ASHA American Speech-Language-Hearing Association

AJSLP

Research Article

Qualitative Analysis of Speech-Language Pathologists' Voice Evaluation Practices and Perspectives

Michael D. Madoule,[a] Katherine L. Marks,[a,b] Kathleen F. Nagle,[c] Eric Kirchgessner,[d] Aideen Gill,[d] Joshua C. Kline,[d] Jenny M. Vojtech,[a,d] and Cara E. Stepp[a,e,f]

[a] Department of Speech, Language & Hearing Sciences, Boston University, MA [b] Department of Otolaryngology, Emory University School of Medicine, Atlanta, GA [c] Department of Speech-Language Pathology, Seton Hall University, South Orange, NJ [d] Delsys, Inc. and Altec, Inc., Natick, MA [e] Department of Otolaryngology—Head & Neck Surgery, Boston University Medical School, MA [f] Department of Biomedical Engineering, Boston University, MA

ARTICLE INFO

Article History:
Received September 20, 2024
Revision received March 3, 2025
Accepted March 31, 2025

Editor-in-Chief: Rita R. Patel

https://doi.org/10.1044/2025_AJSLP-24-00417

ABSTRACT

Purpose: The purpose of this qualitative study was to examine the structure of voice evaluations and gather clinicians' opinions on the barriers to and benefits of using acoustic measures in these evaluations. A secondary goal was to investigate how clinicians assess strain and vocal effort.

Method: Fifteen voice-specialized speech-language pathologists from voice centers around the United States were interviewed to query their current voice evaluation practice patterns and opinions on acoustic measures. They were also asked how they evaluate strain and vocal effort. Thematic analysis was performed by two researchers based on the recorded interviews.

Results: Differences among practitioners were found in almost every component of the evaluation. Four themes related to barriers to and benefits of implementing acoustic measures in a voice evaluation were identified: Collecting and analyzing acoustic measures (a) take time, (b) do not inform therapy patterns, (c) allow for the most accurate comparison, and (d) supplement patient-centered care. Three themes emerged related to evaluating vocal effort and strain: Clinicians (a) lack consensus on objective measures of strain, (b) use more than just auditory perception to evaluate strain, and (c) assess vocal effort in different ways.

Conclusions: Although some speech-language pathologists view acoustic assessment as the gold standard for guiding therapeutic decisions, others believe it may not be strictly necessary for delivering effective voice therapy. Variations in the assessment of strain and vocal effort across voice clinics suggest a need for additional research in this area.

Supplemental Material: https://doi.org/10.23641/asha.29391725

At some point in their lives, one third of the population will experience a voice problem (Stachler et al., 2018). If an individual elects to receive voice therapy, they must first undergo a voice evaluation. A voice evaluation can be performed by a multidisciplinary team consisting of an (oto)laryngologist and a speech-language pathologist (SLP) or by an independent SLP, each of whom serves specific roles in the diagnostic process (Roy et al., 2013). The primary goal of the physician is to form a medical diagnosis relating to the laryngeal pathology and chart the appropriate course of treatment (Schwartz et al., 2009),

Correspondence to Katherine L. Marks: katie.marks@emory.edu.
Disclosure: *Jenny M. Vojtech has received financial compensation through employment at Delsys, Inc., a commercial company that manufactures, markets, and sells electromyographic sensors and other physiological measurement systems. Cara E. Stepp has received consulting fees from Altec, Inc., and Delsys, Inc., both companies focused on developing and commercializing technologies related to human movement, including software for acoustic measurement. Stepp's interests were reviewed and are managed by Boston University in accordance with their conflict-of-interest policies. The other authors have declared that no other competing interests existed at the time of publication.*

whereas the primary goal of an SLP is to assess voice production, its physiological cause, and its impact on daily functioning as it relates to communication (American Speech-Language-Hearing Association [ASHA], 1998, 2004). In the process of assessing vocal function during a voice evaluation, SLPs may use both subjective and objective methods. There is variability in the practice patterns of voice evaluations among clinicians and voice centers. Although ASHA has introduced recommendations for instrumental voice assessment procedures (Patel et al., 2018), at the onset of our study, no publication to our knowledge had formally assessed the adoption of these recommendations. Our goal was to investigate differences in the ways clinicians from different voice centers perform, analyze, and interpret acoustic and aerodynamic measures, with a specific focus on measuring strain and vocal effort.

Comprehensive Voice Assessment

A comprehensive SLP voice assessment typically includes a case history, self-assessment, and oral-peripheral examination, as well as a respiratory, auditory-perceptual, and instrumental evaluation (ASHA, n.d.). A case history often includes a description of the voice problem, a patient's medical status and history, previous voice treatment, and daily vocal health patterns. Patients' self-assessment of vocal quality and its impact on their quality of life can be measured using tools such as the Voice-Related Quality of Life (V-RQOL; Hogikyan & Sethuraman, 1999) or the Voice Handicap Index (VHI; Jacobson et al., 1997). An oral-peripheral examination is often used to assess the motor and sensory functions of the orofacial structures. An assessment of respiration is used to document respiratory patterns and the coordination of respiration with phonation through tasks such as the production of sustained vowels and the duration of /s/ in relation to /z/ (ASHA, n.d.).

Auditory-Perceptual Evaluation

For clinicians, longitudinal monitoring of a patient's voice allows them to better track patient progress throughout the course of voice therapy. Common auditory-perceptual judgments include overall severity or grade, roughness, breathiness, asthenia, and strain (Hirano, 1981; Kempster et al., 2009). However, auditory-perceptual ratings have been shown to be unreliable (Lee et al., 2005). Even experienced listeners such as voice-specialized SLPs have been shown to frequently disagree when voice quality is rated due to the subjective nature of perceptual judgment (Gerratt et al., 1993; Nagle, 2025).

Over the past 20 years, ASHA's Special Interest Group (SIG) 3: Voice and Upper Airway Disorders has published guidelines and recommendations for voice assessment (Patel et al., 2018). As part of this effort, the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V; Kempster et al., 2009) was created, revised, and validated as a tool for SLPs to assess voice quality percepts. These percepts include overall severity, roughness, breathiness, strain, pitch, loudness, and additional perceptual features (e.g., diplopodia, aphonia, tremor). However, the issue of poor interrater variability remains a limitation (Nagle, 2016). The lack of agreement is amplified in auditory-perceptual ratings of strain, as reflected by worse intra- and interrater reliability compared to those of other vocal qualities, such as breathiness or overall severity (Webb et al., 2004; Zraick et al., 2011). Moreover, auditory-perceptual ratings are influenced by external factors, such as hearing status, familiarity with the speaker, and experience with dysphonic voices (Kreiman et al., 1992, 1993; Nagle, 2025). So, although auditory-perceptual ratings of voice continue to be an important aspect of a comprehensive voice evaluation, there are inherent limitations.

Instrumental Assessment

Instrumental assessment includes laryngeal imaging, acoustic analysis, and aerodynamic assessment. The use of all three has been recommended, as together they more fully characterize the components of voice production (Hillman et al., 1997; Mehta & Hillman, 2008; Patel et al., 2018). With the goal of improving the evidence base for voice assessment measures, enabling valid comparisons within and across clients and clinics, and facilitating the evaluation of treatment efficacy, ASHA created an expert panel that developed a recommended protocol for instrumental assessment of vocal function (Patel et al., 2018). These recommendations cover each aspect of instrumental assessment, specifying data acquisition methods, voice and speech tasks, and data analysis to foster uniformity of voice outcomes (Patel et al., 2018).

Acoustic measures are defined as quantitative values derived from recordings of speech that reflect the severity of variations in typical voicing; correspond to the perception of vocal loudness, pitch, and quality; and provide information about the underlying pathophysiology of vocal function indirectly via noninvasive means (Patel et al., 2018). ASHA-recommended acoustic measures include measures of vocal sound pressure level, vocal fundamental frequency (f_o), and noise in the vocal signal (Patel et al., 2018). Recommended measures of vocal sound pressure level include habitual vocal sound pressure level (SPL; dB) and minimum and maximum vocal SPLs. For vocal f_o, the recommended parameters include mean f_o (Hz), f_o standard deviation, and minimum and maximum f_o. Finally, the expert panel recommended vocal cepstral peak prominence (CPP; dB) to measure the level of noise in the vocal signal. Each of these acoustic measures is associated with an auditory-perceptual percept. Mean, minimum, and maximum vocal SPLs serve as quantitative acoustic correlates of vocal loudness. Mean f_o is the objective correlate of pitch, and the standard deviation, minimum, and maximum f_o are quantitative correlates of prosody. Finally, CPP is a quantitative clinical measure that reflects the general balance between periodic and aperiodic energy in the speech signal and has been empirically studied as a correlate of overall voice quality (Patel et al., 2018). CPP has been strongly associated with auditory-perceptual ratings of breathiness in addition to overall severity of dysphonia (Heman-Ackah et al., 2002).

Assessment of Strain and Vocal Effort

Although a common goal of voice therapy is to reduce a patient's level of strain and vocal effort, there are currently no recommended acoustic measures that specifically target the constructs of strain or vocal effort. In a survey by van Mersbergen et al. (2021), two thirds of voice clinicians reported that a majority of their patients present with complaints of vocal effort, and 25% of clinicians further reported that vocal effort is the primary patient complaint they encounter. The terms "vocal effort" and "strain" represent two different but related perspectives of the same phenomenon. According to Hunter et al. (2020), *vocal effort* is defined as the perceived exertion of a vocalist's response to a perceived communication scenario (a perceptual phenomenon experienced by the speaker, not the listener), and *strain* is defined as the auditory-perceptual judgment of vocal effort by a listener. Hunter et al. discussed several aspects that encompass vocal effort found in the existing literature, including an increase in loudness or strain in voicing. Some descriptions of vocal effort include the presence of discomfort while speaking, tightness in the throat, or tearful-sounding speech (Isetti et al., 2014; Paes & Behlau, 2017). Vocal effort has been associated with environmental factors related to voicing, such as distance between speakers, background noise volume, and vocal use (Bermúdez de Alvear et al., 2011; Brungart & Scott, 2001). Changes in posture have been associated with reports of increased vocal effort (Gilman & Johns, 2017). Additionally, mood changes and cognitive load may play a role in vocal effort (Baldner et al., 2015).

Clinicians have found several different ways to assess vocal effort due to its prevalence in patient complaints; however, there is no consensus as to which available effort scales are best (van Mersbergen et al., 2021). Common scales used to measure vocal effort include visual analog scales (VASs; Marks et al., 2021), direct magnitude estimation scales (as used by Chang & Karnell, 2004; Verdolini et al., 1994), the Borg Category Ratio Scale (Baldner et al., 2015), and the OMNI Vocal Effort Scale (OMNI-VES; Shoffel-Havakuk, Marks, et al., 2019). The strengths and limitations of these scales are discussed in detail in van Mersbergen et al. (2021). They conducted a survey to investigate how clinicians measure vocal effort and found that most clinicians reported using the VHI, despite the VHI measuring a different construct than vocal effort (i.e., voice disability) and containing only three questions related to vocal effort. The Vocal Fatigue Index (an inventory of questions that include Likert scales) was another scale used by clinicians to assess vocal effort (van Mersbergen et al., 2021). Despite the number of tools available, there is no current consensus on how best to measure vocal effort. Conversely, vocal strain as an auditory-perceptual phenomenon is most often rated using a VAS such as the scales of the CAPE-V (Kempster et al., 2009) or an ordinal scale such as the Grade, Roughness, Breathiness, Asthenia, Strain scale (GRBAS; Hirano, 1981). However, using these scales presents the same problem described previously: Subjective judgments are unreliable (Lee et al., 2005). For this reason, many researchers have focused their attention on finding an objective measure that correlates with strain and vocal effort.

Potential objective correlates of vocal effort and/or strain can be classified as either physiological measures or acoustic measures. Beginning with physiological aerodynamic measures that quantify how air travels during phonation, Rosenthal et al. (2014) found that subglottal pressure, translaryngeal airflow, and maximum flow declination rate are significantly increased during speech produced with maximal effort as compared to speech produced at comfortable effort levels in speakers with typical voices. Supporting this finding, McKenna et al. (2019) also found subglottal pressure to be a significant predictor of self-ratings of vocal effort. Furthermore, Chang and Karnell (2004) demonstrated that a direct and moderately strong relationship exists between phonation threshold pressure and perceived phonatory effort. Increased mediolateral supraglottic compression (observed during endoscopy) and activation of the suprahyoid muscles (measured using surface electromyography) were also found to be physiological predictors of vocal effort (McKenna et al., 2019). The findings in these studies are significant; however, aerodynamic measures are not always clinically feasible. In addition to these measures being more invasive than acoustic measures, the tools necessary to perform aerodynamic, electromyographic, or endoscopic evaluations are not available to all clinicians.

Acoustically, there is no objective metric that is specific to strain or vocal effort. CPP has been shown to be significantly higher during maximal effort speech compared to comfortable effort speech (Rosenthal et al., 2014). However, CPP is strongly related to breathiness and overall severity and, thus, is not specific to strain or vocal effort (Heman-Ackah et al., 2002). Similarly, Lowell et al. (2012) found that some cepstral- and spectral-based measures are effective in distinguishing strained from unstrained voices, but the same limitation exists: These measures, which include CPP f_o (Hz), CPP (dB), CPP standard deviation (dB), low-to-high (L/H) ratio (dB), and L/H ratio standard deviation (dB), have been primarily used to characterize breathiness and roughness and are thus not specific to strain.

Practice Patterns for Voice Evaluations

Given the variable evidence base and availability of voice outcomes, current practice patterns are not clearly understood. Behrman (2005) collected survey data from 53 experienced SLPs, 100% of whom reported being "very likely" to use auditory-perceptual measures. The survey also found that other subjective measures such as observations of body posture and movement and the patient's ability to alter voice production were common methods of assessment in voice evaluations. Compared to objective measures, such as acoustic, aerodynamic, and electroglottographic data, respondents were significantly more likely to use subjective measurements in their voice evaluations (Behrman, 2005). However, it is important to note that the survey was conducted well before the ASHA-recommended protocol was published (Patel et al., 2018). More recently, following a discussion held by ASHA SIG 3, Estes and Johnson (2023) described clinicians' reports of their experiences implementing vocal function (i.e., acoustic and aerodynamic) assessment in their voice evaluations. The group discussion revealed disparities between the ASHA-recommended protocol for acoustic assessment and clinicians' practice patterns regarding the environment, hardware, software, tasks, and measures used during the evaluation. For example, many clinicians did not have reliable access to a sound-treated environment, used free-standing microphones rather than the recommended head-mounted microphones, or reported that acoustic analysis software was too expensive to implement (Estes & Johnson, 2023). To our knowledge, aside from the community discussion report, no study had formally assessed the adoption of ASHA-recommended measures for clinical voice evaluation at the time our study was conducted. However, a new article by Salgado et al. (2024) used a survey to assess current practices in the assessment of voice across providers from different clinical settings. The need for more formal review motivated our qualitative study, in which we aimed to understand clinicians' practices and opinions on voice evaluation, specifically including strain and vocal effort, within the context of their clinical work environment. Our work is coincidentally complementary to the work of Salgado et al., providing rich contextual information from clinicians that is inaccessible via survey methods.

Thus, the purpose of this study was to broadly investigate differences in voice evaluations across clinicians from different centers. A qualitative approach using a semistructured interview allowed us to gather contextual information from treating SLPs about their current practice patterns, values, and beliefs regarding voice evaluations within the context of their work environment and the practices in place at their respective institutions. We queried the components each SLP included in their voice evaluations, the equipment they used, and which measures were taken. We sought specifically to understand the barriers to and benefits of performing acoustic analysis as part of the voice evaluation process from the perspective of the clinician. Secondarily, we were interested in how clinicians measure strain and vocal effort. We hypothesized that, among the pool of participants, there would be variations in the structure of the voice evaluation as well as differences in opinion on the use of acoustic analysis and the ways in which they measure vocal effort and strain.

Method

Participants

We used purposive sampling (i.e., recruiting respondents most likely to provide relevant and useful information) to reach concept saturation on the barriers to and benefits of acoustic evaluation of voice. Concept saturation is reached when no novel ideas or opinions are identified in the data (Saunders et al., 2018). Over the course of the study, 22 SLPs were asked to participate. Of these, two declined to participate, three did not respond to e-mails, and two agreed to participate but scheduling conflicts prevented them from completing the interviews. We had initially planned to interview 10 clinicians, with half known to work in settings that included acoustic and/or aerodynamic measurement and half known to work in settings that did not. Once the transcript from the 10th participant was coded, however, it was evident that concept saturation had not been reached, as the 10th participant had offered a novel barrier to using acoustic measures. Thus, we continued recruiting and enrolling participants until concept saturation had been reached for each theme of the thematic analysis.

Fifteen SLPs (14 cisgender women and one cisgender man) who specialize in voice disorders participated in the study and were interviewed until concept saturation was reached. The participants practiced at different voice centers around the United States and had at least 3 years of experience treating a caseload that consisted primarily of patients with voice and upper airway disorders. Thirteen participants worked in academic voice centers; one participant worked in a private practice with a laryngologist; and one participant worked in a general ear, nose, and throat practice but exclusively treated voice and upper airway disorders. All participants practiced at different centers. Participants ranged in age from 28 to 61 years (M = 42.1, SD = 10.8). Their years of experience specializing in voice disorders ranged from 3 to 34 years (M = 11.6, SD = 8.4). Eight participants practiced in the Northeast region of the United States, and the other seven practiced in various regions across the United States. Fellowship training also occurred in various regions across the United States. Participants were recruited from the authors' personal contacts who work in the area of voice.

Protocol

Interviews were conducted via an online teleconference platform, and all interviews were recorded. The only individuals present during the interview were the experimenter, who was a master's student in the Speech and Hearing Sciences program at Boston University, and the participant. Although the master's student had been taught ASHA-recommended measures of voice evaluations in the master's-level voice and speech science course, he had not yet practiced in a clinical voice setting, so he was relatively unbiased in the implementation of acoustic measures and evaluating strain and effort.

The experimenter consented the participant as approved by the Boston University Institutional Review Board (No. 2625). Participants were paid $30/hr for participation, resulting in a mean payment of $45, as the mean interview time was 1 hr 22 min. The experimenter then prompted each participant with the following request: "Walk me through your typical voice evaluation process as if I were a patient coming in with muscle tension dysphonia or showing signs of muscle tension dysphonia, starting with any case history or forms and any part of your evaluation that is documented in your evaluation notes." As the participant responded, the experimenter checked off aspects of a voice evaluation that were mentioned using a template that listed potential responses. Any aspect of the participants' voice evaluation process that was not already listed on the table was added. Supplemental Material S1 provides interview prompts and data tables used during the interviews. For the aspects of the voice evaluation that were not mentioned, the experimenter explicitly asked questions to elicit a response (e.g., if the participant did not address laryngeal imaging, the experimenter asked, "Do you perform any laryngeal imaging during your voice evaluation?"). Further specific questions and details about stimuli were noted on a separate template table that listed potential stimuli and speech tasks, with room to add additional reported information.

Next, participants were asked what benefits or barriers they perceive (if any) to implementing acoustic analysis into their evaluation. Following this line of inquiry, the experimenter asked specifically how participants evaluate strain and vocal effort. Strain was defined to them as "a listener's auditory-perceptual judgment of vocal effort," and vocal effort was defined to them as "the self-perceived exertion during vocalization by the speaker" (Hunter et al., 2020). The participants were also asked about how the COVID-19 pandemic has influenced their voice evaluation process to understand lasting changes in protocols. Finally, the participants were asked if they had any additional opinions or thoughts they would like to share about acoustic or aerodynamic analysis.

Analysis

Following data collection, the audio recording from each interview was used to generate a transcript using a Microsoft speech-to-text function. Two researchers were involved in thematic analysis: a senior researcher (K.F.N.), who is a scientist with expertise in qualitative research of voice, and a graduate student researcher (M.D.M.), who is a second-year master's student in speech-language pathology and had taken a graduate course on voice and speech science. The graduate student researcher was trained in thematic analysis by the senior researcher, using previously published qualitative research and by discussing and refining the code book and coding process. One researcher proofread each interview transcript while listening to the recording to add punctuation and correct any spelling or transcription errors. Next, the first two transcripts were reviewed independently by the two researchers for meaning units, defined as words, phrases, sentences, or paragraphs that describe a specific phenomenon (Grossoehme, 2014). After this process, the two researchers collaborated to select the most appropriate meaning units and create a code book by assigning labels to the meaning units derived from the first two transcripts. The code book was then used as a guide for both researchers to code all other transcripts independently, with revisions made as needed. The review for meaning units and coding were performed independently. Following this process, the researchers compared their individual coding for each transcript and reached agreement on which were the most accurate for each transcript. The researchers discussed any codes on which they did not agree until they reached agreement (Coffey & Atkinson, 1996). Next, themes summarizing recurring concepts were derived from these codes independently by one researcher and then solidified only after referencing and collaborating with the second researcher. This thematic analysis was performed specifically in the context of (a) the barriers and benefits of acoustic analysis and (b) the assessment of strain and vocal effort in the voice evaluation.

The concepts of validity, reliability, objectivity, and generalizability are foundational in quantitative research. For this qualitative study, they were represented by the concepts of credibility, dependability, confirmability, and transferability, respectively (Lincoln & Guba, 1985). *Credibility* refers to the extent to which the opinions of the participants are accurately interpreted. To ensure credibility in this qualitative analysis, we selected suitable participants, cited quotations that accurately reflected the data, and selected relevant meaning units using four main strategies: (a) Voice clinicians were defined as suitable participants for a study of acoustic analysis in voice evaluations; (b) the participants' responses were clarified during the interview to ensure that the correct interpretation was being drawn; (c) all transcripts were read and coded separately by two researchers, after which the two coders discussed their findings until agreement was reached; and (d) common themes regarding the use of acoustic and aerodynamic analysis and assessment of strain and vocal effort in voice evaluations were represented using quotes from the participants. *Dependability* is the ability to obtain the same results if the study were to be repeated (Morse, 2015), and *confirmability* refers to the grounding of qualitative findings in the data as opposed to deriving from the researcher's imagination (Enworo, 2023). Dependability and confirmability were established alongside credibility: Both during and after the interviews, the researcher asked the participants to clarify and verify their statements.

Results and Discussion

The mean interview time was 1 hr 22 min, ranging from 1 hr 8 min to 1 hr 45 min. The evaluation structure across clinicians is described below, and the acoustic and aerodynamic measures, stimuli, and software/equipment used are listed in Supplemental Material S2. The results of the thematic analysis are reported in two parts: (a) barriers to and benefits of collecting acoustic measures in clinical practice and (b) measurement of strain and vocal effort. The themes generated from the interviews were not mutually exclusive, meaning that there are instances in which two or more themes may contradict each other. This is due to the diverse set of opinions that existed regarding the use of acoustic assessment within voice evaluations.

Evaluation Structure

All participants interviewed in the present study reported obtaining a case history using case history forms and/or interviewing. Their case history questions primarily pertain to the patient's voice complaint, such as the onset and progression of their voice problem, as well as social and medical history, with common questions referencing smoking and reflux. The extent of the case history portion of their evaluations is largely influenced by the length of the evaluation, which varied across participants and centers, ranging 15–120 min (M = 49, SD = 28). Despite all participants reporting performing a case history, the depth and number of their questions to patients varied significantly, likely due to the differences in time allotted for the voice evaluation.

All participants reported collecting patient-reported outcome measures (PROMs), although the number, variety, and timing of when patients completed them varied. In total, the following PROMs were reported by at least one participant: VHI-10 (Rosen et al., 2004), VHI-30 (Jacobson et al., 1997), Singing VHI (S-VHI; Cohen et al., 2007), Evaluation of the Ability to Sing Easily (Phyland et al., 2015), Reflux Symptom Index (Belafsky et al., 2002), OMNI-VES (Shoffel-Havakuk, Marks, et al., 2019), Cough Severity Index (CSI; Shembel et al., 2013), Dyspnea Index (Gartner-Schmidt et al., 2014), Eating Assessment Tool (EAT-10; Belafsky et al., 2008), Communication Participation Item Bank (Baylor et al., 2013), Glottal Function Index (Bach et al., 2005), V-RQOL (Hogikyan & Sethuraman, 1999), Leicester Cough Questionnaire (Birring et al., 2003), and Voice Catastrophization Index (Shoffel-Havakuk, Chau, et al., 2019). It was commonly noted that although general voice questionnaires, such as the VHI, are provided to every patient, questionnaires targeting specific profiles, such as the CSI, EAT-10, and S-VHI, are administered only when relevant to a specific patient concern. The two most reported ways of administering PROMs are remotely prior to the evaluation (more common in the case of shorter evaluation blocks) or in person when patients arrived at the clinic.

All participants reported performing an auditory-perceptual evaluation. Of the 15 participants included in this study, eight (53.3%) reported using only the CAPE-V for their auditory-perceptual evaluation, five (33.3%) reported using only the GRBAS, and two (13.3%) reported using both scales. Of the 10 participants who used the CAPE-V during the evaluation, six (60%) stated that although they use the general structure of the CAPE-V, they do not use the CAPE-V per the published protocol (Kempster et al., 2009) but rather make modifications, such as not physically ticking the line and then measuring it with a ruler, opting instead to assign ranges of numbers for "mild," "moderate," and "severe" or selecting a number within a range. Although the other four participants did not explicitly report any modifications to the CAPE-V, it was not verified whether they used the published protocol. The lack of fidelity among SLPs when administering the CAPE-V is well documented, as is the resulting risk to reliability and validity (Lodhavia & Kempster, 2024; Nagle, 2025).

Regarding laryngeal imaging, 10 participants (66.6%) reported that a flexible endoscope is used more frequently than a rigid endoscope at their respective practices. Three participants (20%) discussed that the use of a flexible or rigid endoscope depends on the voice quality and patient history alluding to a particular etiology (e.g., flexible scope for a breathy voice quality and rigid scope for a rough voice quality). One participant (6.7%) reported that rigid endoscopes were used more often than flexible, whereas another participant (6.7%) reported that rigid and flexible endoscopes were both used equally. Of the 15 participants, seven (46.7%) reported that the laryngologist or laryngology fellow operated the endoscope all or most of the time. Six of these participants (40%) reported they are typically present in the room (if possible) for the endoscopy, whereas one participant (6.7%) is not present. Five participants (33.3%) consistently perform the endoscopy themselves. One participant (6.7%) reported that endoscopy is performed by either the SLP or the physician, split equally. Two participants (13.3%) reported that they operate the scope only if a patient is referred from an outside otolaryngology clinic.

When describing the components of their voice evaluations, three participants (20%) reported performing laryngeal palpation. It was only later in the interviews, when asked how they assess strain, that three other participants stated that they perform laryngeal palpation as well. Conversely, eight participants (53.3%) did not mention laryngeal palpation at any point during the interview, and the remaining participant explicitly reported that they no longer perform laryngeal palpation following the COVID-19 pandemic. The low-percentage report of laryngeal palpation in the context of a muscle tension dysphonia evaluation is surprising given that palpation is a non-instrumental method for assessing muscle tension, a hallmark feature in muscle tension dysphonia (Morrison & Rammage, 1993), and is included in the ASHA-recommended protocol for voice disorder assessment (ASHA, n.d.). One potential reason for the low percentage of reported palpation may be due to difficulty in the participants' ability to recall every component of their voice evaluation structure. This is supported by the three participants who did not initially report performing palpation but did later on when probed specifically for their strain assessment practices. More discussion on this limitation is to follow.

Out of the 15 participants in the study, eight (53.3%) reported that oral mechanism examinations are a standard component of their assessment process; three (20%) reported they do not include an oral mechanism examination in their voice evaluation; and four (26.7%) reported that an oral mechanism examination is not a standard aspect of their voice evaluation but that they perform one as needed, most commonly for suspicion of a neurological disorder (e.g., if the patient presents with dysarthria). Hearing screenings were seldom part of the participants' evaluations. One participant (6.7%) reported that they perform a hearing screening as a part of their standard voice evaluation. The other 14 participants (93.3%) reported that they do not perform hearing screenings during their evaluations. However, one of these participants noted that all new patients are referred to their clinic's audiologists; thus, they do not need to perform a screening. Another participant reported that they include questions pertaining to hearing during the case history portion of the evaluation and refer to their clinic's audiologist if there is a concern.

Although many of the participants reported collecting acoustic data in a manner consistent with Patel et al. (2018), there was still much variety in the type and total number of measures taken, as well as the software used to gather these measures. A table of the stimuli used to obtain these measures can be found in Supplemental Material S2. The two most commonly reported systems for measuring acoustic parameters were the Computerized Speech Lab (CSL 4400; KayPENTAX) and Praat (Boersma, 2001). Across interviews, the following CSL-compatible software were mentioned: Analysis of Dysphonia in Speech and Voice (Model 5109, Version 3.4.2; KayPENTAX), Multi-Dimensional Voice Program (Kay Elemetrics, 1993), and Real-Time Pitch (Kay Elemetrics, 1994). One participant reported using Phonanium (Maryn, 2017) for telehealth evaluations. Five clinicians (33.3%) reported not using acoustic software at all. This may seem consistent with the findings of Salgado et al. (2024) that 28% of providers did not have access to acoustic equipment. However, it is important to note that Salgado et al. included general SLPs in addition to those who primarily treat voice disorders. Although they reported on barriers such as access to equipment by instrumental evaluation type (acoustic, aerodynamic, and videoendoscopic), the other barriers mentioned were not specific to any type of evaluation. Our results yielded different perceived barriers to using acoustic analysis as part of the voice evaluation, which are subsequently reported in our thematic analysis.

The COVID-19 pandemic had a notable impact across several domains of everyday life and societal functioning. Throughout the pandemic, health care workers continued to provide care, but they had to do so under modified conditions. The most reported lasting changes due to COVID-19 concern the Phonatory Aerodynamic System (PAS), a computer-based hardware and software system used to measure aerodynamic voice characteristics. Six participants (40%) reported a change to their PAS protocol since the pandemic: Four participants (26.7%) reported updates to cleaning protocols involving disassembly or entirely disposable parts, and two participants

Table 1. Themes and subthemes of the barriers to and benefits of collecting acoustic data.

Theme	Subtheme	Definition
Collecting and analyzing acoustic measures take time	Time for clinician Time for patient	Evaluation length and added analysis; Patient fatigue and frustration
Collecting and analyzing acoustic measures do not inform therapy patterns	Acoustic measures do not reflect perception Acoustic tasks do not capture the problem Clinician perception is enough	Incongruence between what someone is hearing and what the values are reflecting; Many acoustic tasks are not reflective of speaking conditions Clinicians can hear what needs to be targeted in therapy without acoustic data
Collecting and analyzing acoustic data allow for the most accurate comparison	Collecting recordings Objective progress	Audio recordings can be used to compare voices at different time points; Acoustic data can be used for progress tracking
Collecting and analyzing acoustic measures supplement patient-centered care	Satisfaction/validation/buy-in Education	Patients like seeing acoustic data; Clinicians use acoustic values to educate patients

(13.3%) reported discontinuing the collection of instrumental aerodynamics entirely. Two additional participants (13.3%) reported that their voice evaluation has returned entirely to prepandemic protocols, whereas another (6.7%) stated they no longer include acoustic or aerodynamic measures in their voice evaluations. Other lasting changes mentioned included the option to perform evaluations via telehealth, adding case history questions specifically related to COVID-19, discontinuing laryngeal palpation to assess tension, placing lower value on acoustic and aerodynamic measures, continuing to wear masks in clinic, and updating endoscope sanitation protocols. COVID-19 precautions were also reported as a barrier to instrumental voice assessment by 14% of the participants in the Salgado et al. (2024) study. Although many of the initial mandatory restrictions and regulations have been relinquished, the COVID-19 pandemic has had varying and, in some cases, lasting impacts on the voice evaluation process in clinics across the United States.

Thematic Analysis

In the following subsections, results of the thematic analysis are described and discussed. A summary of the themes and subthemes with their definitions are provided in Tables 1 and 2. The frequency of reported barriers and the frequency of reported benefits are illustrated as figures in Supplemental Materials S3 and S4, respectively.

Barriers to and Benefits of Including Acoustic Assessment

The main reported barriers to including an acoustic assessment in a voice evaluation were the time it takes to measure acoustic signals and the belief that acoustic measures do not inform therapy decisions. Reported benefits were that acoustic measures serve as the best indicator of pre- to post-therapy change and supplement patient-centered care.

Theme 1: Collecting and analyzing acoustic measures take time.

Time for clinician: An important recurring observation across interviews was the trade-off between time and clinical benefit. One participant reported that this limitation was specific to their work environment, in which they would need to move patients to a different room as well as clean the microphones and equipment. Another participant expressed a similar sentiment that the time for the retrieval of equipment and the setup required outweighed

Table 2. Themes and subthemes describing the evaluation of strain and vocal effort.

Theme	Subtheme	Definition
Clinicians lack consensus on objective measures of strain	N/A	Clinicians use different acoustic and aerodynamic measures as a correlate to strain
Clinicians use more than just auditory perception to evaluate strain	Auditory perception Visual and tactile perception Synesthetic perception	Hearing strain; Seeing strain or feeling strain through touch; Feeling laryngeal muscle tension in their own body after hearing a strained voice
Clinicians assess vocal effort in different ways	Patient report Correlates	Scales, forms, and open-ended questions for self-reported; Clinician-perceived patient behavior indicative of strain

Note. N/A indicates that no subtheme existed.

the benefit of obtaining the measures. They noted that although their clinic owns the CSL equipment, they only use it when they have time, which they reported to be infrequent. In reference to not using the CSL equipment, one participant stated, "It's just we don't have the time, you know. We've got to do the things that are really going to clinically benefit us." A different participant echoed this sentiment by specifying what they do with the additional time: "It just serves everybody better, in my opinion, if I can spend that time helping a patient, or seeing another patient, frankly, rather than spending the time on [aerodynamic and acoustic assessment]." This view was similarly expressed by another participant who emphasized that collecting acoustic measures takes up time not only during the evaluation but also during the day at large: "It's not only just adding more time to my actual evaluation, it's adding time to my analysis and, again, for the fact that it might not change my plan of care ... if I'm seeing, let's say, 10 patients in a day, that's, let's say, [adding] 10 minutes [each], that's an extra 100 minutes added on to my day just to do this." The time constraints reported by these participants are consistent with Estes and Johnson (2023), who also found time to be a commonly reported barrier to vocal function assessments. Likewise, lack of time was reported as a barrier to performing general instrumental evaluations by 22% of the participants in the Salgado et al. (2024) survey.

Time for patients: Several participants described that the additional time added to the evaluation for the collection of acoustic measures may negatively impact the patient's experience at the clinic. One participant reflected on the demeanor of patients as they arrive to the clinic for an evaluation and reported that, in some cases, patients are anxious to discover the nature of their pathology. They stated that the additional time required to collect acoustic measures can lead to patient frustration: "Some patients at the end of the day were like, 'Why are we doing this? This is taking too long. I just want to know [if] I have laryngeal cancer or not. I just want to go straight to the strobe.'" Another participant shared a similar thought that collecting and analyzing acoustic measures extend the length of the evaluation, which can lead to patients' vocal fatigue, potentially resulting in an inaccurate representation of their voice at that time.

Theme 2: Collecting and analyzing acoustic measures do not inform therapy patterns.

Acoustic measures do not reflect perception: A common complaint across participants was that there is sometimes an incongruence between their perception of a patient's vocal quality (e.g., roughness, breathiness) and the acoustic correlate. One participant gave the example: "You don't hear roughness, you hear a little bit of breathiness, but when you [analyze] some acoustic [measures], you're getting a huge noise to harmonic ratio or you're getting a lot of roughness coming out, and the CSID is through the roof." Outside of an evaluation context, several other participants discussed their experience with acoustic measures worsening as the patient perceptually improved. One participant discussed this phenomenon with CPP: "It's not uncommon that maybe the CPP value gets objectively worse, but the patient's symptoms are much better." Another participant reflected on similar experiences with jitter and shimmer: "I have such a long and storied history with jitter and shimmer getting worse when people perceptually get better." This participant went on to explain how this past inconsistency with jitter and shimmer have led them to become more untrusting of the reliability of acoustic measures in general. This same issue holds true for patient perception: One participant explained that there are times when a patient arrives with complaints of vocal fatigue/discomfort/pain but will generate acoustic and aerodynamic values that are within typical limits. Interestingly, although this was a commonly reported barrier, another participant shared a contrasting opinion. Whereas some saw an incongruence between perception and acoustics data as a flaw, they stated that acoustic measures were detecting elements of the voice inaccessible to clinician perception and thus providing insights that clinicians may be missing.

Acoustic tasks do not capture the problem: A third barrier reported was a lack of ecologically valid tasks used to elicit acoustic measures; tasks are "far removed from typical voice use in which the problem is usually encountered." One participant shared that they often face this issue with singers; if a singer comes to the clinic with a specific singing-related voice concern (e.g., "I feel like I'm choking on my high notes"), their complaint is typically not captured in the acoustic battery of tasks. "Even if we're asking them to phonate their highest pitch, because it's not a singing task, and it's brief ... it's not really capturing their concern." Another participant reported that even a reading task can change the quality of a patient's voice: "When someone is sitting and they know that they're being recorded, they are not using their voice in the way that they use it every single day when they're talking to someone, so I think you're getting a skewed assessment of their actual voice function." Another participant explained that although some measures, such as CPP, can be calculated from more naturalistic stimuli, other measures require tasks that are dissimilar to everyday voice use while speaking: "I think something that's terrific about CPP being able to be done in the rainbow passage, versus some of the other measures that we do in a sustained /a/, is that it's more reflective of how somebody might use their voice functionally."

Perception is good enough: Although not all participants held this opinion, six participants (40%) reported that patient and/or clinician perception is good enough for patient care. This subtheme is predicated on two beliefs: (a) Clinicians can perceive what needs to be addressed in therapy without objective data, and (b) patient perception is the most important. Several participants said they can form appropriate therapy treatments and targets without gathering acoustic measures. One participant expressed the opinion:

> Like cepstral peak [prominence] … I already knew they were dysphonic when they opened their mouth, so like, I don't need a number to tell me that, and at the end of all the therapy, I don't need the number to tell me that they got better. The patients are either going to fill out their survey one way or they're not, or they're going to tell me subjectively that they're better or they're not.

Similar opinions were shared among several other participants. Another participant reported that despite collecting acoustic measures, their practice patterns were rarely influenced: "We never did CPP so I can't speak to that, but acoustic data like jitter, shimmer, and noise never affected the way that I treat my patient as a clinician. It never changed the therapy tasks that I chose. It never changed their outcomes." Several participants expressed that therapy targets are ultimately the decision of the patient. One participant rhetorically asked, "I may hear 10,000 on the CAPE-V on strain … but if you tell me that you don't feel strain, you don't feel like you're hurting yourself, and you can do everything you want to do with your voice, do we have to fix it?" Or as another participant put it, "A voice disorder is not defined by the acoustic parameters that we're measuring, it's defined by the way that the patient is experiencing their symptoms." Another expressed this same view: "The really good voice clinician is the 'I'm not fixing what I want to fix; I'm fixing what *you* [the patient] want to fix.'" The same participant added that they have had patients who leave therapy with an improved, albeit dysphonic, voice, and if the patient is satisfied with their progress and their voice, they do not insist that the patient continue therapy.

In almost all cases in which time was a reported barrier, participants also expressed either that analyzing acoustic data does not inform their treatment plan or that acoustic values are not always beneficial to collect. In other words, there seems to be a compounding effect in which clinicians who favor acoustics overlook the added time required, whereas those who disfavor acoustics are more likely to cite time as impedance. Additionally, some of the clinicians who form therapeutic goals without acoustic data noted an ethical problem with the cost incurred on the patient's insurance and the extra time it takes. Because the issue of ethics came up in three interviews, this sentiment may be shared by a number of clinicians in the field, which warrants further discussion or investigation. Below are quotes from three different participants that explain their views:

> I feel like it's almost borderline unethical to spend a lot of time doing [acoustic assessment] when I could make [patients] potentially better, which would increase the likelihood of them coming back for therapy.

> It's more of a time burden, and it almost feels like an unethical thing to bill if it's not going to affect the way that I take care of the patient.

> I think until we have enough research that really demonstrates that [acoustic measures] are really good at showing outcome measures better than asking a patient if they feel better or not, then it's hard for me to say that this is a valuable thing for us to ask people to pay for.

Although Salgado et al. (2024) queried perceived barriers to all instrumental voice assessment types and did not focus specifically on acoustics, they reported similar barriers, including inadequate reimbursement (5%), lack of perceived benefit or usefulness (17%), lack of time (22%), and lack of training or experience (29%). Again, it is important to note that their participants included both general and voice-specialized clinicians, and they did not report barriers by participant setting. However, the mean perceived utility of acoustics on a scale of 0–10, with 10 being extremely useful, was not statistically different between voice-focused and general clinicians (7.31 for voice-focused clinicians and 7.86 for general clinicians). So, although only 13% of participants in their study reported no barriers, the mean perceived utility score for acoustic assessment was moderately high. We did not query our participants explicitly about their training or experience, and it is possible that there is a knowledge gap among them about the interpretation of acoustic data and equipment used to obtain it, including free or low-cost options. Future work may determine how widespread this knowledge gap is and test ways in which the barriers can be addressed.

Theme 3: Collecting and analyzing acoustic data allow for the most accurate comparison.

Collecting recordings: Several participants discussed the value of recording patients' voices. Although recording a patient's voice does not necessitate acoustic analysis, recordings can be used for evaluation purposes. Participants reported that saving an audio recording for later reference is helpful for several different aspects of therapy, such as perceptually comparing voice quality after a patient undergoes treatment or a procedure (e.g., vocal polyp excision) to their baseline prior to treatment or surgery. One participant explained that these comparison recordings allow both them and the patient to assess progress rather more accurately than relying on their memory. They explained that with "the number of patients we all see in a week, if you go a couple weeks between seeing someone, it's hard to really remember how different they sound from the last time" or what they sounded like at the beginning of therapy. They also acknowledged that the patient may not appreciate the change based on memory. Similarly, another participant shared that it can be motivating for a patient to hear what they sounded like before and after treatment.

Objective comparison and progress tracking: The most frequently reported benefit of collecting and analyzing acoustic data during voice evaluations was to gather data midway or at the end of therapy for objective progress tracking and comparison. This benefit was reported by one third of participants. Participants reported that measuring acoustic data is important for showing objective change and ensuring measurable outcomes. One participant offered a specific example: "For example, CPP is very nice for showing any changes in the signal part of the recording, so that if their voice was super noisy to begin with you can show either the richness of the harmonics that have changed or the signal proportion compared to background noise, and the CPP is really great for that." In addition to therapeutic outcomes, two participants reported the importance of measurable outcomes following medical or surgical treatments to document change.

Theme 4: Collecting and analyzing acoustic measures supplement patient-centered care.

Satisfaction/validation/buy-in: One concept that arose throughout the interviews was that analyzing acoustic measures positively influences patients' buy-in to therapy and patient satisfaction. Participants noted that patients appreciate seeing acoustic data when presented and explained in a patient-friendly manner. According to several participants, acoustic data can validate what the patients are experiencing with their voice. One participant explained that it can be reassuring to patients that a computer can validate what they are hearing and feeling about their voice. Another participant specifically noted their experience with patients who want to know if their voice is "normal" or not: "I think that it's beneficial for patients to know, 'Okay, this is within normal limits; you're OK.'" Another participant further specified that the graphs and charts generated by acoustic software are a large component of patient buy-in to therapy. A third participant described that showing patients that acoustic values change due to stimulability is a "proof of concept for the client that behavioral changes can result in vocal changes."

Acoustic measures influencing patient buy-in was an example in which a perceived benefit for some participants was a perceived barrier for other participants. One participant noted that the collection of acoustic data could be a detriment because it can negatively influence buy-in if the patient does not understand the benefit or relevance of the tasks they are performing and the measures that the clinician is obtaining. One participant gave a general insight that interpretation of acoustic data can be confusing for those who lack the training, including both other health care professionals and patients, emphasizing the importance of accurate and easy-to-understand explanations.

Education and understanding: On the other hand, many participants reported that acoustic measures are great tools for patient education. One participant explained that acoustic signals allow patients to "see their voice" and make the purpose and relevance of the voice evaluation more tangible, as opposed to relying on purely perceptual descriptions. Another participant described this process as a "grounding" of patient understanding in why their voice is "doing the things that it is doing." Participants also expressed that acoustic analysis and visualization produced by software can be used as biofeedback for patients, helping patients understand how to adjust their voice. So, although education and patient buy-in were concepts with conflicting opinions, most participants perceived them as benefits of using acoustic measures in their voice evaluations. Further investigation is needed to determine whether the perceived barriers could be addressed with additional clinician training or streamlining of acoustic analyses and reporting.

Evaluation of Strain and Vocal Effort

A secondary goal of this study was to investigate methods of quantifying strain and vocal effort. The query of measuring strain and vocal effort yielded three themes and five subthemes. The interviews resulted in three themes and five subthemes, as discussed below.

Theme 1: Clinicians lack consensus on objective measures of strain.

Eleven of the 15 participants (73.3%) reported that they did not obtain an objective correlate of vocal strain in their voice evaluation. Of the four participants who reported using an objective correlate to measure strain, one participant reported using subglottal pressure and vocal efficiency, and another reported using solely vocal efficiency. A third participant reported referencing CPP values, as well as the definition of harmonics in the spectrogram, to assess strain. Finally, one participant shared that they use habitual pitch and loudness as an acoustic correlate of strain. Although there is research to support some of these correlates (Anand et al., 2019; Lowell et al., 2012), other researchers report that no acoustic indices reliably predict strain (Bhuta et al., 2004). Moreover, clinicians have the least agreement in auditory-perceptual evaluations of strain (Zraick et al., 2011). These findings indicate the need for continued research of an objective measure of strain.

Theme 2: Clinicians use more than just auditory perception to evaluate strain.

Auditory perception: As mentioned previously, most participants reported using either the CAPE-V, the GRBAS, or both to assess auditory-perceptual voice qualities, including strain. Some, but not all, participants described what they are listening for when rating strain using their perceptual scale. Table 3 provides participant quotations that reflect their auditory-perceptual descriptions of strain. The two major patterns that emerged were the perception of strain as a "pushed" quality or as a loss of resonance.

Visual and tactile perception: Although we defined strain to participants as an auditory-perceptual phenomenon, several participants reported pairing a visual-perceptual assessment to form their clinical picture of it. Two participants reported that they observe patient posture and muscle tension as components of strain. In addition to visually observing muscle tension, laryngeal palpation was reported as a part of the clinical picture of strain as well. The most reported visual-perceptual assessment of strain was noting supraglottic compression visualized via endoscopy. Specifically, one third of participants reported looking for patterns of hyperfunction via endoscopy: "I will look for patterns of phonatory hyperfunction, whether that's anterior–posterior, sphincteric, lateral, or … 'splinting,' where it's almost like you're using so much tension you're actually holding the vocal cords apart." Breathing was another visual marker reported by some participants to assess strain. One participant specified that, in terms of respiration, strain is indicated if a patient does not take an adequate number of breaths while speaking.

Synesthetic perception: Several participants reported using their own body's response to auditory perception as a method of assessing strain. One participant stated, "I can listen with my ear and my body responds … I can provide my own facsimile of how it might feel to create a voice." Another participant reported a similar experience: "I feel my mirror neurons kick in … as I'm listening to somebody, my muscles start to like shift … it almost seems to take the shape of what it thinks the other person is doing." Similarly, a third participant reported that they volitionally will mimic a patient's voice to assess the severity of strain: "I'm a big fan of mimicking people's voices to try and figure out the amount of strain that's going on."

Theme 3: Clinicians assess vocal effort in different ways.

Patient report: The most reported method of assessing vocal effort was via open-ended patient explanation. This consisted of asking the patient for a verbal or written explanation of their vocal effort. One participant reported probing vocal effort via analogy: "I ask them to describe

Table 3. Participants' descriptions of their auditory-perceptual evaluation of strain.

Participant	Description of strain
SLP02	What I hear as strain is … a patient sounding like it's effortful. It's sort of like a like a pushed quality. It just sounds like they're working hard.
SLP03	I'm listening for potentially a shift in resonance, from more frontal oral resonance to a more kind of back and down … more of a throaty or throat-focused resonance. Low resonance
SLP05	Like, there's a loss of resonance of the sound. It sounds like it's missing harmonics. I feel like if were to pull up a spectrogram all the harmonics would be gone and there'd just be a couple.
SLP08	Something that sounds like or feels like a pushing
SLP09	Perception of effort or work essentially, or does it sound easy or does it sound hard
SLP11	Were they straining so much that their resonance changed or was their resonance so back-focused that now they're straining? So, I make comments on their resonance.
SLP13	I'm listening for like a pressed sound and to me that indicates more of like a pharyngeal resonance probably.
SLP14	I listen for poor tone focus, I listen for … like I said, whether that pressed quality is there because the pressed nature of the voice can either be, in my opinion, a pull and a squeeze.
SLP15	I mean definitely I'm listening for it and so things like inability to hold out phrases, like short sentences, forced breathiness, roughness, vocal fry, I think any of those things could be equivalent to—or like you know, related to strain.

Note. SLP = speech-language pathologist.

examples . . . 'Is this like picking up a 10-pound weight, is this like picking up a 50-pound weight, like if you had to give the analogy' and 'what it feels like in your limbs . . . how much effort you're putting out in your throat.'" In van Mersbergen et al.'s (2021) survey on clinicians' assessment of vocal effort, they did not explicitly report on the prevalence of this assessment technique. Instead, the researchers examined which scales were most used. However, they did assess how easily patients explained vocal effort to clinicians and vice versa (per clinician report), implying that open-ended descriptions of vocal effort took place routinely in their voice evaluations when appropriate. Participants also reported using rating scales to assess the patient's perception of vocal effort. The most common type of scale that was reported was a Borg scale, in which vocal effort is rated from 0 to 10. Twenty percent of participants reported using the OMNI-VES (Shoffel-Havakuk, Marks, et al., 2019), a modified Borg scale on which 0 represents that it is *extremely easy* to use their voice and 10 represents that it is *extremely hard* to use their voice, with pictorial representations of increased weight by weightlifters. Forty percent of participants reported using an ordinal scale by asking the patient to rate their voice from 0 or 1 to 10, using no standardized form. Another participant reported using specific questions on the VHI (Jacobson et al., 1997), which contains questions with Likert ratings, specifically mentioning the prompt, "I use a great deal of effort to speak." Two participants reported using direct magnitude estimation scales, asking, "How much effort are you putting out on average compared to normal? Twice as much would be a 2, three times as much is a 3, half as much is [.5]." These reports are also accounted for by van Mersbergen et al. However, the proportion of diagnostic measures reported differs between the results found in the current study. Whereas van Mersbergen et al. found the VHI to be the most commonly used diagnostic measure to quantify vocal effort and Borg scales to be the least used, the current study found the VHI to be the least commonly reported and Borg scales to be the most prevalent. It is important to note that van Mersbergen et al.'s study included a much larger and more varied sample size (N = 86) and, therefore, may be a more accurate representation of practice patterns.

Correlates of vocal effort: Three of the 15 participants mentioned the use of both perceptual and objective measures related to vocal effort. Two participants mentioned that visible tension around the patient's larynx adds to their clinical construct of vocal effort. One participant described, "I'm looking at visible tension . . . that appears to onset at the time voice onsets. So maybe I see like they have a like a tight jaw in general, but if I see, does it get tighter as they start to speak, or can I see the muscles move and the larynx rise as they start to speak? So, I look for some of those things that seem to indicate some extra muscular work." Another participant noted that their observation of the patient's respiration also informs their perception of vocal effort. This includes gasping for breath or running out of breath while speaking.

Despite us providing definitions of strain and vocal effort (Hunter et al., 2020), some participants appeared to use the constructs interchangeably, evidenced by the way in which they reported measuring them. For example, even though clinician-perceived voice quality fell under the definition of strain, two participants reported that auditory perception of vocal effort is something that they consider as well, specifically listening for "stridor" or "perceptually listening for the sound of somebody like lifting a piano." Regarding quantitative correlates of vocal effort, one participant reported hyperfunction via endoscopy as a component of their assessment of patient vocal effort. This same participant was the only participant to report objective measures for vocal effort, which they reported as the same for strain (i.e., subglottal pressure and vocal efficiency). This variability is consistent with prior research that found lack of standardization for a measure of vocal effort (van Mersbergen et al., 2021). Given the variable nature of the assessment of strain and vocal effort and the prevalence of those symptoms as patient complaints, there remains a need to identify quantitative measures that correlate with both clinician perception of strain and patient perception of vocal effort.

Limitations

Due to the open-ended nature of the qualitative prompts, the participants' responses were bound by their memories in answering questions. In other words, for any given question, it is possible that they did not give an exhaustive list of all their opinions. For example, for questions regarding their perceived benefit of acoustic measures, they might have listed three benefits. However, if presented with an additional fourth benefit, they might have agreed with the fourth as an additional benefit, despite not having mentioned it during their initial response. This limitation may apply not only to participants' opinions but also to their recall of their voice evaluation process. Some participants were able to quickly access evaluation forms and protocols to reference, but many described the evaluation structure and acoustic measures from memory. During subsequent parts of the interview, several participants spontaneously remembered a fact that they had unintentionally omitted earlier. Others expressed difficulty in recalling information. Given these instances, it is likely that some information was lost, resulting in a less than complete description of their evaluation process. In future work, direct observations of voice evaluation practices may elucidate more accurate accounts of voice evaluation practices.

Participants included in this study all practiced at voice centers that exclusively treat voice disorders or voice and swallowing and/or airway disorders at the time of their interviews. As such, viewpoints of voice clinicians who work in smaller private practice or primarily telepractice settings were not assessed; however, these clinicians may offer unique opinions on voice evaluations in the context of their respective settings. Future qualitative studies may extend this work by including clinicians from such settings. More work is needed to determine how pervasive the themes that resulted from this study are among a larger pool of voice clinicians across the United States, with the goal of informing continued implementation of acoustic measures into clinical voice assessment, as well as refining clinician training and graduate education on acoustic analysis.

Conclusions

This qualitative study offers detailed accounts of voice evaluation practices, opinions, and beliefs of 15 voice-specialized SLPs within the context of each clinician's work environment. Participants provided details regarding (a) the structure of their voice evaluations, (b) barriers to and benefits of collecting and analyzing acoustic data in a voice evaluation, and (c) methods for assessing strain and vocal effort. Results revealed diverse opinions regarding the use of acoustic measures and to what extent they should be used when making therapeutic decisions and making comparisons across treatment. For those who collect acoustic measures, there was no consensus about which measures are correlated with strain or vocal effort. Future work may include a survey of clinicians across the country to investigate how pervasive these practices, opinions, and beliefs are to inform future implementation of evidence-based objective measures to clinical practice.

Data Availability Statement

Data are not available to share because of the highly personal nature of the interviews and the inability to anonymize the content.

Acknowledgments

This work was funded in part by National Institute on Deafness and Other Communication Disorders Grant R01DC015570 (awarded to Cara E. Stepp). The authors wish to thank Daniel Buckley, Nichole Houle, and Laura Toles who graciously participated in practice interviews and provided pilot data for this project. The authors are also grateful to the 15 voice clinicians in this study for sharing their time, energy, and expertise.

References

American Speech-Language-Hearing Association. (1998). *The roles of otolaryngologists and speech-language pathologists in the performance and interpretation of strobovideolaryngoscopy* [Relevant paper]. https://www.asha.org/policy/rp1998-00132/

American Speech-Language-Hearing Association. (2004). *Preferred practice patterns for the profession of speech-language pathology* [Preferred practice patterns]. https://www.asha.org/policy/pp2004-00191/

American Speech-Language-Hearing Association. (n.d.). *Voice disorders* [Practice portal]. https://www.asha.org/Practice-Portal/Clinical-Topics/Voice-Disorders

Anand, S., Kopf, L. M., Shrivastav, R., & Eddins, D. A. (2019). Objective indices of perceived vocal strain. *Journal of Voice, 33*(6), 838–845. https://doi.org/10.1016/j.jvoice.2018.06.005

Bach, K. K., Belafsky, P. C., Wasylik, K., Postma, G. N., & Koufman, J. A. (2005). Validity and reliability of the Glottal Function Index. *Archives of Otolaryngology—Head & Neck Surgery, 131*(11), 961–964. https://doi.org/10.1001/archotol.131.11.961

Baldner, E. F., Doll, E., & van Mersbergen, M. R. (2015). A review of measures of vocal effort with a preliminary study on the establishment of a vocal effort measure. *Journal of Voice, 29*(5), 530–541. https://doi.org/10.1016/j.jvoice.2014.08.017

Baylor, C., Yorkston, K., Eadie, T., Kim, J., Chung, H., & Amtmann, D. (2013). The Communicative Participation Item Bank (CPIB): Item bank calibration and development of a disorder-generic short form. *Journal of Speech, Language, and Hearing Research, 56*(4), 1190–1208. https://doi.org/10.1044/1092-4388(2012/12-0140)

Behrman, A. (2005). Common practices of voice therapists in the evaluation of patients. *Journal of Voice, 19*(3), 454–469. https://doi.org/10.1016/j.jvoice.2004.08.004

Belafsky, P. C., Mouadeb, D. A., Rees, C. J., Pryor, J. C., Postma, G. N., Allen, J., & Leonard, R. J. (2008). Validity and reliability of the Eating Assessment Tool (EAT-10). *Annals of Otology, Rhinology & Laryngology, 117*(12), 919–924. https://doi.org/10.1177/000348940811701210

Belafsky, P. C., Postma, G. N., & Koufman, J. A. (2002). Validity and reliability of the Reflux Symptom Index (RSI). *Journal of Voice, 16*(2), 274–277. https://doi.org/10.1016/S0892-1997(02)00097-8

Bermúdez de Alvear, R. M., Barón, F. J., & Martínez-Arquero, A. G. (2011). School teachers' vocal use, risk factors, and voice disorder prevalence: Guidelines to detect teachers with current voice problems. *Folia Phoniatrica et Logopaedica, 63*(4), 209–215. https://doi.org/10.1159/000316310

Bhuta, T., Patrick, L., & Garnett, J. D. (2004). Perceptual evaluation of voice quality and its correlation with acoustic measurements. *Journal of Voice, 18*(3), 299–304. https://doi.org/10.1016/j.jvoice.2003.12.004

Birring, S. S., Prudon, B., Carr, A. J., Singh, S. J., Morgan, M. D. L., & Pavord, I. D. (2003). Development of a symptom specific health status measure for patients with chronic cough: Leicester Cough Questionnaire (LCQ). *Thorax, 58*(4), 339–343. https://doi.org/10.1136/thorax.58.4.339

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5*(9), 341–345.

Brungart, D. S., & Scott, K. R. (2001). The effects of production and presentation level on the auditory distance perception of speech. *The Journal of the Acoustical Society of America, 110*(1), 425–440. https://doi.org/10.1121/1.1379730

Chang, A., & Karnell, M. P. (2004). Perceived phonatory effort and phonation threshold pressure across a prolonged voice loading task: A study of vocal fatigue. *Journal of Voice, 18*(4), 454–466. https://doi.org/10.1016/j.jvoice.2004.01.004

Coffey, A., & Atkinson, P. (1996). *Making sense of qualitative data: Complementary research strategies*. Sage.

Cohen, S. M., Jacobson, B. H., Garrett, C. G., Noordzij, J. P., Stewart, M. G., Attia, A., Ossoff, R. H., & Cleveland, T. F. (2007). Creation and validation of the Singing Voice Handicap Index. *Annals of Otology, Rhinology & Laryngology, 116*(6), 402–406. https://doi.org/10.1177/000348940711600602

Enworo, O. C. (2023). Application of Guba and Lincoln's parallel criteria to assess trustworthiness of qualitative research on indigenous social protection systems. *Qualitative Research Journal, 23*(4), 372–384. https://doi.org/10.1108/QRJ-08-2022-0116

Estes, C. M., & Johnson, A. M. (2023). Practical considerations for instrumental acoustic and aerodynamic assessment of voice: Discussion points from an open forum of clinicians. *Perspectives of the ASHA Special Interest Groups, 8*(6), 1354–1362. https://doi.org/10.1044/2023_PERSP-23-00039

Gartner-Schmidt, J. L., Shembel, A. C., Zullo, T. G., & Rosen, C. A. (2014). Development and validation of the Dyspnea Index (DI): A severity index for upper airway–related dyspnea. *Journal of Voice, 28*(6), 775–782. https://doi.org/10.1016/j.jvoice.2013.12.017

Gerratt, B. R., Kreiman, J., Antonanzas-Barroso, N., & Berke, G. S. (1993). Comparing internal and external standards in voice quality judgments. *Journal of Speech and Hearing Research, 36*(1), 14–20. https://doi.org/10.1044/jshr.3601.14

Gilman, M., & Johns, M. M. (2017). The effect of head position and/or stance on the self-perception of phonatory effort. *Journal of Voice, 31*(1), 131.E1–131.E4. https://doi.org/10.1016/j.jvoice.2015.11.024

Grossoehme, D. H. (2014). Overview of qualitative research. *Journal of Health Care Chaplaincy, 20*(3), 109–122. https://doi.org/10.1080/08854726.2014.925660

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. https://doi.org/10.1016/S0892-1997(02)00067-X

Hillman, R. E., Montgomery, W. W., & Zeitels, S. M. (1997). Current diagnostics and office practice: Appropriate use of objective measures of vocal function in the multidisciplinary management of voice disorders. *Current Opinion in Otolaryngology & Head and Neck Surgery, 5*(3), 172–175. https://doi.org/10.1097/00020840-199706000-00005

Hirano, M. (1981). Psycho-acoustic evaluation of voice. In M. Hirano (Ed.), *Clinical examination of voice* (pp. 81–84). Springer-Verlag.

Hogikyan, N. D., & Sethuraman, G. (1999). Validation of an instrument to measure Voice-Related Quality of Life (V-RQOL). *Journal of Voice, 13*(4), 557–569. https://doi.org/10.1016/S0892-1997(99)80010-1

Hunter, E. J., Cantor-Cutiva, L. C., van Leer, E., van Mersbergen, M., Nanjundeswaran, C. D., Bottalico, P., Sandage, M. J., & Whitling, S. (2020). Toward a consensus description of vocal effort, vocal load, vocal loading, and vocal fatigue. *Journal of Speech, Language, and Hearing Research, 63*(2), 509–532. https://doi.org/10.1044/2019_JSLHR-19-00057

Isetti, D., Xuereb, L., & Eadie, T. L. (2014). Inferring speaker attributes in adductor spasmodic dysphonia: Ratings from unfamiliar listeners. *American Journal of Speech-Language Pathology, 23*(2), 134–145. https://doi.org/10.1044/2013_AJSLP-13-0010

Jacobson, B. H., Johnson, A., Grywalski, C., Silbergleit, A., Jacobson, G., Benninger, M. S., & Newman, C. W. (1997). The Voice Handicap Index (VHI): Development and validation. *American Journal of Speech-Language Pathology, 6*(3), 66–70. https://doi.org/10.1044/1058-0360.0603.66

Kay Elemetrics. (1993). *Multi-Dimensional Voice Program (MDVP)* [Computer program].

Kay Elemetrics. (1994). *CSL Computerized Speech Lab: Model 4300B* (Version 5.X) [Computer software].

Kempster, G., Gerratt, B., Barkmeier-Kraemer, J., & Hillman, R. (2009). Consensus Auditory-Perceptual Evaluation of Voice: Development of a standardized clinical protocol. *American Journal of Speech-Language Pathology, 18*(2), 124–132. https://doi.org/10.1044/1058-0360(2008/08-0017)

Kreiman, J., Gerratt, B. R., Kempster, G. B., Erman, A., & Berke, G. S. (1993). Perceptual evaluation of voice quality: Review, tutorial, and a framework for future research. *Journal of Speech and Hearing Research, 36*(1), 21–40. https://doi.org/10.1044/jshr.3601.21

Kreiman, J., Gerratt, B. R., Precoda, K., & Berke, G. S. (1992). Individual differences in voice quality perception. *Journal of Speech and Hearing Research, 35*(3), 512–520. https://doi.org/10.1044/jshr.3503.512

Lee, M., Drinnan, M., & Carding, P. (2005). The reliability and validity of patient self-rating of their own voice quality. *Clinical Otolaryngology, 30*(4), 357–361. https://doi.org/10.1111/j.1365-2273.2005.01022.x

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage. https://us.sagepub.com/en-us/nam/book/naturalistic-inquiry

Lodhavia, A., & Kempster, G. B. (2024). Fidelity to the Consensus Auditory-Perceptual Analysis of Voice (CAPE-V): A pilot study. *Journal of Voice*. Advance online publication. https://doi.org/10.1016/j.jvoice.2023.12.009

Lowell, S. Y., Kelley, R. T., Awan, S. N., Colton, R. H., & Chan, N. H. (2012). Spectral- and cepstral-based acoustic features of dysphonic, strained voice quality. *Annals of Otology, Rhinology & Laryngology, 121*(8), 539–548. https://doi.org/10.1177/000348941212100808

Marks, K. L., Verdi, A., Toles, L. E., Stipancic, K. L., Ortiz, A. J., Hillman, R. E., & Mehta, D. D. (2021). Psychometric analysis of an ecological vocal effort scale in individuals with and without vocal hyperfunction during activities of daily living. *American Journal of Speech-Language Pathology, 30*(6), 2589–2604. https://doi.org/10.1044/2021_AJSLP-21-00111

Maryn, Y. (2017). Practical acoustics in clinical voice assessment: A Praat primer. *Perspectives of the ASHA Special Interest Groups, 2*(3), 14–32. https://doi.org/10.1044/persp2.SIG3.14

McKenna, V. S., Diaz-Cadiz, M. E., Shembel, A. C., Enos, N. M., & Stepp, C. E. (2019). The relationship between physiological mechanisms and the self-perception of vocal effort. *Journal of Speech, Language, and Hearing Research, 62*(4), 815–834. https://doi.org/10.1044/2018_JSLHR-S-18-0205

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. https://doi.org/10.1097/MOO.0b013e3282fe96ce

Morrison, M. D., & Rammage, L. A. (1993). Muscle misuse voice disorders: Description and classification. *Acta Oto-Laryngologica, 113*(3), 428–434. https://doi.org/10.3109/00016489309135839

Morse, J. M. (2015). Critical analysis of strategies for determining rigor in qualitative inquiry. *Qualitative Health Research, 25*(9), 1212–1222. https://doi.org/10.1177/1049732315588501

Nagle, K. F. (2016). Emerging scientist: Challenges to CAPE-V as a standard. *Perspectives of the ASHA Special Interest Groups, 1*(3), 47–53. https://doi.org/10.1044/persp1.SIG3.47

Nagle, K. F. (2025). Clinical use of the CAPE-V scales: Agreement, reliability and notes on voice quality. *Journal of Voice 39*(3)*,* 685–698. https://doi.org/10.1016/j.jvoice.2022.11.014

Paes, S. M., & Behlau, M. (2017). Dosage dependent effect of high-resistance straw exercise in dysphonic and non-dysphonic women. *CoDAS, 29*(1), Article e20160048. https://doi.org/10.1590/2317-1782/20172016048

Patel, R. R., Awan, S. N., Barkmeier-Kraemer, J., Courey, M., Deliyski, D., Eadie, T., Paul, D., Švec, J. G., & Hillman, R. (2018). Recommended protocols for instrumental assessment of voice: American Speech-Language-Hearing Association expert panel to develop a protocol for instrumental assessment of vocal function. *American Journal of Speech-Language Pathology, 27*(3), 887–905. https://doi.org/10.1044/2018_AJSLP-17-0009

Phyland, D. J., Pallant, J. F., Thibeault, S. L., Benninger, M. S., Vallance, N., & Smith, J. A. (2015). Measuring vocal function in professional music theater singers: Construct validation of the Evaluation of the Ability to Sing Easily (EASE). *Folia Phoniatrica et Logopaedica, 66*(3), 100–108. https://doi.org/10.1159/000366202

Rosen, C. A., Lee, A. S., Osborne, J., Zullo, T., & Murry, T. (2004). Development and validation of the Voice Handicap Index-10. *The Laryngoscope, 114*(9), 1549–1556. https://doi.org/10.1097/00005537-200409000-00009

Rosenthal, A. L., Lowell, S. Y., & Colton, R. H. (2014). Aerodynamic and acoustic features of vocal effort. *Journal of Voice, 28*(2), 144–153. https://doi.org/10.1016/j.jvoice.2013.09.007

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*(2), 212–226. https://doi.org/10.1044/1058-0360(2012/12-0014)

Salgado, S., Schils, S. A., Childes, J. M., Crino, C., & Palmer, A. D. (2024). Current practices in the assessment of voice: A comparison of providers across different clinical settings. *Journal of Voice*. Advance online publication. https://doi.org/10.1016/j.jvoice.2024.08.007

Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., Burroughs, H., & Jinks, C. (2018). Saturation in qualitative research: Exploring its conceptualization and operationalization. *Quality & Quantity, 52*(4), 1893–1907. https://doi.org/10.1007/s11135-017-0574-8

Schwartz, S. R., Cohen, S. M., Dailey, S. H., Rosenfeld, R. M., Deutsch, E. S., Gillespie, M. B., Granieri, E., Hapner, E. R., Kimball, C. E., & Krouse, H. J. (2009). Clinical practice guideline: Hoarseness (dysphonia). *Otolaryngology–Head and Neck Surgery, 141*(Suppl. 1), 1–31. https://doi.org/10.1016/j.otohns.2009.06.744

Shembel, A. C., Rosen, C. A., Zullo, T. G., & Gartner-Schmidt, J. L. (2013). Development and validation of the Cough Severity Index: A severity index for chronic cough related to the upper airway. *The Laryngoscope, 123*(8), 1931–1936. https://doi.org/10.1002/lary.23916

Shoffel-Havakuk, H., Chau, S., Hapner, E. R., Pethan, M., & Johns, M. M., III. (2019). Development and validation of the Voice Catastrophization Index. *Journal of Voice, 33*(2), 232–238. https://doi.org/10.1016/j.jvoice.2017.09.026

Shoffel-Havakuk, H., Marks, K. L., Morton, M., Johns, M. M., III, & Hapner, E. R. (2019). Validation of the OMNI Vocal Effort Scale in the treatment of adductor spasmodic dysphonia. *The Laryngoscope, 129*(2), 448–453. https://doi.org/10.1002/lary.27430

Stachler, R. J., Francis, D. O., Schwartz, S. R., Damask, C. C., Digoy, G. P., Krouse, H. J., McCoy, S. J., Ouellette, D. R., Patel, R. R., & Reavis, C. W. (2018). Clinical practice guideline: Hoarseness (dysphonia) (update). *Otolaryngology–Head and Neck Surgery, 158*(Suppl. 1), S1–S42. https://doi.org/10.1177/0194599817751030

van Mersbergen, M., Beckham, B. H., & Hunter, E. J. (2021). Do we need a measure of vocal effort? Clinician's report of vocal effort in voice patients. *Perspectives of the ASHA Special Interest Groups, 6*(1), 69–79. https://doi.org/10.1044/2020_PERSP-20-00258

Verdolini, K., Titze, I. R., & Fennell, A. (1994). Dependence of phonatory effort on hydration level. *Journal of Speech and Hearing Research, 37*(5), 1001–1007. https://doi.org/10.1044/jshr.3705.1001

Webb, A. L., Carding, P. N., Deary, I. J., MacKenzie, K., Steen, N., & Wilson, J. A. (2004). The reliability of three perceptual evaluation scales for dysphonia. *European Archives of Oto-Rhino-Laryngology and Head & Neck, 261*(8), 429–434. https://doi.org/10.1007/s00405-003-0707-7

Zraick, R. I., Kempster, G. B., Connor, N. P., Thibeault, S., Klaben, B. K., Bursac, Z., Thrush, C. R., & Glaze, L. E. (2011). Establishing validity of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V). *American Journal of Speech-Language Pathology, 20*(1), 14–22. https://doi.org/10.1044/1058-0360(2010/09-0105)