fold abduction/adduction) can also be observed during videostroboscopy.

Appropriate knowledge, skills, and training in this method of assessment are critical for speech-language pathologists performing the specific procedure(s) discussed in this document. Education and training for speech-language pathologists' performance of laryngeal videoendoscopic assessment are outlined in the Knowledge and Skills on Vocal Tract Visualization and Imaging (ASHA, 2004b), Vocal Tract Visualization and Imaging Technical Report (ASHA, 2004c), and ASHA Position Statement (ASHA, 2004d). Office-based endoscopy is within the scope of practice of speech-language pathologists involved in the management of individuals with voice disorders as defined in the ASHA's Position Statement on Vocal Tract Visualization and Imaging (ASHA, 2004d) and in the document entitled, "The Roles of Otolaryngologists and Speech-Language Pathologist in the Performance and Interpretation of Strobovideolaryngoscopy" (ASHA, 1998).

Data Acquisition

Laryngeal videoendoscopic and videostroboscopic examinations are performed by pairing an appropriate light source with a rigid 70° or 90° endoscope, a standard flexible fiberoptic endoscope, and/or a videoscope (flexible endoscope with a digital chip image sensor at the distal tip). Overall, the laryngeal videoendoscopic and videostroboscopic recordings should have adequate quality (e.g., image brightness and resolution) to make judgments of structural and vibratory ratings. Valid ratings of vocal fold vibration require at least three consecutive videostroboscopic glottal cycles (Hertegard, 2005). A videostroboscopic glottal cycle consists of opening, closing, and closed phases (Timcke, Von Leden, & Moore, 1958).

Downloaded from: https://pubs.asha.org Katherine Hoener on 10/07/2021, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

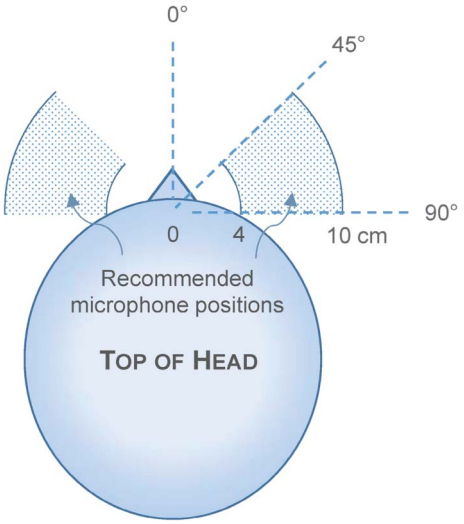
Technical Specifications

Videostroboscopic imaging. Stroboscopic sampling provides an optical illusion of slowing down the rapid vocal fold motion that the naked eye is unable to perceive. Numerous system specifications need to be met to create this apparent slow motion (Hillman & Mehta, 2010). The stroboscopic capture system should meet the following specifications: (a) The stroboscopic effect for videostroboscopic imaging should be provided either by controlling a flashing light source with constant image acquisition (currently the most common approach) or by controlling the timing of image exposure (similar to a camera shutter) coupled with a constant light source. (b) The recommended image integration time, which is either the duration of the strobe flash light impulse or the length of a single-frame image exposure time with constant light source, is 125 microseconds (μs) or less (Deliyski, Powell, Zacharias, Gerlach, & de Alarcon, 2015). The system must deliver only one light flash/exposure per captured video frame to maintain bright, sharp, and artifact-free images across the entire vocal frequency range, particularly at high vocal pitches. (c) The use of "fast" stroboscopy mode (1.5 Hz) is recommended for the typical recording. Optimally, a minimum of 16 images per cycle are required to adequately assess the vibratory motion (Deliyski, 2010). A range of 1–2 Hz for the stroboscopy mode is acceptable. (d) The stroboscopic system should be able to track frequency in the range of 60–1000 Hz using a contact microphone, electroglottographic, or other valid sensor.

Simultaneous acoustic recording. It is recommended that signals from the acoustic recording and laryngeal videoendoscopic/videostroboscopic recording be acquired simultaneously during the assessment protocol. The acoustic recording can be obtained by placing a small microphone on the camera or at a fixed distance of 4–10 cm, at an angle of 45°–90° away from the front of the mouth (Figure 1). It is recommended that the gain setting and microphone distance be documented for each recording and be kept the same for subsequent recordings of the same client for comparison purposes (e.g., pretreatment vs. posttreatment).

Image capture and playback. Overall, the laryngeal videoendoscopic and videostroboscopic recordings should have adequate quality to make judgments of structural and vibratory measures. In addition, the following should be considered: (a) The camera used for image capture should provide an image that is color balanced with a spatial resolution of at least 720 × 480 pixels and a signal-to-noise ratio (SNR) ≥ 42 dB (effective 7-bit dynamic range). To obtain such an adequate SNR, the camera should be paired with a light source and a laryngoscope of sufficient brightness. (b) The video recording and storage systems should be adequate to preserve the image quality produced by the imaging system. If data compression is needed, the use of standard compression that preserves the image quality and dynamics to render clinical judgments is recommended (e.g., MPEG 2). (c) The video monitor should be able to display image quality that is the same or better than what is produced by the camera and the recording system, and (d) the video playback should include capabilities for real-time playback, frame-by-frame video playback, and various slower frame rates for adequate judgments of laryngeal videoendoscopic and videostroboscopy parameters.

Figure 1. Schematic showing the recommended acceptable range of locations for the head-mounted microphone used to record the acoustic signal. The microphone should be positioned at a fixed distance of 4–10 cm from the lips at an angle of 45°–90° away from the front of the mouth.



Examination Procedures

Image orientation. Ideally, the client should be positioned or the endoscope should be placed so that the larynx is shown in the center of the image with the anterior commissure of the vocal folds pointing straight down during abduction and the artifact (e.g., size distortion) due to endoscope angle being minimized. The entire length of the vocal folds during phonation should be visible.

Use of topical anesthesia. The use of topical anesthesia is recommended when adequate image acquisition is compromised in a client who has an uncontrolled gag response or is sensitive to the placement of the endoscope in the oral or nasal cavity (Peppard & Bless, 1991). Clinicians must follow local (e.g., institutional and state), regional, and ASHA guidance when using topical anesthesia (ASHA, 2004b, 2004c). Some general recommendations when using a rigid endoscope include administering topical anesthetic to the pharyngeal wall, faucial arches, dorsum and base of the tongue, and/or the soft palate. For flexible endoscopes (fiberoptic and distal chip tip), some general recommendations include application of topical lidocaine to the nasal mucosa unilaterally. Decongestants like phenylephrine could be combined with lidocaine and administered in a spray form to the unilateral nasal passage (Burton, Altman, & Rosenfield, 2012; Hirano & Bless, 1993).

Downloaded from: https://pubs.asha.org Katherine Hoener on 10/07/2021, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Client/Subject Tasks

The following basic tasks are recommended for the evaluation of gross vocal fold motion and structure and of vibratory characteristics (Table 1).

Evaluation of Laryngeal Structure and Nonvibratory Motion

Rest breathing involves observing the laryngeal structures during quiet breathing. Either rigid or flexible endoscopes can be used to make observations of vocal fold position at rest and of anatomical structures. At a minimum, the rest breathing should involve three complete breath cycles (i.e., inhalation and exhalation).

Laryngeal diadochokinetic task involves producing /ʔiʔiʔiʔiʔi/ without breaths in between to evaluate vocal fold adduction and abduction integrity, rhythm, and pace using rapid production of glottal stops. Either rigid or flexible endoscopes can be used to make observations during this task.

Maximum range of adduction and abduction with sniffing. This task involves production of "/i/-sniff…/i/-sniff" or "/i/-quick inhale…/i/-quick inhale" at a self-selected rate to evaluate the extent of adduction and abduction of the vocal folds and the integrity of the muscles involved in these actions. The integrity of only the posterior cricoarytenoid muscle also could be evaluated by using only the "sniff" (quick inhalation) gesture. The use of a flexible endoscope is preferred for this task, although individuals may tolerate quickly inhaling between productions of /i/ when stroboscopic recordings are obtained using a rigid scope.

Evaluation of Vocal Fold Vibratory Characteristics

It is recommended that clients/subjects be instructed to produce sustained phonation of the vowel /i/ (Kitzing, 1985) to obtain a minimum of three consecutive videostroboscopic vibratory cycles. Maintenance of a constant pitch and loudness is recommended except when performing pitch and loudness glides. A minimum of one trial should be produced at each of the following levels: (a) phonation with stable normal (i.e., typical) pitch and loudness, (b) pitch variations (i.e., high pitch, low pitch), (c) loudness variations (i.e., loud voice, quiet voice), and, (d) when possible, variations in pitch and loudness that may better elucidate the client's problem (e.g., if the client presents with difficulty in achieving high-pitch phonation at reduced loudness levels, then attempting to capture an example of this would be highly warranted). During the examination, it is also critical to visually check both the fundamental frequency (f_o) indicator and the vocal fold images being displayed and recorded to ensure that the system is detecting a valid/stable f_o and that this coincides with stable/continuous stroboscopic imaging/tracking of vocal fold motion (Titze et al., 2015).

Table 1. Core tasks and measures for laryngeal imaging with valid regularity.

Tasks	Light source	
	Continuous light	Strobe light
Rest breathing • Three complete breath cycles (inhalation and exhalation)	Vocal fold edge	Vocal fold edge
Laryngeal diadochokinetic task /ʔiʔiʔiʔiʔi/	Gross-level vocal fold mobility	Gross-level vocal fold mobility
Maximum-range vocal fold adduction and abduction during alternated /i:/-sniff or /i:/-quick inhale	Vocal fold mobility maximum range	Vocal fold mobility maximum range
Sustained phonation of /i:/ at stable typical pitch and loudness • At least three consecutive glottal cycles	Supraglottic compression	• Supraglottic compression • Regularity • Amplitude • Mucosal wave • Left/right phase symmetry • Vertical level • Glottal closure pattern • Glottal closure duration
Sustained phonation of /i:/ at varied pitches (e.g., high, low pitch) • At least three consecutive glottal cycles for each pitch variation	Supraglottic compression	• Supraglottic compression • Regularity • Amplitude • Mucosal wave • Left/right phase symmetry • Vertical level • Glottal closure pattern • Glottal closure duration
Sustained phonation of /i:/ at varied loudness levels (e.g., loud voice, quiet voice production) • At least three consecutive glottal cycles for each loudness variation	Supraglottic compression	• Supraglottic compression • Regularity • Amplitude • Mucosal wave • Left/right phase symmetry • Vertical level • Glottal closure pattern • Glottal closure duration

Downloaded from: https://pubs.asha.org Katherine Hoener on 10/07/2021, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

The examiner must also ensure that a minimum of three consecutive videostroboscopic glottal cycles (opening, closing, and closed phases) are visible for typical, high-pitch, low-pitch, loud voice, and quiet voice recordings.

Data Analysis

It is recommended that both laryngeal videoendoscopic (structural/gross movement) and videostroboscopic (vibratory motion) measures of laryngeal function be obtained in individuals with voice complaints. Note that it may not be possible to obtain reliable and valid ratings of the laryngeal endoscopic ratings described below if recordings exhibit poor image quality (e.g., blurry image, view of the glottis partially obstructed by other structures). Similarly, valid laryngeal videostroboscopic recordings may not be possible in individuals with a moderate-to-severe dysphonia characterized by an absence of vocal fold vibration (aphonia) or the presence of irregular vocal fold vibration resulting in tracking errors (Olthoff, Woywod, & Kruse, 2007; Patel, Dailey, & Bless, 2008). In all of these instances, the examiner should indicate "could not judge (CNJ)" and note the reasons why the measures could not be rated (Poburka, Patel, & Bless, 2017). It is recommended that clinicians use slow playback rate and/or frame-by-frame analysis to perform the videoendoscopic/videostroboscopic recording visual perceptual ratings described below. Because of the visual perceptual nature of evaluating the various endoscopic and stroboscopic features, it is critical that the reliability of ratings is established. Clinically, this can be achieved by using the same rater for a given client or using a consensus approach for ratings. For research purposes, it is critical to report interrater and intrarater reliability when reporting research findings related to visual–perceptual analysis of laryngeal imaging. In addition, the experience level of the raters, the number of raters, and the statistical methods for interrater and intrarater reliability should be includedwhen reporting research findings on visual–perceptual ratings of laryngeal imaging (Poburka et al., 2017; Rosen, 2005; Woo, 1996). With improvement in technology, quantitative evaluation of vibratory function may also be available in the future for routine measurement of salient vocal fold movements (Just, Tyc, & Niebudek-Bogusz, 2015; Niebudek-Bogusz et al., 2017; Saadah, Galatsanos, Bless, & Ramos, 1998; Verikas, Uloza, Bacauskiene, Gelzinis, & Kelertas, 2009; Woo, 1996).

Evaluation of Laryngeal Structure and Nonvibratory Motion

Measures of laryngeal structure and nonvibratory motion can be obtained from either videoendoscopic or videostroboscopic recordings if appropriate tasks are performed (i.e., rest breathing, laryngeal diadochokinetic task [i.e., repetition of /ʔiʔiʔiʔiʔi/], and maximum range of adduction and abduction [i.e., sniffing (or quick inhalation) alternated with phonation of /i/]).

Vocal fold edges. This involves rating the appearance of the free edges of the membranous vocal folds (i.e., anterior commissure to the vocal process) when the vocal folds are in the abducted (rest breathing) position (Bless et al., 1987). The vocal fold edges should be rated as smooth and straight, bowed, irregular, or rough. Ratings of the vocal fold edges should be separately reported for each of the left and right vocal folds.

Vocal fold mobility. This refers to the movement of each of the vocal folds toward and away from the midline when producing the laryngeal adduction/abduction and maximum adduction–abduction tasks. Vocal fold mobility should be rated as normal, reduced, or absent. It is recommended that vocal fold mobility be separately reported for the left and right vocal folds.

Supraglottic activity. This refers to the degree of compression of the supraglottic structures during sustained phonation. When present, the supraglottic compression can be classified as predominantly unilateral or bilateral medial compression, predominantly anteroposterior compression, or sphincteric compression. The degree of the type of compression can be rated as mild, moderate, or severe.

Evaluation of Vocal Fold Vibratory Characteristics

These measures cannot be obtained from videoendoscopic recordings and need to be obtained from videostroboscopic recordings. Three consecutive videostroboscopic cycles from the middle of a stable phonation should be viewed as a basis for rating the following parameters. At minimum, ratings for these parameters should be reported for typical/normal phonation (typical pitch and loudness) and any other voice production tasks that help to elucidate the client's problem.

Regularity. This addresses the reliability of the stroboscopic tracking of glottal cycles and reflects the degree to which one videostroboscopic glottal cycle is consistent with successive videostroboscopic glottal cycles in terms of period and phase. On the basis of the extent to which the videostroboscopic system can reliably track the frequency and phase of vocal fold vibration, regularity can be rated as regular strobe tracking, intermittent strobe tracking, or irregular strobe tracking. If strobe tracking is intermittent or irregular, further estimates/ratings of vocal fold vibratory measures are not valid. Newer evaluation techniques, such as laryngeal high-speed videoendoscopy (Deliyski & Hillman, 2010; Patel et al., 2008) and videokymography (Švec & Schutte, 1996; Švec & Šram, 2011; Švec, Šram, & Schutte, 2009), may be used in instances where strobe tracking is intermittent or irregular to obtain further estimates of vocal fold vibratory function.

Amplitude. This refers to the extent of lateral movement of the vibrating portion of the vocal fold in the medial plane during phonation. Ratings provide an estimate of the typical medial-to-lateral excursion (amplitude) of the midmembranous portion of the fold during phonation from 0% to 100% using 25% increments, where 100% corresponds to the total visible width of the vocal folds.

Downloaded from: https://pubs.asha.org Katherine Hoener on 10/07/2021, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Amplitude should be separately reported for the left and right vocal folds.

Mucosal wave. This refers to the independent lateral movement of the mucosa over the body of the vocal fold (Bless et al., 1987). The mucosal wave originates on the medial surface of the vocal folds during the closure portion of the glottal cycle. It becomes visible during laryngeal videostroboscopy as it travels laterally across the superior surface of the vocal folds from the medial edge. The mucosal wave extent should be rated as the observation of mucosal wave movement from the medial edge toward the lateral surface of the vocal fold in increments of 25%, ranging from 0% to 100%, where 100% refers to the total visible width of the vocal fold. Mucosal wave should be separately reported for the left and right vocal folds.

Left/right phase symmetry. This is the degree to which the vocal folds appear as mirror images of each other during an apparent (videostroboscopic) glottal cycle in the timing of opening, closing, and maximum lateral–medial excursion (Bless et al., 1987; Bonilha, Deliyski, & Gerlach, 2008). Phase symmetry between the vocal folds is rated as symmetric or nearly symmetric, intermittently asymmetric, and consistently asymmetric.

Vertical level. This refers to the level difference in the vertical plane between the two vocal folds during the maximum closed phase of an apparent glottal cycle. The relative vertical level of the two vocal folds is rated as being similar or different. In the presence of a level difference between the vocal folds, report which of the vocal folds is above or below the other vocal fold.

Glottal closure pattern. This refers to the glottal configuration during maximum closure. The glottal configuration during maximum closure (i.e., closed phase of the glottal cycle) should be classified as one of the following: (a) complete closure, which occurs when there is no gap evident on maximal closure; (b) anterior gap, which occurs when closure is accomplished in the posterior part of the larynx but a gap remains at some point in the anterior third; (c) irregular closure, which occurs when the degree of closure varies along the length of the vocal folds—in some places, closure may be complete, whereas in other places, a gap may be observed, and the glottal space will not appear as a straight line but exhibits an irregular contour; (d) spindle-shaped gap, which occurs when there is a gap along the membranous portion of the vocal folds with vocal fold approximations at the vocal processes and near the anterior commissure; (e) posterior gap, which occurs when closure is accomplished along the anterior and mid-membranous portions of the vocal folds but a gap remains at the posterior glottis—if present, the posterior gap can be of two types: cartilaginous gap only or cartilaginous gap extending into the membranous portion; (f) hourglass gap, in which configuration occurs when closure is accomplished somewhere along the membranous portion of the vocal folds but the gaps are seen both anteriorly and posteriorly to the point of closure; (g) absence of closure, in which a lack of glottal closure exists between the vocal fold along the entire length of the vocal folds including the cartilaginous portion and the membranous portion during maximal approximation; and (h) variable closure, that is, when more than one glottal closure pattern is observed within an examination, the pattern should be rated as variable and the predominant closure pattern should be identified (Bless et al., 1987).

Glottal closure duration. This refers to the relative portion of each apparent glottal cycle that the glottis is closed. The closure duration is rated as closed phase missing, open phase predominant, closed phase predominant, or approximately equal.

Recommendations for Acoustic Assessment

Acoustic measures of vocal function are generally viewed as quantitative noninvasive metrics that (a) are sensitive to the severity of disturbances in voice production; (b) are often reported as being related to the perceptual parameters of vocal loudness, pitch, and quality; and (c) can provide indirect inferences regarding the underlying pathophysiology of voice disorders. Although there is an extensive literature describing a plethora of acoustic voice measures (Buder, 2000), there have been only limited attempts to reach consensus on the proper use (Titze, 1995) and reporting (Titze et al., 2015) of some of these measures, and even these efforts have not had a clear impact on the development of evidence-based protocols for voice assessment.

The following recommendations take into account the long-standing goal of having quantitative acoustic measures that associate with the auditory–perceptual parameters of vocal loudness (sound pressure level [SPL]), pitch (f_o), and quality (signal periodicity and/or spectral-based measures). The core set of parameters recommended for measuring vocal sound level includes habitual vocal SPL (decibels) and minimum and maximum vocal SPLs (decibels). Mean vocal f_o (hertz), vocal f_o standard deviation (hertz), and minimum and maximum vocal f_o (hertz) are the basic parameters recommended for measuring vocal frequency. For measuring the overall level of noise in the vocal signal, the recommendation is to use a measure of the vocal cepstral peak prominence (CPP; in decibels; Heman-Ackah, Michael, & Goding, 2002; Heman-Ackah et al., 2014; Hillenbrand, Cleveland, & Erickson, 1994; Noll, 1964). The most contentious discussions in both the panel and involving feedback from the field were related to the choice of an acoustic measure/correlate of voice quality. After much discussion, CPP (a measure of the relative amplitude of the CPP) was chosen as a general measure of dysphonia that reflects the global relationship of periodic versus aperiodic energy in a signal.

Data Acquisition

It is recommended that the acoustic signal of the client/subject productions of voice and speech tasks be recorded with a system that uses a calibrated head-mounted

Downloaded from: https://pubs.asha.org Katherine Hoener on 10/07/2021, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

microphone, a microphone preamplifier, and conversion to a digital format for storage and analysis.

Technical Specifications

Microphone. Ideally, recordings should be made with a head-mounted omnidirectional microphone that is positioned at a distance of 4–10 cm from the lips at an angle of 45°–90° away from the front of the mouth (Švec & Granqvist, 2010; Winholtz & Titze, 1997b; Figure 1). An omnidirectional microphone has the same sensitivity regardless of the direction of the sound source (i.e., it receives signals from all directions). Headset microphones are strongly recommended because they provide improved SNRs (due to the short distance from the lips) and maintain a consistent mouth-to-microphone distance. A unidirectional microphone (a microphone that picks up sound with high gain from a single direction) can be used to further reduce the impact of environmental noise when necessary, but SPL-based measures and spectral measures could be somewhat compromised because of the proximity effect (Švec & Granqvist, 2010). The microphone should meet the following specifications: (a) There should be a flat frequency response (i.e., variation of less than 2 dB) across the frequency range between the lowest expected f_o of voice and the highest spectral component of interest (approximately 50–8000 Hz), (b) noise level should be at least 10 dB lower than the sound level of the quietest vocal sound (Šrámková, Granqvist, Herbst, & Švec, 2015), and (c) the upper limit of the dynamic range should be above the sound level of the loudest phonations (i.e., can record the loudest voice production without saturation/clipping; Švec & Granqvist, 2010). The tutorial by Švec and Granqvist (2010) is an excellent resource for understanding the basics related to selecting microphones for evaluation of voice.

Microphone preamplifier. A preamplifier is an electronic device that amplifies a weak signal. The preamplifier specifications should match or exceed those for the microphone. In addition, it should have (a) an input impedance at least as high as the minimum terminating impedance required by the microphone, (b) gain adjusted so that the levels of the loudest phonations remain slightly below the saturation/clipping level, (c) no equalizers or bass/treble knobs (to prevent modification of the sound spectrum), and (d) no use of automatic gain control or any other smart signal processing circuits, such as noise cancellation, which can modify the original microphone signal (Švec & Granqvist, 2010).

Digital recording. The analog-to-digital conversion that is required to record the microphone signal can be done using an internal high-quality computer sound card (e.g., for a desktop computer) or with an external analog-to-digital device (often combined with a microphone preamplifier), which connects to a computer via USB or some other port (e.g., for a laptop computer). It is preferable to have external versus internal computer hardware. Minimum specifications include a sampling rate of ≥ 44.1 kHz (International Electrotechnical Commission, 1999), a minimum resolution of 16 bits (24 bits preferred for an increased dynamic range), a noise level of at least 10 dB lower than the sound level of the quietest phonations, an adjustable gain to ensure that the levels of the loudest phonation remain slightly below the saturation/clipping level of the analog-to-digital converter (Švec & Granqvist, 2010), and an audio file format that has no compression or lossless compression format (e.g., a recommended format is .wav).

Examination Procedures

SPL calibration. The level of the recorded signal is affected by the distance of the microphone from the mouth (must be held constant) and recording system gain/amplification, including any set scaling/gain that is internal to the computer/software being used (SPL values produced by computer software programs are relative and not absolute measures). Thus, the entire recording system must be calibrated to enable measurement of absolute SPL values. This process includes the additional recommendation that all SPL measurements be related to the often cited standard distance of 30 cm from the mouth (Schutte & Seidner, 1983) and that this distance is specified along with measures of SPL (see examples below). A suggested approach entails simultaneously measuring SPL with a sound-level meter at 30 cm from a client's/subject's lips while he or she sustains /a:/ vowel. The SPL of the vowel calibration signal captured with the head-mounted microphone can then be made equal to that shown by the sound-level meter (Švec & Granqvist, 2018; Winholtz & Titze, 1997a). If, for example, the computer software reports 75 dB SPL for a 70-dB calibration signal (as measured using a sound level meter) at 30 cm, it is then appropriate to subtract 5 dB from the computer result for any future signal (assuming no change in recording methods and no clipping; Švec & Granqvist, 2018). In this case, the calibrated head-mounted microphone signal provides SPLs as if placed at a 30-cm distance.

Alternatively, a calibration method with the sound-level meter placed in direct proximity of the head-mounted microphone could be used to make the voice SPLs captured by the head-mounted microphone identical to those measured by the sound-level meter (Maryn & Zarowski, 2015). In this case, the measured SPLs can be recalculated for the standard distance using the "distance law" relationship, for example, $SPL@30\ cm = SPL@d - 20 \log(30/d)$, where d is the distance of the head-mounted microphone (in centimeters) from the center of the mouth. This method may be easier to implement in clinical practice but can introduce inaccuracies of a few decibels due to distance measurement uncertainty at the proximity of the mouth and due to the fact that the distance law (6-dB decrease of SPL per distance doubling) is not accurate for mouth-to-microphone distances comparable with the size of mouth opening. The inaccuracies decrease with increasing mouth-to-microphone distance; therefore, the microphone distance of 10 cm from the center of the mouth is preferred over shorter distances when noise conditions allow. For

Downloaded from: https://pubs.asha.org Katherine Hoener on 10/07/2021, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

SPL correction from 10 to 30 cm, it holds SPL@30 cm = SPL@10 cm − 9.5 dB (Švec & Granqvist, 2018).

It is recommended that C frequency weighting is used for the sound-level meter, both for calibrating the recording system and for measuring the noise level in the recording environment (see below). C-weighting is preferred for measures of vocal sound level because it (a) measures uniformly over the frequency range (up to approximately 10 kHz) and (b) will not discriminate against low frequencies such as those often found in the f_o of speech and most singing. When calibrated, the unweighted microphone/computer recording will normally show SPLs that are close to the C-weighted SPL of the sound-level meter when the microphone response is flat (Švec & Granqvist, 2010) and there is no direct current (DC) component in the recorded signal.

If circumstances require that different frequency weighting be used (e.g., A-weighting to reduce the influence of low-frequency environmental noise on measured SPL), then the results can only be compared with SPL measures gathered using the same weighting, including calibration and comparisons with normative data. The difference in SPL between the head-mounted microphone (as measured with the computer and software being used for data analysis) and the sound-level meter at a 30-cm distance can then again be used to convert SPL measures to sound levels at 30 cm. Any uncalibrated changes in system gain/amplification would require recalibration. Examples for reporting the SPL measurements in voice and speech are the following:

- Mouth-to-microphone calibration distance: 30 cm
- Equivalent SPL of speech: SPLeq@30 cm = 70 dB (C-weighted)
- Quietest sustained voice SPL, vowel [a:]: SPL@30 cm = 50 dB (C-weighted, 1-s time averaged)
- Background noise level: 25 dB (A-weighted), 38 dB (C-weighted)

More details on the SPL measurement and calibration can be found in a complementary tutorial article by Švec and Granqvist (2018).

Recording environment. The background (environmental) noise levels substantially influence the quality of the acoustic signal that is recorded. To ensure accurate recording of the acoustic signal for measurement purposes, the following specifications are recommended: (a) The ambient noise level should be at least 10 dB weaker than the level of the quietest phonations (optimally < 38 dB [C-weighted or unweighted] or alternatively < 25 dB [A-weighted] for measurements at a 30-cm distance or < 35 dBA and < 48 dBC for measurements with the omnidirectional head-mounted microphones in proximity of the mouth; Šrámková et al., 2015). Background noise levels should be recorded when the client is asked to be quiet for about 5 s to document these levels for signal quality verifications; (b) the SNR for vocal signal quality measurements should be ≥ 30 dB (ideal is > 42 dB; Deliyski, Shaw, & Evans, 2005). Special precautions should be taken to eliminate sources of nonstationary noise, such as talking inside or outside the room, having open windows, playing music, and moving elevators. If these recommendations cannot be met in a quiet ordinary room, access to a soundproof or sound-treated environment should be considered; (c) the reverberation should be kept to a minimum (e.g., avoid reflective/hard surfaces), and the reverberation radius (distance at which the room-reflected sound becomes stronger than the direct sound from the mouth [Everest, 2001; Howard & Angus, 2009; Kuttruff, 2000]) should be at least twice as far as the mouth-to-microphone distance.

Client/Subject Tasks

Below are the recommended tasks (Table 2) and client/subject instructions for acoustic analysis of voice:

> Sustained vowels: Sustain the vowel /a:/ at a habitual level (habitual pitch and loudness) holding pitch and loudness as constant as possible for 3–5 s on one comfortable breath. Repeat this task three times.
>
> Standard reading passage: Read a typed passage (adults: first paragraph of the Rainbow Passage [Fairbanks, 1960]; children who can read: "The Trip to the Zoo" [Fletcher, 1972]) at comfortable pitch and loudness.
>
> Loudness range: (a) Sustain the vowel /a:/ as quietly as possible for at least 2 s without whispering—do this three times. (b) Sustain the vowel /a:/ as loudly as possible for at least 2 s. It is recommended that this task be repeated three times.
>
> Pitch range: (a) Sustain the vowel /a:/ as high in pitch as possible (including falsetto/loft) for at least 2 s. Repeat this task three times. (b) Sustain the vowel /a:/ as low in pitch as possible (in the modal register without the inclusion of fry/pulse register) for at least 2 s. Repeat this task three times. Note that the highest and lowest pitches also may be obtained either by using a pitch glide or in a stepwise fashion (Zraick, Nelson, Montague, & Monoson, 2000).

Data Analysis

Acoustic analysis should be performed with software that has some level of validation (e.g., use of standard [documented] algorithms and/or formally validated via comparison with other commonly used programs).

Measures of Vocal Sound Level

These correlate with the auditory perception of vocal loudness and are measured as the SPL in decibels at a specified distance from the mouth.

Habitual vocal SPL (decibels). This refers to the typical sound level of the voice during connected speech as the mean (time-averaged SPL [also known as equivalent continuous sound level]; American National Standards Institute, 1985; International Electrotechnical Commission, 2002). When a sound-level meter is used rather than a computer analysis, the most frequently observed SPL on

Downloaded from: https://pubs.asha.org Katherine Hoener on 10/07/2021, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Table 2. Core tasks and measures for acoustic analysis.

Tasks	Acoustic measures
Sustained vowel for 3- to 5-s duration • /a:/	• Cepstral peak prominence (CPP_{vowel})
Standard reading passage • Rainbow Passage (adults) • "The Trip to the Zoo" passage (children)	• Mean vocal frequency (Hz) • Habitual vocal SPL (dB) • Vocal frequency standard deviation (Hz) • Cepstral peak prominence (CPP_{speech})
Loudness range • Loudness glide on the vowel /a/, sustaining the loudest and quietest sounds for 1 s	• Maximum vocal SPL (dB) • Minimum vocal SPL (dB)
Pitch range • Pitch glide on the vowel /a/, sustaining the highest and lowest pitches for 1 s	• Maximum vocal frequency (Hz) • Minimum vocal frequency (Hz)

the meter (modal SPL) can also be used for determining the habitual SPL. In this case, the slow time weighting is recommended to be set on the sound-level meter. The SPL measures are extracted from the reading passage to control for potential phonemic effects/variations that might be observed in spontaneous speech tasks. When possible, it is recommended that these measures be based on an analysis of the entire reading passage. If this is not possible, a consistently selected subsegment (e.g., second and third sentences) of the Rainbow Passage that is at least 5 s long can be analyzed to obtain estimates of the measures.

Minimum and maximum vocal SPLs (decibels). These refer to SPL values for the quietest and loudest sustainable phonations. These measures also can be used to calculate the maximum range for vocal SPL (decibels). SPL is extracted as an average (equivalent level) across a 1-s segment that encompasses the lowest or highest SPL values (depending on the task being performed) for each of the three vowels produced for each task. It is recommended that only the single lowest and single highest values of the three trials for each task are reported and used to calculate the maximum SPL (decibels) range.

Measures of Vocal Frequency

These are correlated with the auditory perception of vocal pitch and measured as f_o in cycles per second or hertz. The f_o generally appears as the lowest harmonic frequency in the voice signal that spectrally presents itself as the frequency spacing between the harmonics.

Mean vocal f_o (hertz). This refers to the average of the estimates of the f_o for an acoustic signal recorded during connected speech, provided that all these estimates are obtained from windows (i.e., time frames) of the same duration covering the entire acoustic signal. An alternative interpretation is the total number of fundamental periods in the acoustic signal divided by the sum of those fundamental periods in the units of seconds. These measures are extracted from the reading passage (to control for potential phonemic effects/variations in spontaneous speech). When possible, it is recommended that these measures be based on an analysis of the entire reading passage. If this is not possible, a consistently selected subsegment (e.g., second and third sentences) of the Rainbow Passage that is at least 5 s long can be analyzed to obtain estimates of the measures (Zraick, Birdwell, & Smith-Olinde, 2005).

Vocal f_o standard deviation (hertz). This refers to the standard deviation (i.e., average variation) of the estimates of the f_o for an acoustic signal recorded during connected speech, provided that all these estimates are obtained from windows (i.e., time frames) of the same duration covering the entire acoustic signal. Like the mean vocal f_o (hertz), these measures are also extracted from the reading passage. When possible, it is recommended that these measures be based on an analysis of the entire reading passage. If this is not possible, a consistently selected subsegment (e.g., second and third sentences) of the Rainbow Passage that is at least 5 s long can be analyzed to obtain estimates of the measures (Zraick, Wendel, & Smith-Olinde, 2005).

Minimum and maximum vocal f_o (hertz). These refer to f_o values for the lowest-pitched (in modal register) and highest-pitched (including falsetto/loft register) sustainable phonations. These measures also can be used to calculate the phonational range for vocal f_o in semitones. The f_o is extracted as an average across a 1-s segment that encompasses the lowest or highest f_o values (depending on the task being performed) for each of the three /a/ vowels produced for each task. It is recommended that only the single lowest and single highest values of the three trials for each task are reported and used to calculate the phonational f_o (semitones) range.

Measures of Noise in the Vocal Signal

These refer to measures that are correlated with the auditory perception of voice quality and are based on estimating levels of periodic and/or aperiodic energy in the voice acoustic signal during sustained vowels and/or connected speech. A cepstral-based measure is recommended based on growing evidence that such measures are viable for analyzing the entire range of dysphonia severity in sustained vowels and connected speech (Maryn, Roy, De Bodt, Van Cauwenberge, & Corthals, 2009). This is an advantage over some more traditional measures (e.g., jitter and shimmer), which are only valid for mild-to-moderate dysphonia and for relatively long-duration

Downloaded from: https://pubs.asha.org Katherine Hoener on 10/07/2021, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

sustained vowel contexts in which the client is attempting a relatively steady pitch and loudness production (Awan, Roy, & Dromey, 2009; Zhang & Jiang, 2008). In addition, cepstral measures have recently become available in readily available software programs for clinical use (Watts, Awan, & Maryn, 2017).

Vocal CPP (decibels). This refers to a measure of the relative amplitude of the peak in the cepstrum (computed via a Fourier transform of the power spectrum of the voice signal) that represents the dominant rahmonic of the voice acoustic signal (in normal and Type I voices, the first harmonic/vocal f_o; Noll, 1964). CPP measures should be extracted from both sustained vowel and connected speech sample and clearly labeled as such, for example, CPP_{vowel} or CPP_{speech} (Awan, Roy, Jette, Meltzner, & Hillman, 2010). For vowels, CPP is extracted from a minimum of 1 s taken from the steadiest portion (most constant waveform amplitude) of the middle of each of the three /a:/ vowel productions. The final CPP value is averaged across the three /a:/ vowel productions. CPP measures for connected speech should be extracted from a consistently selected subsegment of a reading passage (e.g., The Rainbow Passage). CPP_{speech} measures could also be obtained from the CAPE-V sentences (Kempster et al., 2009; especially all voiced sentences) that are typically recorded for the auditory–perceptual assessment of voice quality.

Recommendations for Aerodynamic Assessment

Aerodynamic measures are designed to obtain noninvasive estimates of basic glottal aerodynamic parameters (i.e., both respiratory and laryngeal systems) that are required to produce phonation. The recommended measures include average glottal airflow rate (estimated from oral airflow rate during vowel production of /pi:pi:pi:pi:pi/ production) and average subglottal air pressure (estimated from intraoral air pressure during stop consonant production), acquired simultaneously with estimates of mean acoustic vocal SPL and f_o (Schutte, 1986).

Data Acquisition

The signals from airflow, air pressure, and microphone systems are acquired simultaneously during phonatory tasks. This can be done using a facemask that fits tightly over the nose and mouth to collect and direct the oral air stream through a pneumotachograph for measuring airflow and volumes. A small catheter is passed through a hole in the mask and positioned between the lips to measure intraoral air pressure during bilabial stop consonant production (Rothenberg, 1977; Smitheran & Hixon, 1981). The acoustic signal is picked up by a microphone that is appropriately positioned near the aerodynamic measurement system.

Technical Specifications

Airflow system. Pneumotachograph devices provide estimates of oral airflow by using a differential pressure transducer to measure the pressure difference across the flow resistance offered by a wire screen or mesh placed in the airstream; the pressure differential is calibrated to the flow. The estimates of air volumes are determined by integrating or cumulating for a specified time across flow. The pneumotachograph may be built into a tube that is attached to a facemask made of solid material, or the mask itself may have multiple holes (circumferentially vented) that are covered with mesh (Rothenberg, 1977). The following specifications for the airflow system are recommended: (a) a minimum bandwidth of DC to 75 Hz for average flow measures; (b) a maximum range of 0–2 L/s for voice production tasks; (c) for systems that provide specifications for several ranges of airflow, the accuracy should be within ±5 ml/s for 0.0–0.5 L/s, ±10 ml/s for 0.5–1.0 L/s, ±25 ml/s for 1.0–1.5 L/s, and ±65 ml/s for 1.5–2.0 L/s, whereas for systems that provide one specification, the accuracy should be within ±20 ml/s; and (d) the recommended noise (root-mean-square [rms]) is < 4 ml/s.

Air pressure system. Intraoral air pressure is usually measured directly with a pressure transducer attached to the oral catheter. The following specifications for the air pressure system are recommended: (a) a minimum bandwidth of DC to 60 Hz; (b) a maximum range of 0–75 cmH_2O; (c) based on specifications for two ranges of air pressures, the accuracy for 0–25 cmH_2O should be ±5 mmH_2O, and for 25–75 cmH_2O, it should be ±10 mmH_2O; and (d) the recommended noise is < 1 cmH_2O.

Microphone system. The microphone is positioned perpendicular to the air stream at the far end of the pneumotachograph tube or outside a circumferentially vented mask to reduce air pressure pulses/overpressures. Microphone characteristics should meet the requirements to capture a signal that is adequate to reliably extract measures of f_o and SPL (see section on "microphone" in the protocol for acoustic measures).

Examination Procedures

Calibration. The airflow system is calibrated for each client/subject to ensure accurate measurement. The most likely sources of variation in system performance and sensitivity are changes in the flow resistance of the wire screen or mesh of the pneumotachograph that result from repeated use and cleaning. Calibration is carried out using a known volume-flow source (e.g., a large syringe of known volume or another metered flow source) that is coupled in an airtight manner to the pneumotachograph.

Mask placement. The airflow mask is correctly positioned over the nose and mouth and pressed against the face to ensure that there are no air leaks between the rim of the airflow mask and the face. Even relatively small leaks can significantly reduce estimates of average airflow (underestimation).

Placement of the oral catheter. The oral catheter is correctly sized and positioned so that its open end is not occluded during lip closure for the /p/ sound (e.g., lips, tongue, or buildup of saliva blocking the end of the tube). Signs of obstruction are obvious in significantly reduced

Downloaded from: https://pubs.asha.org Katherine Hoener on 10/07/2021, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

(or almost absent) air pressure signals and air pressures that do not vary as expected (e.g., return to zero when there is no applied air pressure).

Environment. The environment must be quiet enough so that the system does not track f_o during nonphonatory segments and clearly tracks these parameters during phonatory segments (particularly f_o and SPL during quiet/minimal voice production). It should be quiet enough to record an acoustic signal that can be reliably analyzed for SPL (including quiet/minimal voice production) and f_o. In addition, any transient noise sources should be identified and avoided during data acquisition. Ideally, the SNR (signal-to-background noise) should be > 10 dB SPL.

Client/Subject Tasks

Clients/subjects are instructed to produce short utterances each composed of minimally five /p + vowel/ syllables at a rate approximating 1.5–2 syllables per second (Holmberg, Hillman, Perkell, Guiod, & Goldman, 1995; Holmberg, Perkell, & Hillman, 1984; Smitheran & Hixon, 1981). Faster rates of production also have been recommended in some circumstances and are acceptable if the criteria for the airflow and air pressure signals are met (see below). Syllabic rate may be controlled via the use of a metronome. The /i:/ vowel is recommended (e.g., pi:pi:pi:pi:pi) because its high tongue position is associated with a more consistent velopharyngeal (VP) closure, but other vowels may also be employed to satisfy the requirements of additional analysis methods (e.g., vowels with more neutral tongue positions are typically used when the intent is to extract additional measures from the airflow signal using inverse filtering [Holmberg, Hillman, & Perkell, 1988; Smitheran & Hixon, 1981]). Each string of at least five syllables should be produced on one exhalation, as a continuous utterance (i.e., legato, no pauses between syllables—like sustaining an /i:/ vowel with the /p/ sounds inserted; Plexico & Sandage, 2012; Plexico, Sandage, & Faver, 2011; Smitheran & Hixon, 1981) and holding pitch and loudness as constant as possible. There should be complete VP closure and no respiratory pumping or puffing of the cheeks, allowing the syllable strings to be produced as smoothly as possible. In cases of VP incompetence, occlusion of the nose (e.g., nose clip) may be considered to help attain valid measures. Full lip closure is necessary for each /p/. The syllable strings should be produced for a minimum of three times each at comfortable (typical or normal) loudness and raised loudness (as if to be heard across a room; approximately a 6-dB increase; Table 3). If the clients/subjects cannot produce five /pi/ syllables on one exhalation, the number can be reduced to four or three syllables per exhalation. Measures should still only be taken from the middle of each syllable string (avoiding the first and last syllables).

Because measures are taken from the three middle syllables of each syllable string and there are three syllable strings produced per voice condition, this results in nine measures per voice condition (3 measures per syllable string × 3 repetitions). The shorter syllable strings should be repeated enough times to yield measures from a minimum of nine middle syllables per voice production condition.

Signal Criteria

The airflow and air pressure signals should be visually monitored to ensure that the airflow signal attains a steady state during the /i:/ vowel productions (relatively flat horizontal line) and the air pressure signal attains a steady state during the /p/ stop consonant production (the peak pressures during lip closures should appear relatively flat on top while the airflow approximates zero; Figure 2). If clients/subjects cannot be trained to produce steady-state airflow and air pressure signals (relatively flat horizontal signals), then the assumptions underlying the indirect estimation of glottal aerodynamic parameters are not met and any measures that are extracted from these signals may not be valid (Lofqvist, Carlborg, & Kitzing, 1982; Rothenberg, 1977; Smitheran & Hixon, 1981).

Data Analysis

Measure of Average Glottal Airflow

Average glottal airflow rate (liters per second or milliliters per second). This measure is estimated from the oral airflow rate during vowel production. Measures should not be obtained from the first and last syllables. The middle three syllables from each string of five syllables are chosen for analysis. A measure of average glottal airflow rate is taken from the middle steady-state portion of each vowel (Holmberg, Hillman, & Perkell, 1989; Holmberg et al., 1988).

Measure of Subglottal Air Pressure

Average subglottal air pressure (centimeters of water or kilopascals). This measure is estimated from the intraoral air pressure produced during the repetition of stop consonants in syllable strings (Rothenberg, 1977; Smitheran & Hixon, 1981). An estimate of average air pressures is taken from the middle three syllables by linearly interpolating (essentially drawing a line connecting the peak pressures) between the peak air pressures produced during lip closures for adjacent consonant productions (i.e., at the same time points that the airflow measures are taken; Holmberg et al., 1989, 1988).

Measures of Mean Vocal SPL and f_o

These measures are extracted from the simultaneously recorded acoustic signal to facilitate the interpretation of airflow and air pressure measures. This SPL measure should not be interpreted as a standard SPL obtained without a mask because the mask may introduce considerable changes to the vocal signal. Estimates of average SPL and f_o are taken from the acoustic/microphone signal (or wideband airflow signal for a circumferentially vented mask system) at the same time points in each vowel that airflow measurements are obtained.

Calculation of Measures

Data values are averaged across syllable strings within each voice production condition (comfortable, loud, and

Downloaded from: https://pubs.asha.org Katherine Hoener on 10/07/2021, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Table 3. Core tasks and measures for aerodynamic analysis.

Tasks	Aerodynamic measures
• /pi:pi:pi:pi:pi/ at habitual pitch and loudness levels at ~1.5–2 syllables/s	• Average glottal airflow rate (L/s or ml/s) • Average interpolated air pressure (cmH_2O or kPa) • Mean vocal SPL (dB) and vocal frequency (Hz) during the task
• /pi:pi:pi:pi:pi/ at raised loudness levels (e.g., increased by 6 dB SPL) at ~1.5–2 syllables/s	• Average glottal airflow rate (L/s or ml/s) • Average interpolated air pressure (cmH_2O or kPa) • Mean vocal SPL (dB) and vocal frequency (Hz) during the task

Note. SPL = sound pressure level.

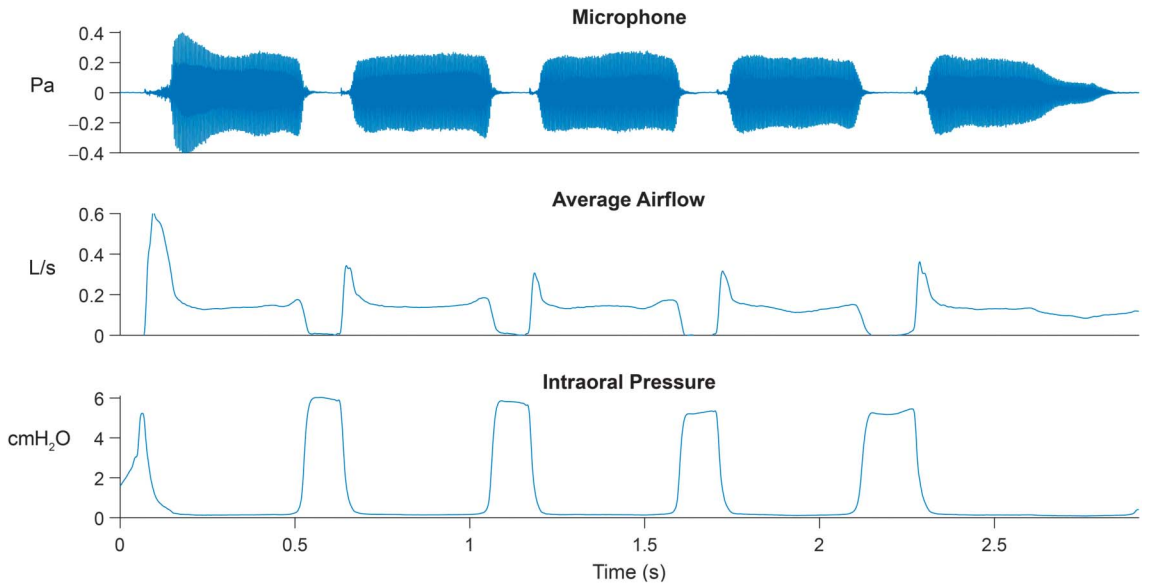
quiet), and the results are reported separately for comfortable, loud, and quiet voice productions as estimates of average glottal airflow rate (liters per second or milliliters per second), average subglottal air pressure (centimeters of water or kilopascals), average SPL (decibels), and average f_o (hertz). Units of measurement should be indicated clearly and used consistently within and across clients/subjects.

Discussion

Comprehensive evaluation of individuals with voice disorders entails obtaining a thorough case history and a battery of assessments including laryngeal imaging, acoustic measures, aerodynamic measures, auditory–perceptual evaluation, and patient self-report measures. This combination of assessments is designed to evaluate the impact of the voice disorder on the various subsystems of voice production as well as the impact of the voice disorder on an individual's daily function and quality of life. The product of a previous effort sponsored by ASHA Special Interest Division 3 (now SIG 3), the CAPE-V (Kempster et al., 2009), is now being widely used for clinical and research purposes, thereby increasing the validity of comparisons across clinics/clinicians and research studies and increasing the potential impact of future meta-analyses of the evidence base for the clinical management of voice disorders. ASHA SIG 3 also sponsored the current effort to develop core recommendations for instrumental voice assessments (laryngeal imaging, acoustics, and aerodynamics) with the similar intent to further improve the evidence base for assessing and treating voice disorders.

A combination of existing scientific evidence and expert consensus (supplemented with several cycles of review/feedback from the field) was used in developing these ASHA-IVAP recommended protocols for instrumental assessment of voice production using laryngeal endoscopic

Figure 2. Examples of acceptable low-pass filtered airflow and air pressure signals during production of a /pi:pi:pi:pi:pi/ syllable string for estimation of average airflow (milliliters per second) and average subglottal air pressure (centimeters of water). Note that the airflow signal attains a steady state during the /i:/ vowel productions (relatively flat horizontal line) and the air pressure signal attains a steady state during the /p/ stop consonant production (the peak pressures appear relatively flat on top). During production of the /p/ stop consonant, the airflow signal becomes 0 ml/s, ensuring a full bilabial seal during consonant production and a tight facial mask seal against the face without air leakage.



Downloaded from: https://pubs.asha.org Katherine Hoener on 10/07/2021, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

imaging, acoustic, and aerodynamic methods. As noted previously, this type of informal expert consensus is commonly used in medical specialties to establish a starting point for developing standards of care when there is insufficient scientific evidence (Fink et al., 1984). It is readily acknowledged, however, that this is not a perfect process. Although the public vetting of the recommendations had an impact and an attempt was made to balance the makeup of the expert panel (speech-language pathology, speech science, and otolaryngology/laryngology), the final result still reflects the interpersonal dynamics and biases of the panel. There are more formal approaches to consensus like the "Delphi method" (Dalkey & Helmer, 1963) that could have been employed to mitigate issues related to lack of anonymity of participants and lack of structure regarding the flow of information.

We also feel compelled to again emphasize that the recommended protocols are meant to produce a core set of well-defined measures using instrumental approaches that can be universally interpreted and compared. It is not the intent of these recommendations to preclude the use of additional measures or protocols that individual clinics/clinicians or researchers deem useful in assessing vocal function, including the use of noninstrumental methods. For example, there was some support for including additional aerodynamic measures such as laryngeal airway resistance (derived from air pressure and airflow) and phonation threshold pressure, which could be easily added to the core set of measures as deemed necessary by individual clinicians and researchers. The most contentious discussions in both the panel and involving feedback from the field were related to the choice of an acoustic measure/correlate of voice quality. After much discussion, CPP was chosen as a general measure of dysphonia that reflects the global relationship of periodic versus aperiodic energy in a signal. Potential insights into the different sources of aperiodicity that may affect the CPP may be provided by other acoustic measures including spectral-based measures (e.g., measures of spectral tilt) or (for sustained vowel contexts) more traditional measures (e.g., jitter, shimmer). Therefore, our current recommendation does not preclude the use of a variety of measures for the purpose of documenting vocal quality disturbances, provided that the core protocol measures are obtained and the user recognizes the limitations of the measures being used (e.g., the use of traditional perturbation measures) and follows appropriate procedures and precautions (e.g., use of traditional perturbation measures based on signal typing [Titze, 1995]).

The present recommendations do not include measurement norms. Although normative references are available from various sources such as textbooks and research publications in the areas of laryngeal imaging (Biever & Bless, 1989; Bless, Glaze, Lowery-Biever, Campos, & Peppard, 1993; Hirano & Bless, 1993; Woo, 1996), acoustic analysis (Awan et al., 2010; Baken & Orlikoff, 2007; Colton, Casper, & Leonard, 2011; Maturo et al., 2012), and aerodynamics (Weinrich, Brehm, Knudsen, McBride, & Hughes, 2013; Zraick, Smith-Olinde, & Shotts, 2012), it is often challenging for clinicians and researchers to determine which set of normative data to use due to variable descriptions of how the data were collected across the various studies. The recommended protocols could be used to systematically develop normative data from a reference population, against which the client/subject findings could be compared. Instrumental measures cannot be compared across different instrumentations and algorithms.

Conclusions

The recommended protocols for instrumental assessment of voice production using laryngeal endoscopic imaging, acoustic, and aerodynamic methods will enable clinicians and researchers to collect a uniform set of valid and reliable measures that can be compared across assessments, clients, and facilities. There is an ongoing need to expand the scientific evidence base for these measures and to potentially revise the recommended protocols as warranted by future changes in the evidence base.

Acknowledgments

This work was supported by the American Speech-Language-Hearing Association, Resolution No. BOD11-2012. In the Czech Republic, the participation of Jan Švec was supported by the Czech Science Foundation (GA CR) Project No. GA16-01246S. The authors thank Daryush Mehta for his assistance with figures and everyone who provided feedback during the various stages of the development of the protocol.

References

American National Standards Institute. (1985). *ANSI S1.4-1983. American National Standard: Specification for sound level meters*. Melville, NY: Author.

American Speech-Language-Hearing Association. (1998). *The roles of otolaryngologists and speech-language pathologists in the performance and interpretation of strobovideolaryngoscopy* [Relevant paper]. Retrieved from http://www.asha.org/policy

American Speech-Language-Hearing Association. (2004a). *Evidence-based practice in communication disorders: An introduction* [Technical report]. Retrieved from http://www.asha.org/policy

American Speech-Language-Hearing Association. (2004b). *Knowledge and skills for speech-language pathologists with respect to vocal tract visualization and imaging* [Knowledge and skills]. Retrieved from http://www.asha.org/policy

American Speech-Language-Hearing Association. (2004c). *Vocal tract visualization and imaging* [Technical report]. Retrieved from http://www.asha.org/policy

American Speech-Language-Hearing Association. (2004d). *Vocal tract visualization and imaging* [Position statement]. Retrieved from http://www.asha.org/policy

Awan, S., Barkmeier-Kraemer, J., Courey, M., Deliyski, D., Švec, J., Patel, R., . . . Paul, D. (2014). *Standard clinical protocols for endoscopic, acoustic and aerodynamic voice assessment: Recommendations from ASHA expert committee seminar*. Paper presented at the annual convention of the American Speech-Language-Hearing Association, Orlando, FL.

Downloaded from: https://pubs.asha.org Katherine Hoener on 10/07/2021, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Awan, S., Roy, N., & Dromey, C. (2009). Estimating dysphonia severity in continuous speech: Application of a multi-parameter spectral/cepstral model. *Clinical Linguistics & Phonetics, 23*(11), 825–841.

Awan, S., Roy, N., Jette, M. E., Meltzner, G. S., & Hillman, R. E. (2010). Quantifying dysphonia severity using a spectral/cepstral-based acoustic index: Comparisons with auditory-perceptual judgements from the CAPE-V. *Clinical Linguistics & Phonetics, 24*(9), 742–758.

Baken, R. J., & Orlikoff, R. F. (2007). *Clinical measurement of speech and voice* (2nd ed.). San Diego, CA: Singular.

Behrman, A. (2005). Common practices of voice therapists in the evaluation of patients. *Journal of Voice, 19*(3), 454–469.

Biever, D. M., & Bless, D. (1989). Vibratory characteristics of the vocal folds in young adult and geriatric women. *Journal of Voice, 3*(2), 120–131.

Bless, D., Glaze, L. E., Lowery-Biever, D., Campos, G., & Peppard, R. C. (1993). Stroboscopic, acoustic, aerodynamic, and perceptual analysis of voice production in normal speaking adults. *National Center for Voice and Speech Status and Progress Report, 4,* 121–134.

Bless, D., Hirano, M., & Feder, R. J. (1987). Videostroboscopic evaluation of the larynx. *Ear, Nose, & Throat Journal, 66*(7), 289–296.

Bonilha, H. S., Deliyski, D. D., & Gerlach, T. T. (2008). Phase asymmetries in normophonic speakers: Visual judgments and objective findings. *American Journal of Speech-Language Pathology, 17*(4), 367–376.

Buder, E. (2000). *Acoustic analysis of voice quality: A tabulation of algorithms 1902–1990*. San Diego, CA: Singular.

Burton, M. J., Altman, K. W., & Rosenfield, R. M. (2012). Topical anaesthetic or vasoconstrictor preparation for flexible fiber-optic nasal pharyngoscopy and laryngoscopy. *Otolaryngology–Head and Neck Surgery, 146,* 694–697.

Colton, R., Casper, J., & Leonard, R. (2011). *Understanding voice problems: A physiological perspective for diagnosis and treatment* (4th ed.). Baltimore, MD: Lippincott Williams & Wilkins.

Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management Science, 9*(3), 458–467.

Dejonckere, P. H., Bradley, P., Clemente, P., Cornut, G., Crevier-Buchman, L., Friedrich, G., . . . Woisard, V. (2001). A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. Guideline elaborated by the Committee on Phoniatrics of the European Laryngological Society (ELS). *European Archives of Oto-rhino-laryngology, 258*(2), 77–82.

Deliyski, D. (2010). Laryngeal high-speed videoendoscopy. In K. A. Kendall & R. J. Leonard (Eds.), *Laryngeal evaluation: Indirect laryngoscopy to high-speed digital imaging* (pp. 245–270). New York, NY: Thieme Medical.

Deliyski, D., & Hillman, R. E. (2010). State of the art laryngeal imaging: Research and clinical implications. *Current Opinion in Otolaryngology & Head and Neck Surgery, 18*(3), 147–152.

Deliyski, D., Hillman, R. E., & Mehta, D. D. (2015). Laryngeal high-speed videoendoscopy: Rationale and recommendation for accurate and consistent terminology. *Journal of Speech, Language, and Hearing Research, 58*(5), 1488–1492.

Deliyski, D., Powell, M. E. G., Zacharias, S. R. C., Gerlach, T. T., & de Alarcon, A. (2015). Experimental investigation on minimum frame rate requirements of high-speed videoendoscopy for clinical voice assessment. *Biomedical Signal Processing and Control, 17,* 21–28.

Deliyski, D., Shaw, H. S., & Evans, M. K. (2005). Influence of sampling rate on accuracy and reliability of acoustic voice analysis. *Logopedics, Phoniatrics, Vocology, 30*(2), 55–62.

Eller, R., Ginsburg, M., Lurie, D., Heman-Ackah, Y., Lyons, K., & Sataloff, R. (2008). Flexible laryngoscopy: A comparison of fiber optic and distal chip technologies. Part 1: vocal fold masses. *Journal of Voice, 22*(6), 746–750.

Everest, F. A. (2001). *Master handbook of acoustics*. New York, NY: McGraw-Hill.

Fairbanks, G. (1960). *Voice and articulation drillbook* (2nd ed.). New York, NY: Harper & Row.

Fink, A., Kosecoff, J., Chassin, M., & Brook, R. H. (1984). Consensus methods: Characteristics and guidelines for use. *American Journal of Public Health, 74*(9), 979–983.

Fletcher, S. G. (1972). Contingencies for bioelectric modification of nasality. *Journal of Speech and Hearing Disorders, 37,* 329–346.

Heman-Ackah, Y. D., Michael, D. D., & Goding, G. S., Jr. (2002). The relationship between cepstral peak prominence and selected parameters of dysphonia. *Journal of Voice, 16*(1), 20–27.

Heman-Ackah, Y. D., Sataloff, R. T., Laureyns, G., Lurie, D., Michael, D. D., Heuer, R., . . . Hillenbrand, J. (2014). Quantifying the cepstral peak prominence, a measure of dysphonia. *Journal of Voice, 28*(6), 783–788.

Hertegard, S. (2005). What have we learned about laryngeal physiology from high-speed digital videoendoscopy? *Current Opinion in Otolaryngology & Head and Neck Surgery, 13*(3), 152–156.

Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research, 37*(4), 769–778.

Hillman, R., & Mehta, D. (2010). The science of stroboscopic imaging. In K. A. Kendall & R. J. Leonard (Eds.), *Laryngeal evaluation: Indirect laryngoscopy to high-speed digital imaging* (pp. 101–209). New York, NY: Thieme Medical.

Hillman, R., Montgomery, W., & Zeitels, S. M. (1997). Appropriate use of objective measures of vocal function in the multidisciplinary management of voice disorders. *Current Diagnostic and Office Practice, 5,* 172–175.

Hirano, M. (1989). Objective evaluation of the human voice: Clinical aspects. *Folia Phoniatrica, 41,* 89–144.

Hirano, M., & Bless, D. (1993). *Videostroboscopic examination of the larynx*. San Diego, CA: Singular.

Holmberg, E. B., Hillman, R., & Perkell, J. S. (1989). Glottal airflow and transglottal air pressure measurements for male and female speakers in low, normal and high pitch. *Journal of Voice, 3*(4), 294–305.

Holmberg, E. B., Hillman, R. E., & Perkell, J. S. (1988). Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice. *The Journal of the Acoustical Society of America, 84*(2), 511–529.

Holmberg, E. B., Hillman, R. E., Perkell, J. S., Guiod, P. C., & Goldman, S. L. (1995). Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice. *Journal of Speech and Hearing Research, 38*(6), 1212–1223.

Holmberg, E. B., Perkell, J. S., & Hillman, R. (1984). Methods for using a noninvasive technique for estimating glottal functions from oral measurements. *The Journal of the Acoustical Society of America, 75*(S1:S7), 30.

Howard, D. M., & Angus, J. (2009). *Acoustics and psychoacoustics* (4th ed.). Oxford, United Kingdom: Focal.

International Electrotechnical Commission. (1999). *International standard IEC 60908: Audio recording—Compact disc digital audio system*. Geneva, Switzerland: Author.

Downloaded from: https://pubs.asha.org Katherine Hoener on 10/07/2021, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

International Electrotechnical Commission. (2002). *International standard IEC 61672-1: Electroacoustics—Sound level meters—Part 1: Specifications*. Geneva, Switzerland: Author.

Just, M., Tyc, M., & Niebudek-Bogusz, E. (Eds.). (2015). Objectivization of high-speed digital phonoscopic (HSDP) and laryngovideostroboscopic (LVS) vocal fold imaging. In K. Izdebski, Y. Yan, R. R. Ward, B. J. F. Wong, & R. M. Cruz (Eds.), *Normal and abnormal vocal folds kinematics: High-speed digital phonoscopy (HSDP), optical coherence tomography (OCT) & narrow band imaging (NBI®)* (Vol. I: Technology, pp. 185–203). San Francisco, CA: Pacific Voice and Speech Foundation (PVSF e-Q&A-b).

Kempster, G. B., Gerratt, B. R., Verdolini Abbott, K., Barkmeier-Kraemer, J., & Hillman, R. E. (2009). Consensus auditory–perceptual evaluation of voice: Development of a standardized clinical protocol. *American Journal of Speech-Language Pathology, 18*(2), 124–132.

Kitzing, P. (1985). Stroboscopy—A pertinent laryngological examination. *Journal of Otolaryngology, 14*(3), 151–157.

Kuttruff, H. (2000). *Room acoustics* (4th ed.). Oxford, United Kingdom: Spon.

Lofqvist, A., Carlborg, B., & Kitzing, P. (1982). Initial validation of an indirect measure of subglottal pressure during vowels. *The Journal of the Acoustical Society of America, 72*(2), 633–635.

Maryn, Y., Roy, N., De Bodt, M., Van Cauwenberge, P., & Corthals, P. (2009). Acoustic measurement of overall voice quality: A meta-analysis. *The Journal of the Acoustical Society of America, 126*(5), 2619–2634.

Maryn, Y., & Zarowski, A. (2015). Calibration of clinical audio recording and analysis systems for sound intensity measurement. *American Journal of Speech-Language Pathology, 24*(4), 608–618.

Maturo, S., Hill, C., Bunting, G., Ballif, C., Maurer, R., & Hartnick, C. (2012). Establishment of a normative pediatric acoustic database. *Archives of Otolaryngology–Head & Neck Surgery, 138*(No. 10), 956–961.

Mehta, D. D., & Hillman, R. E. (2008). Voice assessment: Updates on perceptual, acoustic, aerodynamic, and endoscopic imaging methods. *Current Opinion in Otolaryngology & Head and Neck Surgery, 16*(3), 211–215.

Mehta, D. D., & Hillman, R. E. (2012). Current role of stroboscopy in laryngeal imaging. *Current Opinion in Otolaryngology & Head and Neck Surgery, 20*(6), 429–436.

Mendelsohn, A. H., Remacle, M., Courey, M. S., Gerhard, F., & Postma, G. N. (2013). The diagnostic role of high-speed vocal fold vibratory imaging. *Journal of Voice*, *27*(5), 627–631.

Niebudek-Bogusz, E., Kopczynski, B., Strumillo, P., Morawska, J., Wiktorowicz, J., & Sliwinska-Kowalska, M. (2017). Quantitative assessment of videolaryngostroboscopic images in patients with glottic pathologies. *Logopedic, Phoniatrics, Vocology, 42*(2), 73–83.

Noll, A. M. (1964). Short-term spectrum and "cepstrum" techniques for vocal pitch detection. *The Journal of the Acoustical Society of America, 41,* 293–309.

Olthoff, A., Woywod, C., & Kruse, E. (2007). Stroboscopy versus high-speed glottography: A comparative study. *Laryngoscope, 117*(6), 1123–1126.

Patel, R., Dailey, S., & Bless, D. (2008). Comparison of high-speed digital imaging with stroboscopy for laryngeal imaging of glottal disorders. *Annals of Otology, Rhinology & Laryngology, 117*(6), 413–424.

Paul, B. C., Chen, S., Sridharan, S., Fang, Y., Amin, M. R., & Branski, R. C. (2013). Diagnostic accuracy of history, laryngoscopy, and stroboscopy. *Laryngoscope, 123*(1), 215–219.

Peppard, R. C., & Bless, D. (1991). The use of topical anesthesia in videostroboscopic examination of the larynx. *Journal of Voice, 5,* 57–63.

Plexico, L. W., & Sandage, M. J. (2012). Influence of vowel selection on determination of phonation threshold pressure. *Journal of Voice, 26*(5), 673.e7–673.e12.

Plexico, L. W., Sandage, M. J., & Faver, K. Y. (2011). Assessment of phonation threshold pressure: A critical review and clinical implications. *American Journal of Speech-Language Pathology, 20*(4), 348–366.

Poburka, B. J., Patel, R. R., & Bless, D. M. (2017). Voice-vibratory assessment with laryngeal imaging (VALI) form: Reliability of rating stroboscopy and high-speed videoendoscopy. *Journal of Voice, 31*(4), e1513–e1514.

Rosen, C. A. (2005). Stroboscopy as a research instrument: Development of a perceptual evaluation tool. *Laryngoscope, 115*(3), 423–428.

Rothenberg, M. (1977). Measurement of airflow in speech. *Journal of Speech and Hearing Research, 20*(1), 155–176.

Roy, N., Barkmeier-Kraemer, J., Eadie, T., Sivasankar, M. P., Mehta, D., Paul, D., & Hillman, R. (2013). Evidence-based clinical voice assessment: A systematic review. *American Journal of Speech-Language Pathology, 22,* 212–226.

Saadah, A. K., Galatsanos, N. P., Bless, D., & Ramos, C. A. (1998). Deformation analysis of the vocal folds from videostroboscopic image sequences of the larynx. *The Journal of the Acoustical Society of America, 103*(6), 3627–3641.

Sataloff, R. T., Spiegel, J. R., & Hawkshaw, M. J. (1991). Strobovideolaryngoscopy: Results and clinical value. *Annals of Otology, Rhinology & Laryngology, 100*(9, Pt. 1), 725–727.

Schutte, H. K. (1986). Aerodynamics of phonation. *Acta Oto-rhino-laryngolica Belgica, 40,* 344–357.

Schutte, H. K., & Seidner, W. (1983). Recommendation by the Union of European Phoniatricians (UEP): Standardizing voice area measurement/phonetography. *Folia Phoniatrica et Logopaedica, 35*(6), 286–288.

Schwartz, S. R., Cohen, S. M., Dailey, S. H., Rosenfeld, R. M., Deutsch, E. S., Gillespie, M. B., … Patel, M. M. (2009). Clinical practice guideline: Hoarseness (dysphonia). *Annals of Otology, Rhinology, and Laryngology, 141*(3, Suppl. 2), S1–S31.

Smitheran, J., & Hixon, T. J. (1981). A clinical method for estimating laryngeal airway resistance during vowel production. *Journal of Speech and Hearing Disorders, 46*(1), 138–146.

Šrámková, H., Granqvist, S., Herbst, C. T., & Švec, J. G. (2015). The softest sound levels of the human voice in normal subjects. *The Journal of the Acoustical Society of America, 137*(1), 407–418. https://doi.org/10.1121/1.4904538

Švec, J., & Granqvist, S. (2010). Guidelines for selecting microphones for human voice production research. *American Journal of Speech-Language Pathology, 19*(4), 356–368.

Švec, J., & Granqvist, S. (2018). Tutorial and guidelines on measurement of sound pressure level in voice and speech. *Journal of Speech, Language, and Hearing Research, 61,* 331–461.

Švec, J., & Schutte, H. K. (1996). Videokymography: High-speed line scanning of vocal fold vibration. *Journal of Voice, 10*(2), 201–205.

Švec, J., & Šram, F. (2011). Videokymographic examination of voice. In E. P. M. Ma & E. M. L. Yiu (Eds.), *Handbook of voice assessments* (pp. 129–146). San Diego, CA: Plural.

Švec, J., Šram, F., & Schutte, H. K. (2009). Videokymography. In M. P. Fried & A. Ferlito (Eds.), *The larynx* (3rd ed., Vol. 1). San Diego, CA: Plural.

Downloaded from: https://pubs.asha.org Katherine Hoener on 10/07/2021, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Timcke, R., Von Leden, H., & Moore, P. (1958). Laryngeal vibrations: Measurements of the glottic wave: I. The normal vibratory cycle. *AMA Archives of Otolaryngology, 68*(1), 1–19.

Titze, I. R. (1995). *Summary statement: Workshop on acoustic voice analysis*. National Center for Voice and Speech. Retrieved from http://ncvs.org/

Titze, I. R., Baken, R. J., Bozeman, K., Granqvist, S., Henrich, N., Herbst, C. T., . . . Wolfe, J. (2015). Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization. *The Journal of the Acoustical Society of America, 137*(5), 3005–3007.

Verikas, A., Uloza, V., Bacauskiene, M., Gelzinis, A., & Kelertas, E. (2009). Advances in laryngeal imaging. *European Archives of Oto-rhino-laryngology, 266*(10), 1509–1520.

Ward, P. H., Hanson, D. G., Gerratt, B. R., & Berke, G. S. (1989). Current and future horizons in laryngeal and voice research. *Annals of Otology, Rhinology & Laryngology, 98*(2), 145–152.

Watts, C. R., Awan, S. N., & Maryn, Y. (2017). A comparison of cepstral peak prominence measures from two acoustic analysis programs. *Journal of Voice*, *31*(3), 387.e1–387.e10.

Weinrich, B., Brehm, S. B., Knudsen, C., McBride, S., & Hughes, M. (2013). Pediatric normative data for the KayPENTAX phonatory aerodynamic system model 6600. *Journal of Voice, 27*(1), 46–56.

Winholtz, W. S., & Titze, I. R. (1997a). Conversion of a head-mounted microphone signal into calibrated SPL units. *Journal of Voice, 11*(4), 417–421.

Winholtz, W. S., & Titze, I. R. (1997b). Miniature head-mounted microphone for voice perturbation analysis. *Journal of Speech, Language, and Hearing Research, 40*(4), 894–899.

Woo, P. (1996). Quantification of videostrobolaryngoscopic findings—Measurements of the normal glottal cycle. *Laryngoscope, 106*(3, Pt. 2 Suppl. 79), 1–27.

Zhang, Y., & Jiang, J. J. (2008). Nonlinear dynamic mechanism of vocal tremor from voice analysis and model simulations. *Journal of Sound and Vibration, 316*(1–5), 248–262.

Zraick, R. I., Birdwell, K. Y., & Smith-Olinde, L. (2005). The effect of speaking sample duration on determination of habitual pitch. *Journal of Voice, 19*(2), 197–201.

Zraick, R. I., Nelson, J. L., Montague, J. C., & Monoson, P. K. (2000). The effect of task on determination of maximum phonational frequency range. *Journal of Voice, 14*(2), 154–160.

Zraick, R. I., Smith-Olinde, L., & Shotts, L. L. (2012). Adult normative data for the KayPENTAX Phonatory Aerodynamic System Model 6600. *Journal of Voice, 26*(2), 164–176.

Zraick, R. I., Wendel, K., & Smith-Olinde, L. (2005). The effect of speaking task on perceptual judgment of the severity of dysphonic voice. *Journal of Voice, 19*(4), 574–581.

Downloaded from: https://pubs.asha.org Katherine Hoener on 10/07/2021, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Appendix A

Template for Laryngeal Imaging

Name: _____ **Date:** _____
Endoscope used: _____ **Gain setting:** _____
Task/s: _____ **Mouth-to-microphone distance and angle:** _____

Vocal fold edges

Left vocal fold	*Smooth/straight*	*Irregular*	*Rough*	*Bowed*	**CNJ*
Right vocal fold	*Smooth/straight*	*Irregular*	*Rough*	*Bowed*	*CNJ*

Vocal fold mobility

Left vocal fold	*Normal*	*Reduced*	*Absent*	*CNJ*
Right vocal fold	*Normal*	*Reduced*	*Absent*	*CNJ*

Supraglottic activity

Lateral compression	*Anteroposterior compression*	*Sphincteric compression*	*CNJ*
	Mild	Moderate	Severe

Regularity

Regular *Intermittent* *Irregular* *CNJ*

Amplitude

Left vocal fold	*0% ≤ 25% ≤ 50% ≤ 75% ≤ 100%*	*CNJ*
Right vocal fold	*0% ≤ 25% ≤ 50% ≤ 75% ≤ 100%*	*CNJ*

Mucosal wave movement

Left vocal fold	*0% ≤ 25% ≤ 50% ≤ 75% ≤ 100%*	*CNJ*
Right vocal fold	*0% ≤ 25% ≤ 50% ≤ 75% ≤ 100%*	*CNJ*

Glottal closure

Complete closure *Anterior gap* *Irregular closure*
Spindle shaped gap *Posterior gap* *Hourglass gap*
Absence of closure *Variable closure (identify predominant pattern)* *CNJ*

Left/right phase symmetry

Symmetric *Intermittently asymmetric* *Consistently asymmetric* *CNJ*

Vertical level

Same level *Different levels (identify which vocal fold is above the other vocal fold)* *CNJ*

Closure duration

Open phase predominant *Closed phase predominant* *Approximately equal* *CNJ*

**CNJ = could not judge. Identify reason if rating CNJ.*

Downloaded from: https://pubs.asha.org Katherine Hoener on 10/07/2021, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Appendix B

Template for Acoustic Analysis

Name: _____ **Date:** _____
Microphone distance (cm/angle): _____ **Sampling rate (Hz):** _____
Quantization rate (bits): _____
Software(s) used for acoustic analysis: _____
Background noise level (dB SPL) during 5 s of silence: _____
SPL frequency weighting for background noise level: C-weighting / A-weighting / no weighting
SPL calibration: yes / no
Measures of vocal sound level

1. **Habitual vocal sound pressure level (dB SPLeq@30 cm, C-weighted):** _____
   - Task: Standard reading passage
2. **Vocal sound pressure level (dB SPL) range:**
   - Task: Glide on the vowel /a/
   - **Maximum vocal dB SPLeq@30 cm, C-weighted:** _____
   - **Minimum vocal dB SPLeq@30 cm, C-weighted:** _____

Measures of vocal frequency

1. **Mean vocal frequency (Hz):** _____
   - Task: Standard reading passage
2. **Vocal frequency standard deviation (Hz):** _____
   - Task: Standard reading passage
3. **Vocal frequency (Hz) range:**
   - Task: Glide on the vowel /a/
   - **Maximum vocal frequency (Hz):** _____
   - **Minimum vocal frequency (Hz):** _____

Measure of noise in the vocal signal

1. **Vocal cepstral peak prominence (CPP in decibels)**
   - Task: Sustained vowel /a:/ for 3–5 s (**CPP_{vowel}**): _____
   - Task: Standard reading passage (**CPP_{speech}**): _____

Appendix C

Template for Aerodynamic Analysis

Name: _____ **Date:** _____
Background noise level (dB SPL) during 5 s of silence: _____
SPL frequency weighting for background noise level: C-weighting / A-weighting / no weighting
SPL calibration: yes / no

	Tasks	
Aerodynamic measures	**Habitual loudness (pi:pi:pi:pi:pi)**	**Raised loudness (pi:pi:pi:pi:pi)**
Average glottal airflow rate (L/s or ml/s)		
Average subglottal air pressure (cmH_2O or kPa)		
Mean frequency (Hz)		
Mean vocal dB SPLeq@30 cm, C-weighted		

Downloaded from: https://pubs.asha.org Katherine Hoener on 10/07/2021, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions