

2024

Clinical translation of an acoustic measure of vocal strain: a mixed methods study

<https://hdl.handle.net/2144/48764>

"Downloaded from OpenBU. Boston University's institutional repository."

BOSTON UNIVERSITY
SARGENT COLLEGE OF HEALTH AND REHABILITATION SCIENCES

Thesis

**CLINICAL TRANSLATION OF AN
ACOUSTIC MEASURE OF VOCAL STRAIN:
A MIXED METHODS STUDY**

by

MICHAEL D. MADOULE

B.A., Boston College, 2022

Submitted in partial fulfillment of the
requirements for the degree of
Master of Science

2024

Approved by

First Reader

Cara E. Stepp Ph.D.
Professor of Speech, Language & Hearing Sciences
Sargent College of Health and Rehabilitation Sciences

Professor of Biomedical Engineering
College of Engineering

Professor of Otolaryngology – Head and Neck Surgery
Aram V. Chobanian & Edward Avedisian School of Medicine

Second Reader

Katherine L. Marks, Ph.D., CCC-SLP
Postdoctoral Research Fellow
Stepp Lab for Sensorimotor Rehabilitation Engineering

Third Reader

Kathleen F. Nagle, Ph.D., CCC-SLP
Associate Professor of Speech-Language Pathology
Seton Hall University

Fourth Reader

Jennifer M. Vojtech, Ph.D.
Research Assistant Professor of Speech, Language & Hearing Sciences

Fifth Reader

Daniel P. Buckley, M.S., CCC-SLP
Lecturer in Speech, Language & Hearing Sciences
Sargent College of Health and Rehabilitation Sciences

Assistant Professor of Otolaryngology – Head and Neck Surgery
Aram V. Chobanian & Edward Avedisian School of Medicine

ACKNOWLEDGMENTS

I would first like to thank Dr. Cara Stepp who served as my thesis advisor. From the beginning, Cara embraced my ideas and helped me fine-tune them into the research presented here. As someone new to research, I am forever grateful for the guidance and feedback you have given me throughout this whole experience. Above all, you have been so supportive and encouraging, and are the reason that I have the opportunity to continue on investigating my research questions beyond graduation. Words cannot express my gratitude. I would also like to thank Dr. Katie Marks for her invaluable mentorship. You have been incredibly generous with your time in supporting my growth. You have taught me skills that I will carry with me for my life-time, and I am lucky to have been able to research under your guidance.

Additionally, I owe an incredible thanks to Daniel Buckley whose passion for the field of voice led me to my thesis topic. To say you sparked my interest in voice would be an understatement. As I am sure anyone who has met Daniel would agree, his enthusiasm is infectious, and I hope to inspire others in the future the way that Daniel has inspired me. I would like to also thank the other members of my thesis committee: Drs. Kathy Nagle and Jenny Vojtech. Kathy taught me how to perform qualitative analysis from the ground-up. In many ways, this portion of my research became the most important and interesting aspect of my project. Without your commitment and encouragement as a mentor and fellow coder, this work would not be possible. Similarly, as the major contributor to the software on which my research is based, this project would not have been possible without Jenny's incredible knowledge and ingenuity. Your

commitment to advancing the assessment and treatment of sensorimotor disorders of the voice, is one that I hope to emulate in the future.

Finally, I would like to thank my family and friends for the positivity and support they have shown me over the last year and a half. What my family lacked in knowledge about acoustics, they made up for in their ever-present love and support throughout this project. They helped the most when the lows were low and celebrated me generously when the highs were high. To my cohort, thank you for your friendship and your continual curiosity in my research. And to all the members of the STEPP Lab that engaged without me in any capacity during this project, having the opportunity to explain my successes and frustrations to others that shared similar experiences has made all the difference.

**CLINICAL TRANSLATION OF AN
ACOUSTIC MEASURE OF VOCAL STRAIN:
A MIXED METHODS STUDY**

MICHAEL D. MADOULE

ABSTRACT

Purpose: The purpose of this study was to assess the accuracy of a newly developed software, the Automatic RFF Calculator (ARC), in measuring relative fundamental frequency (RFF) compared to current semi-automated RFF estimation algorithms (aRFF). To gain a better understanding of how ARC might fit into clinical voice evaluations, the general structure of voice evaluations across clinicians was investigated in addition to their opinions on the use of acoustic assessment as well. Furthermore, the study sought clinician feedback on improvements to ARC that would make it more applicable to the clinical setting.

Method: The first research question was assessed by comparing the effect sizes of RFF values calculated via ARC with those calculated by aRFF. To address the remaining research questions, one-on-one interviews with clinicians were conducted during which they were probed specifically for their assessment of strain and vocal effort as well as their opinions on the benefits and barriers of using acoustic measures in the voice evaluation process. Additionally, during the interviews, the clinicians were introduced to ARC and RFF and then prompted to share their opinions on which features of ARC they enjoyed as well as features they thought should be added or adjusted.

Results: This study found that there were no statistically significant differences between the effect sizes calculated via ARC and aRFF. Additionally, the clinicians provided useful feedback on improvements to be made to ARC to increase its clinical utility. Most clinicians who already include acoustic assessment in their voice evaluations were willing to incorporate ARC and RFF into their clinical protocol, whereas those who do not use acoustic assessment were less willing to use the measure. Across interviews, clinicians reported a desire to see more research surrounding the clinical utility of ARC and RFF.

Conclusion: The quantitative results suggests that RFF estimation via ARC is as accurate as the current semi-automated estimation method. These results implicate ARC as a useful clinical tool as an acoustic correlate of vocal strain and/or effort. The feedback gleaned from the clinician interviews is beneficial in directing the future development of ARC.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	iv
ABSTRACT.....	vi
TABLE OF CONTENTS.....	viii
LIST OF TABLES	x
LIST OF FIGURES	xi
Introduction.....	1
Quantitative Methods.....	14
Participants	14
Protocol	17
Data Analysis	18
Statistical Analysis	20
Quantitative Results	20
Quantitative Discussion	24
Qualitative Methods.....	25
Participants	26
Protocol	26
Analysis.....	30
Qualitative Results	32
Evaluation Structure.....	32
Benefits and Barriers to Acoustics	38
Evaluation of Strain and Vocal Effort.....	48
Opinions on RFF and ARC	53
Qualitative Discussion	65
Evaluation Structure.....	65
Benefits and Barriers to Acoustics	67

Evaluation of Strain and Vocal Effort.....	71
Opinions on RFF and ARC	72
Limitations and Future Study Directions	74
Conclusion	78
APPENDIX.....	80
BIBLIOGRAPHY.....	82
CURRICULUM VITAE.....	91

LIST OF TABLES

Table	Title	Page
1	Participant Demographics and Overall Severity of Dysphonia Rating of the Effort Group	15
2	Participant Demographics and Overall Severity of Dysphonia Rating of the Therapy Group	16
3	Cohen's <i>d</i> Values for Offset 10 and Onset 1 Using ARC and aRFF for the Effort Group	22
4	Cohen's <i>d</i> Values for Offset 10 and Onset 1 Using ARC and aRFF for the Therapy Group	23
5	Summary of the Acoustic and Aerodynamic Measures in Each Participant's Voice Evaluation	36–37
6	Themes and Sub-themes Describing the Benefits and Barriers of Performing Acoustics	38
7	Themes and Sub-themes Describing the Evaluation of Strain and Vocal Effort	48
8	Participants' Descriptions of their Auditory-Perceptual Evaluation of Strain	50

LIST OF FIGURES

Figure	Title	Page
1	Mean relative fundamental frequency (RFF) values for control participants and participants with vocal hyperfunction	10
2	The effect sizes for Offset Cycle 10 and Onset Cycle 1 for ARC and aRFF for the Effort Group. Error bars indicate 95% confidence intervals.	24
3	The effect sizes for Offset Cycle 10 and Onset Cycle 1 between ARC and aRFF for the Therapy Group. Error bars indicate 95% confidence intervals.	24
4	Voice evaluation checklist used while conducting qualitative interviews.	27
5	Voice evaluation checklist for specification used while conducting qualitative interviews.	28
6	The OMNI-Vocal Effort Scale	52
7	ARC Dashboard	54
8	Waveform and RFF Graph of an Analyzed Sample in ARC	54
9	RFF Graph and Value Table of an Analyzed Sample in ARC	55

Introduction

Vocal hyperfunction is a type of voice disorder characterized by increased muscle tension in and around the laryngeal framework (Dworkin, Meleca, & Abkarian, 2000; Koufman & Blalock, 1991; Morrison et al., 1983; Roy et al., 1996). When vocal hyperfunction presents without any apparent pathology on or around the vocal folds (e.g., nodules or polyps), it is referred to as non-phonotraumatic vocal hyperfunction, commonly called primary muscle tension dysphonia (MTD; Hillman et al., 2020). The excessive muscle tension can result in dysphonic voice production that is often perceived as “strained” by the listener and experienced as “vocal effort” by the speaker (Hunter et al., 2020). For clinicians, longitudinal monitoring of a patient’s voice allows them to better track patient progress throughout the course of voice therapy, which is the recommended treatment for MTD. Traditionally, patient-reported outcomes or clinician-rated auditory-perceptual assessment are used for tracking changes in vocal function. Common judgments include grade/overall severity, roughness, breathiness, asthenia, and strain (Minoru Hirano, 1981; Kempster et al., 2009). Unfortunately, in both cases, these perceptual ratings are unreliable. Lee et al. (2005) found that, when using the Grade, Roughness, Breathiness, Asthenia, Strain scale (GRBAS; M Hirano, 1981), not only did patients have lower intra-rater reliability than clinicians, but the inter-rater agreement between patients and clinicians was worse than chance. Furthermore, even trained listeners, like speech-language pathologists (SLPs), have been shown to frequently disagree when voice quality is rated due to the subjective nature of perceptual judgment (Nagle, 2022). The lack of agreement is amplified in auditory-perceptual ratings of strain,

as reflected by worse intra- and inter-rater reliability compared to those of other vocal qualities, such as breathiness or overall severity (Webb et al., 2004; Zraick et al., 2011).

There are several acoustic measures that are currently used in a standard voice evaluation that may correlate with auditory-perceptual judgments of vocal quality. One such objective measure is harmonics-to-noise ratio (HNR; Yumoto et al., 1982). During phonation, there are two components of the sound that exit the vocal tract. One is the periodic vibration of the vocal folds that form a harmonic series. The other component is the aperiodic noise that comes from turbulent air in the vocal tract. HNR is the ratio between the periodic vibration of the vocal folds and any aperiodic sounds. When a typical speaker produces a vowel, the HNR is typically high due to the lack of turbulence. However, any turbulent airflow caused by vocal hyperfunction or pathology at the glottis can reduce the periodic component. This results in a lower HNR and is associated with a breathy quality in the voice (Krom, 1995). Other clinical objective measures often used in a voice evaluation include habitual vocal sound pressure level (SPL), vocal fundamental frequency (f_0), and cepstral-spectral measures (Patel et al., 2018). Mean, minimum, and maximum vocal SPL (in decibels [dB SPL]) serve as quantitative acoustic correlates of vocal loudness. Mean f_0 is the acoustic correlate of vocal pitch (Patel et al., 2018), and the standard deviation (SD), minimum, and maximum f_0 are acoustic correlates of prosody. Finally, cepstral peak prominence (CPP) is an acoustic clinical measure that has been empirically studied as a correlate of voice quality (Patel et al., 2018). Specifically, CPP has been strongly associated with auditory-perceptual ratings of breathiness as well as overall severity of dysphonia (Heman-Ackah et al., 2002). These measures are useful

tools, as they provide an objective representation of the auditory-perceptual parameters that have previously demonstrated poor reliability, and they allow clinicians to assess voice quality more consistently and efficiently.

The terms “vocal effort” and “strain” represent two different but related perspectives of the same phenomenon. Hunter et al. (2020) discussed several aspects that encompass vocal effort found in the existing literature, including an increase in loudness or strain in voicing (Brandt et al., 1969; Lien et al., 2015). The researchers also discussed how certain changes in posture were associated with reports of increased vocal effort in the literature (Gilman & Johns, 2017). Some definitions included the presence of discomfort while speaking, tightness in the throat, or tearful-sounding speech (Isetti et al., 2014; Paes & Behlau, 2017). Additionally, researchers have also found that mood changes and cognitive load play a role in vocal effort (Baldner et al., 2015). Finally, 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). Going forward, Hunter et al. (2020) recommend the definition of vocal effort as “the perceived exertion of a vocalist’s response to a perceived communication scenario” (a perceptual phenomenon experienced by the speaker and not the listener) and strain as the auditory-perceptual judgment of vocal effort by a listener.

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). One of the most common ways

clinicians reported quantifying vocal effort is by using a visual analog scale (VAS). When using this type of scale, the rater indicates the intensity of a perceptual phenomenon along a line. Some use this scale without a physical stimulus, whereas others use anchors (e.g. 0 = no effort, 100 = severe effort; Marks et al., 2021). One issue with visual analog scales is that they cannot be mathematically manipulated, so statistical inferences cannot be readily made (van Mersbergen et al., 2021). The direct magnitude estimation (DME) scale is a nonlinear perceptual scale that has been used to measure vocal effort in studies such as Chang and Karnell (2004) and Verdolini et al. (1994). When using a DME scale, the perceiver gives a numerical judgment for the sensory level of a physical stimulus, without a visual line such as those used with VAS. This judgment can then be adjusted to correspond to different perceptions. The Borg category-ratio (Borg CR) scale is another scale that can be used to measure effort (Baldner et al., 2015). It combines components of the VAS and DME scales, namely the usability from the former and the mathematical precision from the latter. The Borg CR10 has demonstrated sensitivity to changes in mood (Bermúdez de Alvear et al., 2011; Brungart & Scott, 2001; van Mersbergen & Delany, 2014; van Mersbergen et al., 2017; Van Mersbergen et al., 2008), cognitive status (van Mersbergen et al., 2020), and post-therapy outcomes (van Leer & van Mersbergen, 2017). However, the Borg CR10 lacks granularity that is required to detect small changes in vocal effort. Likert scales (and Likert-type scales) involve the participant rating their level of agreement with a statement. They are usually given choices such as “strongly agree” and “strongly disagree” at the extremes as well as a neutral option at the midpoint. These types of scales enforce strict criteria when

assigning the rating levels such that mathematical comparisons can be made between statements as well as between individuals (van Mersbergen et al., 2021).

Recently, van Mersbergen et al. (2021) conducted a survey to further investigate how clinicians measure vocal effort. Based on survey results, most clinicians used the Voice Handicap Index (VHI), despite the VHI measuring a different construct (i.e. voice disability) and only containing three questions related to vocal effort. The researchers also found that the Vocal Fatigue Index (a Likert scale) was another scale used by clinicians to assess vocal effort (van Mersbergen et al., 2021). Vocal strain as an auditory-perceptual phenomenon is most often rated using a VAS like the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V; Kempster et al., 2009) or an ordinal scale like the GRBAS scale (M Hirano, 1981). Despite the range of perceptual scales available, there is still no standard way to quantify vocal effort and/or strain (van Mersbergen et al., 2021). Additionally, using these scales presents the same problem described previously: subjective judgments by both clinicians and patients are unreliable (Lee et al., 2005). Until an objective correlate is adopted clinically, tracking vocal effort and/or strain throughout therapy will remain unstandardized.

Just as with breathiness and overall severity of dysphonia, many researchers have focused their attention on finding an objective measure of vocal effort and/or strain. Generally, these objective measures 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 (MFDR) are significantly

increased during maximal effort speech as compared to speech at comfortable effort levels in healthy speakers. 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. Mediolateral supraglottic compression (observed during endoscopy) and normalized percent activation of the suprahyoid muscles (measured using electromyography) were found to be physiological predictors of vocal effort, as well (McKenna et al., 2019). The findings in these studies are significant; however, the measures that were tested are of limited clinical use. In addition to these measures being more invasive than acoustic measures to obtain, the tools necessary to perform aerodynamic, EMG, or endoscopic evaluations are not available to all clinicians.

Acoustically, CPP measures have been shown to be significantly higher during maximal effort speech compared to comfortable effort speech (Rosenthal et al., 2014). However, as discussed earlier, CPP has also been used to quantify breathiness and overall severity, and is therefore not specific to strain or vocal effort. Similarly, Lowell et al. (2012) found that some cepstral- and spectral-based measures are effective in distinguishing strain from unstrained voices, but the same limitation exists; these measures have been used to characterize breathiness and roughness and are thus not specific to strain.

Another acoustic measure that is associated with vocal effort is relative fundamental frequency (RFF); however, RFF is currently only used in research and has

not yet been implemented clinically. During a vowel-voiceless consonant-vowel (VCV) production, there is a change in the instantaneous f_o for each individual voicing cycle prior to and following the consonant (Vojtech et al., 2019). This change is captured as RFF, which is measured using the 10 cycles prior to (“offset cycles”) and following (“onset cycles”) the voiceless consonant; therefore, an RFF estimate from one speaker is composed of 20 individual cycles which can either be analyzed individually as an estimate of laryngeal tension or in reference to each other to plot out a particular speaker’s trends in devoicing/voicing. When reporting RFF values, each voicing cycle is labeled serially from 1 to 10 on either side of the intervocalic consonant. The stimuli typically used for estimating RFF include the VCV utterances /ifi/, /afa/, and /ufu/ (Lien et al., 2014).

The RFF value of any one voicing cycle is estimated by first calculating the instantaneous f_o of the cycle in Hertz (Hz). This instantaneous f_o is normalized relative to the voicing cycle closest to the steady-state vowel. For the offset cycles, the pseudo steady-state f_o is represented by Offset Cycle 1, whereas for the onset cycles, this value is represented by Onset Cycle 10. Finally, RFF is plotted in semitones (ST) relative to a steady-state f_o with Offset Cycle 1 and Onset Cycle 10 set as 0 ST, using the following Equation 1:

$$RFF (ST) = 12 \times \log_2 \left(\frac{f_o}{f_o^{ref}} \right)$$

The conversion and normalization allow for comparisons to be made both within the same speaker and across different speakers. Employing this standardization, the RFF

values in ST represent a change in instantaneous f_0 of any given voicing cycle from the steady-state vowel. Therefore, an RFF value of 0 ST indicates no change in f_0 , whereas a positive RFF value indicates a higher f_0 for a particular cycle relative to the steady-state vowel (quicker vocal fold vibration for that one cycle) and a negative RFF value indicates a lower f_0 relative to the steady-state vowel (slower vocal fold vibration for that one cycle).

Initial calculations of RFF were made using a manual technique that involved recording an audio sample and using a speech analysis software such as Praat to line up generated pulses with the peaks or troughs of the waveform by hand (Boersma, 2001). For manual RFF calculation, a trained technician must analyze the transitional waveforms into and out of devoicing to identify the boundary cycles; then, the 10 closest vocal cycles to voiceless consonants on either side can be determined. Using software like Praat (Boersma, 2001) to identify the boundary cycles and estimate the 10 closest vocal cycles is tedious as the technician must consider environmental noise, concurrent aspiration, and frication from coarticulation when trying to identify the boundary cycles (Vojtech et al., 2019). The tedious nature of manual estimation results in trained technician requiring 20–40 minutes to obtain one set of usable RFF values (Vojtech et al., 2019).

Improving upon the manual method, Lien et al. (2017) developed a semi-automated RFF algorithm that removed much of the manual work that characterized the previous method. The semi-automated RFF algorithm is implemented in a custom MATLAB script that identifies the 10 vocal cycles prior to and following the voiceless consonant, computes the inverse of each period, and compares the calculated f_0 in ST to

the steady-state f_0 for normalization (Lien & Stepp, 2013). This program allows the technician to adjust the cycles for more accurate estimates when needed. Lien et al. (2017)'s semi-automated algorithm was further refined by Vojtech et al. (2019), allowing for differences in voice sample characteristics to be taken into account (referred to aRFF-AP; current version). The algorithm accounts for overall severity of dysphonia and signal acquisition quality using pitch strength — “the degree of tonality in a sound or the saliency of pitch sensation on a scale of weak to strong” (Kopf et al., 2017) — as an objective measure of both characteristics (Vojtech et al., 2019). Thus, rather than use the same acoustic feature thresholds to identify boundary cycles across samples, a voice sample can be evaluated according to its specific attributes using thresholds most optimal for that voice sample (Vojtech et al., 2019).

The RFF pattern for typical healthy speakers has been well established. The offset cycles start at 0 ST (the reference Offset 1 cycle) and remain relatively stable in younger speaker and decrease slightly in older speakers closer to the voiceless consonant (Watson, 1998). Conversely, the onset cycles start high and decrease to 0 ST (the reference Offset 10 cycle; Robb & Smith, 2002; Watson, 1998). This typical RFF pattern is depicted in Figure 1 by the “control” data points. This pattern is hypothesized to be due to interactions between the aerodynamics of the system and the tension present in the laryngeal muscles when abducting and adducting the vocal folds for the production and cessation of the voiceless consonant, respectively (Stepp et al., 2010). RFF has also been used to examine purported differences in voice mechanisms between VH and controls. In contrast to typical speakers, for individuals with vocal hyperfunction, on average, the

RFF values for all cycles are decreased compared to typical speakers, most evident for those closest to the voiceless consonant (Stepp et al., 2010). As depicted in Figure 1, speakers with vocal hyperfunction show a similar RFF pattern of gradual decrease toward Offset Cycle 10 and Onset Cycle 1. However, the RFF values for the three vocal hyperfunction groups (polyps, nodules, and MTD) are relatively lower than the control group for both the onset and offset cycles. Researchers hypothesize that the reason for this lowered RFF pattern is the already-present tension in the laryngeal muscles of those with vocal hyperfunction (Stepp et al., 2010). Because vocal hyperfunction is characterized by excessive muscle tension which speakers often report as sensation of vocal effort, these relatively decreased RFF values represent a numerical reflection of the increased effort.

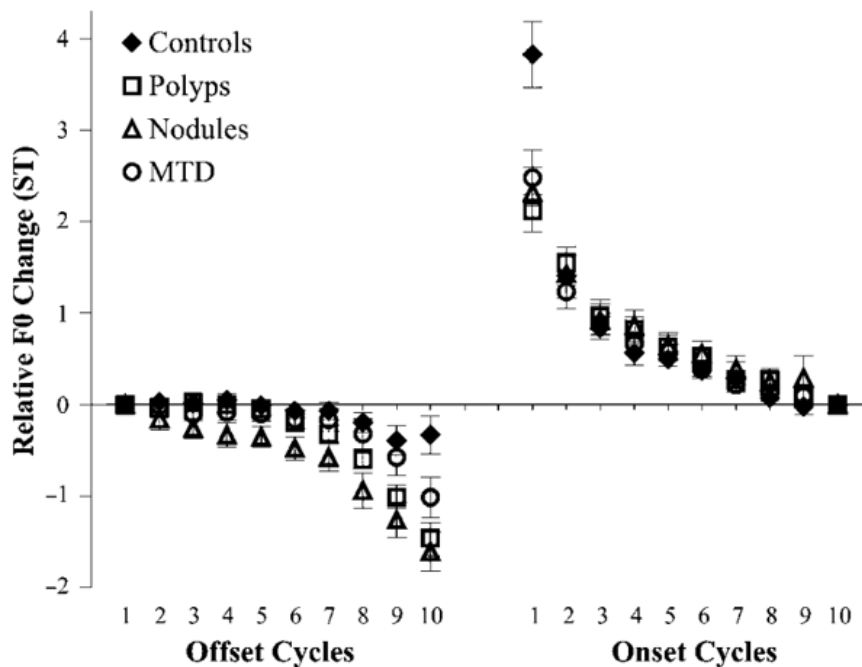


Figure 1. Mean relative fundamental frequency (RFF) values for control participants and participants with vocal hyperfunction

How RFF patterns change as a result of voice therapy and voluntary fluctuations in effort has also been studied. Stepp et al. (2011) compared pre- and post-therapy measurements of RFF in 16 participants who had voice disorders related to vocal hyperfunction. Their results showed that participants had lower than typical RFF values prior to therapy but that these values increased towards the pattern typically seen in individuals without voice disorders post-therapy. Additionally, Lien et al. (2015) found that RFF changed in typical speakers producing voice at different levels of vocal effort. To ensure modulation in effort, the participants were given feedback from the researcher in addition to real-time visual feedback based on their intraoral pressure and SPL; they were instructed to increase their intraoral pressure while keeping a constant SPL to increase the strain in their voice. When participants spoke with a typical voice, their RFF patterns were similar to those observed in healthy speakers from previous studies (Robb & Smith, 2002; Watson, 1998). However, when they were asked to speak with greater levels of vocal effort, their offset RFF values decreased whereas their onset RFF remained stable. This pattern matches those observed in individuals with vocal hyperfunction (Stepp et al., 2010). Additionally, McKenna et al. (2016) found that RFF predicted kinematic estimates of laryngeal stiffness in speakers without voice disorders. In support of these findings, Park et al. (2020) found that by synthetically lowering the RFF values of typical speakers, listener-perceived strain increased. In contrast, by synthetically raising the RFF values of participants speaking with greater levels of vocal effort, perceived strain decreased.

Although these findings are promising, they demonstrate the efficacy of RFF values as an objective measure of strain only in controlled laboratory environments. There are several reasons that RFF measurements have yet to be translated into clinical application. One major obstacle is the aperiodic nature of dysphonic voices. Time-based acoustic measures such as RFF require specification of where the cycle-to-cycle boundaries are during voicing. The identification of these boundaries becomes increasingly more difficult as the signal becomes more aperiodic (McKenna & Stepp, 2018). Therefore, in cases of dysphonia caused by severe hyperfunction, the signal may lack the periodicity required to properly analyze the VCV segments (Stepp et al., 2011). Consistent with this hypothesis, Roy et al. (2016) found that as the severity of dysphonia increased, the number of usable RFF instances decreased. Furthermore, Vojtech et al. (2019) noted that, although not an inherent issue with RFF measurements, the lack of RFF research on sample populations that are representative of those in the clinical practice detracts from the clinical relevance of RFF (Vojtech et al., 2019). They note that in their study, 3% of the examined population was diagnosed with vocal polyps, whereas the prevalence of vocal polyps in adults with voice disorders was closer to 12% by other estimates (Martins et al., 2016). Similarly, they report that their study population was made up of 37% individuals with Parkinson's disease and 33% individuals with muscle tension dysphonia, which may bias their results toward those two groups (Vojtech et al., 2019).

Another important contributor to the lack of clinical adoption of RFF measurements is that standardized values for typical and disordered voices have yet to be

established (McKenna et al., 2022); without a clear consensus on which cycle(s) should be used when calculating RFF or what the typical values should be, the clinical use of RFF is uncertain. Additionally, the added time it takes to calculate RFF using current methods may outweigh its usefulness as a tool for clinicians. The manual estimation method takes a trained technician 20–40 minutes to obtain one set of usable RFF values, which is particularly time-consuming, whereas the semi-automated algorithm requires an expensive subscription to MATLAB (Vojtech et al., 2019). Neither option is readily implementable into clinical environments. Thus, a new method of calculating RFF that is as accurate as the presently available semi-automated algorithm was needed.

The Automated RFF Calculator (ARC; Gill et al., 2023) is software designed to address the drawbacks of current semi-automated RFF estimation methods. ARC requires less training to operate, does not require technicians to access MATLAB, and is less time-consuming to use. Gill et al. (2023) demonstrated a reduction in average analysis time from 10.1 \pm 2.5 seconds in the semi-automated RFF algorithm to 2.5 \pm 1.1 seconds in ARC. The decreased time needed to complete the RFF analysis makes ARC potentially more clinically. Additionally, it allows for fully automated RFF analysis of both real-time audio inputs and prerecorded audio files. Thus, the purpose of this study was to assess the accuracy of ARC in measuring RFF values in typical speakers and speakers with VH compared to current semi-automated RFF estimation algorithms. We hypothesized that there would be no statistically significant difference between the effect sizes of RFF calculated via ARC compared to RFF calculated via the semi-automated algorithm. This would indicate that ARC estimates RFF with a similar accuracy to the semi-automated

algorithm. We also hypothesized that this new method of estimating RFF would be perceived as clinically useful by voice clinicians. Additionally, we sought clinician feedback on what features may be added to ARC to make it more applicable to the clinic.

Furthermore, we aimed more broadly to assess the general structure of the voice evaluation process and how it differs among clinicians across locations. We examined which components are included in the voice evaluation, what equipment is used, and which measures are taken, with a special focus on how clinicians evaluate strain. We also sought to understand the barriers and benefits to having acoustics as part of the voice evaluation process from the perspective of the clinician. We hypothesized that, among the pool of participants, there would be variations in the structure of the voice evaluation as well as a difference in opinion on the use of acoustics. We contextualized our findings with the goal of understanding the clinical utility of RFF in voice evaluations.

Quantitative Methods

Participants

Two datasets of speakers were used to compare RFF values calculated via the semi-automated RFF algorithm (Vojtech et al., 2019) and RFF values calculated via ARC (Gill et al., 2023). All participants provided informed, written consent in compliance with the Boston University Institutional Review Board. All were fluent in English and had typical hearing.

The first dataset — called the “effort” group — included voice samples from 13 adult speakers (seven females and six males; gender information was unavailable).

Participants ranged in age from 19–91 years ($M = 57$ years, $SD = 23.24$ years). These speakers had no prior speech, language, or hearing disorders. A certified SLP specializing in voice screened all participants in this group to ensure healthy vocal function. This screening was performed using auditory-perceptual assessment and flexible videostroboscopy. The overall severity of dysphonia for each participant is listed in Table 1.

Table 1. Participant Demographics and Overall Severity of Dysphonia Rating of the Effort Group

Participant	Age	Sex	Overall Severity of Dysphonia Rating
Effort_01	63	F	11.7
Effort_02	67	M	4.3
Effort_03	53	M	8.3
Effort_04	78	F	19.7
Effort_05	76	M	16.8
Effort_06	46	M	2.6
Effort_07	26	F	9.7
Effort_08	22	F	14.1
Effort_09	19	F	6.1
Effort_10	91	M	23.4
Effort_11	54	M	2.1
Effort_12	66	F	8.5
Effort_13	80	F	20.9

The second dataset — called the “therapy” group — contained audio recordings of 13 speakers who had been diagnosed with vocal hyperfunction (10 cisgender females and three cisgender males) aged 19–64 years ($M = 34.54$ years, $SD = 13.80$ years) who were recorded at two time points: before and after voice therapy. The Voice-Related Quality of Life (V-RQOL; Hogikyan & Sethuraman, 1999) scores, period of therapy, and vocal hyperfunction (VH) type for each participant (either phonotraumatic [P] or non-phonotraumatic [NP]) are listed in Table 2.

Table 2. Participant Demographics and Overall Severity of Dysphonia Rating of the Therapy Group

Participant	Age	Sex	VRQOL Score	VH Type	Therapy Sessions
Therapy_01	45	F	11	PVH	4 sessions
Therapy_02	35	F	14	NPVH	5 sessions
Therapy_03	21	F	13	PVH	6 weeks
Therapy_04	37	F	12	NPVH	4 months
Therapy_05	19	F	13	PVH	12 sessions
Therapy_06	48	M	11	PVH	10 sessions
Therapy_07	33	F	14	NPVH	10 sessions
Therapy_08	19	M	12	NPVH	8 sessions
Therapy_09	64	F	37	NPVH	15 sessions
Therapy_10	21	F	17	PVH	10 sessions
Therapy_11	47	F	15	NPVH	6 weeks
Therapy_12	24	M	13	NPVH	5 sessions
Therapy_13	36	F	Not Completed	NPVH	4 sessions

Protocol

The participants in the effort group were a subset selected from a larger dataset, which was published in Vojtech et al. (2021). Per their protocol, each participant underwent training prior to data collection. During this phase, the participants were trained to produce /ifi/ utterances. This particular utterance was chosen because the phoneme /i/ results in a more open pharynx, allowing for better visualization of the vocal folds during nasoendoscopy (McKenna et al., 2016); additionally, the phoneme /f/ has been shown to reduce intra-speaker variations in RFF values (Lien et al., 2014). First, the participants were instructed to produce eight /ifi/ utterances, taking a breath after the first four to divide the set into two halves. They were then trained to produce the eight /ifi/ utterances using typical and maximal levels of effort. To elicit typical levels of effort, the participants were instructed speak as they would comfortably. To modulate the participant's vocal effort, they were instructed to "increase your effort during your speech as if you are trying to push your air out" with specific instructions to use "as much effort as you can while still having a voice" to elicit the maximal effort condition.

Following the training, participants sat in a sound-treated booth with a directional headset microphone (Shure SM35 XLR) placed 45° from the midline and 7cm from the lips. The microphone signals were pre-amplified (Xenyx Behringer 802 Preamplifier) and digitalized at 30 kHz (National Instruments 6312 USB). A flexible endoscope (Pentax, Model FNL-10RP3, 3.5-mm) was inserted transnasally through the inferior nasal turbinate, superior to the velum, and into the hypopharynx throughout the data collection process. For participants who reported discomfort or whose nasal anatomy impeded

image acquisition with the routine endoscope, a pediatric endoscope (Pentax, Model FNL-7RP3, 2.4-mm) was used to minimize discomfort. Although numbing agents were not used as to not influence typical laryngeal functioning (Dworkin, Meleca, Simpson, et al., 2000), a nasal decongestant was offered to each participant prior to insertion of the endoscope to reduce any potential discomfort caused by the endoscope as it moved through the nasal cavity (for additional details, see Vojtech et al. (2021)). With the scope in place, the participants were instructed to produce eight /ifi/ utterance at both typical and maximal effort levels. RFF onset and offset cycles were averaged per person per effort level.

The participants in the therapy group were a subset selected from an unpublished dataset (Abur et al., in progress). The same procedures were followed as the effort group to obtain these participants' voice samples. Members of this group were instructed to produce seven to fourteen consecutive /afa/, /ifi/, or /ufu/ utterances spoken at an evenly emphasized rate with varying speeds across files. They were instructed to produce only typical effort and were not asked to produce variable levels of effort. Recordings were made at two time periods: before and after voice therapy. RFF onset and offset cycles were averaged per person pre- and post-therapy.

Data Analysis

A set of 223 audio files across both datasets were analyzed via the semi-automated RFF algorithm implemented in MATLAB (Version R2018a, The MathWorks, Natick, MA; Vojtech et al., 2019). The program estimated RFF using a 9-step process that included: (1) identifying the voiceless consonant and vowels in each production

using high-to-low energy ratios in the acoustic signal, (2) manually confirming and/or adjusting the location of the voiceless consonant and vowels in the signal, (3) estimating the average f_o and pitch strength of the vowels using the Auditory-SWIPE algorithm (Camacho, 2012), (4) categorizing the voice signal based on average pitch strength (Kopf et al., 2017), (5) identifying the peaks and troughs of potential vocal cycles related to the vowels in the signal, (6) estimating a series of acoustic features during the transition into and out of the voiceless consonant, (7) locating the boundary between each vowel and the voiceless consonants by applying category-based thresholds to the acoustic feature vectors, (8) rejecting VCV tokens that do not meet certain criteria (e.g., less than 10 onset or offset cycles, glottalization, misarticulation, voicing during the voiceless consonant, etc.), and (9) calculating RFF according to Equation 1: $RFF (ST) = 12 \times \log_2 \left(\frac{f_o}{f_o^{ref}} \right)$ (Vojtech et al., 2021). For the effort group, semi-automated RFF analysis was calculated for each VCV utterance at each level of effort. This analysis resulted in the RFF values for the 10 offset and 10 onset cycles surrounding the voiceless consonant. For the therapy group, semi-automated RFF analysis was calculated for VCV utterances pre- and post-therapy. RFF for all effort and therapy audio files were then automatically calculated via ARC. The RFF values of the cycles closest to the voiceless consonant — Offset Cycle 10 and Onset Cycle 1 — that were derived from each of these methods of estimation were then statistically compared. These two cycles were chosen because previous research has found that they are promising as diagnostic markers of vocal hyperfunction (Stepp et al., 2011).

Statistical Analysis

Cohen's d values were calculated as the effect size of the difference in typical effort and maximal effort in the effort group and the difference pre-post-therapy in the therapy group for both Offset Cycle 10 and Onset Cycle 1. A paired t -test was used to assess the difference in Cohen's d values for the effort group when calculated using the semi-automated method compared to those calculated using ARC. A second paired t -test was used to assess the difference in Cohen's d values for the therapy group when RFF values were calculated using the semi-automated method compared to those calculated using ARC.

Quantitative Results

Although audio files were taken from 13 speakers for both the effort and therapy groups, some participants were excluded in the t -test analysis. A participant's data were excluded if a Cohen's d value could not be calculated for either the Offset Cycle 10 value or the Onset Cycle 1 value by either ARC or the semi-automated algorithm due to inability to estimate one or both RFF values comparing typical and maximal effort or pre- and post-therapy. There are several reasons that either algorithm may not be able to assign RFF values for a particular VCV token. This may occur if there is (1) no steady-state vowel to which the instantaneous fundamental frequencies of the onset or offset cycles could be compared (2) glottalization in the sample, (3) f_0 aperiodicity, and (4) sharp f_0 transitions (Vojtech & Heller Murray, 2019).

Four participants (Effort_04, Effort_09, Effort_10, and Effort_13) were excluded from the *t*-test on Cohen's *d* values for Offset Cycle 10. For Effort_04, Effort_10, and Effort_13, Cohen's *d* values for Offset Cycle 10 could not be calculated via ARC. For Effort_09, Cohen's *d* values for Offset Cycle 10 could not be calculated via aRFF. Three participants (Effort_04, Effort_10, and Effort_13) were excluded from the *t*-test on Cohen's *d* values for Onset Cycle 1. For Effort_04 and Effort_13, Cohen's *d* values for Onset Cycle 1 could not be calculated via ARC. For Effort_10, Cohen's *d* values for Onset Cycle 1 could not be calculated via either ARC or aRFF.

One participant (Therapy_09) was excluded from the *t*-test on Cohen's *d* values for Offset Cycle 10. For this participant, Cohen's *d* values for Offset Cycle 10 could not be calculated via aRFF. Two participants (Therapy_09 and Therapy_12) were excluded from the *t*-test on Cohen's *d* values for Onset Cycle 1. For Therapy_09, Cohen's *d* values for Onset Cycle 1 could not be calculated via ARC. For Therapy_12, Cohen's *d* values for Onset Cycle 1 could not be calculated via aRFF. The coupled Cohen's *d* values used in the *t*-test and subsequent statistics are highlighted in Table 3 and Table 4. All excluded Cohen's *d* values are represented by empty plots in both tables, resulting in a total *n* of nine for the "effort" group and 11 for the "therapy" group in the analysis.

Table 3. Cohen's *d* Values for Offset 10 and Onset 1 Using ARC and aRFF for the Effort Group

Participant	ARC Offset Cohen's <i>d</i>	aRFF Offset Cohen's <i>d</i>	ARC Onset Cohen's <i>d</i>	aRFF Onset Cohen's <i>d</i>
Effort_01	0.72	0.76	0.69	0.24
Effort_02	-1.55	-0.04	-0.35	-0.34
Effort_03	1.74	0.55	0.16	1.00
Effort_04	N/A	-1.51	N/A	0.33
Effort_05	1.21	0.15	0.33	0.36
Effort_06	0.05	0.60	1.68	0.81
Effort_07	0.87	1.00	0.04	0.80
Effort_08	0.34	-0.01	-0.03	-1.79
Effort_09	7.22	N/A	1.03	1.12
Effort_10	N/A	27.31	N/A	N/A
Effort_11	0.61	1.30	0.02	-0.18
Effort_12	1.29	3.83	0.03	0.18
Effort_13	N/A	-1.20	N/A	-0.43

Table 4. Cohen’s *d* Values for Offset 10 and Onset 1 Using ARC and aRFF for the Therapy Group

Participant	ARC Offset Cohen’s <i>d</i>	aRFF Offset Cohen’s <i>d</i>	ARC Onset Cohen’s <i>d</i>	aRFF Onset Cohen’s <i>d</i>
Therapy_01	-0.44	-0.14	0.50	0.38
Therapy_02	0.06	0.24	0.03	-0.76
Therapy_03	-0.30	-0.28	-0.07	0.42
Therapy_04	0.89	0.56	0.31	0.03
Therapy_05	1.06	-0.88	-0.13	0.96
Therapy_06	0.32	0.10	-0.15	0.04
Therapy_07	0.03	0.34	0.17	-0.04
Therapy_08	1.14	0.14	0.36	-0.56
Therapy_09	-4.06	N/A	N/A	-1.57
Therapy_10	0.35	0.06	0.20	0.07
Therapy_11	0.53	0.75	0.86	0.02
Therapy_12	0.69	0.02	-0.11	N/A
Therapy_13	-0.02	0.06	-0.29	-0.04

There was no statistical difference in Cohen’s *d* effect sizes for ARC and the semi-automated method for Offset Cycle 10 or Onset Cycle 1 when comparing the effect of changes in vocal effort (typical vs. maximal effort) in individuals without voice disorders ($p > .05$). There was also no statistical difference in Cohen’s *d* effect sizes for ARC and the semi-automated method for Offset Cycle 10 or Onset Cycle 1 when comparing the effect of voice therapy ($p > .05$). These data are shown in Figure 2 and Figure 3, respectively.

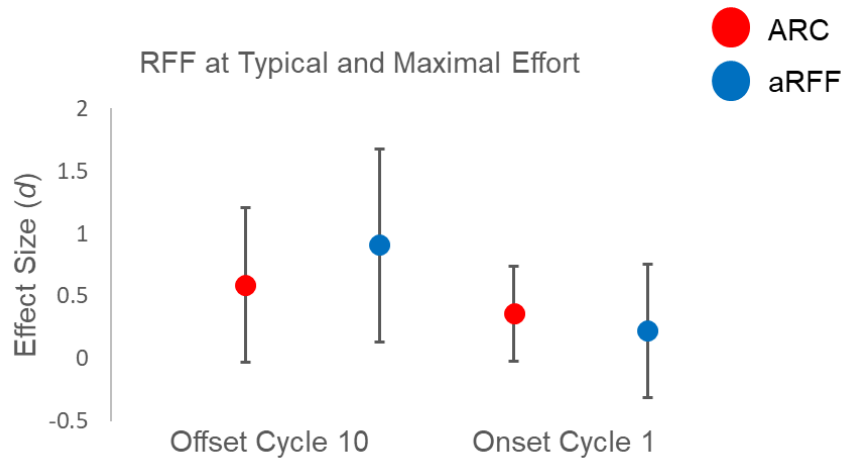


Figure 2. The effect sizes for Offset Cycle 10 and Onset Cycle 1 for ARC and aRFF for the Effort Group. Error bars indicate 95% confidence intervals.

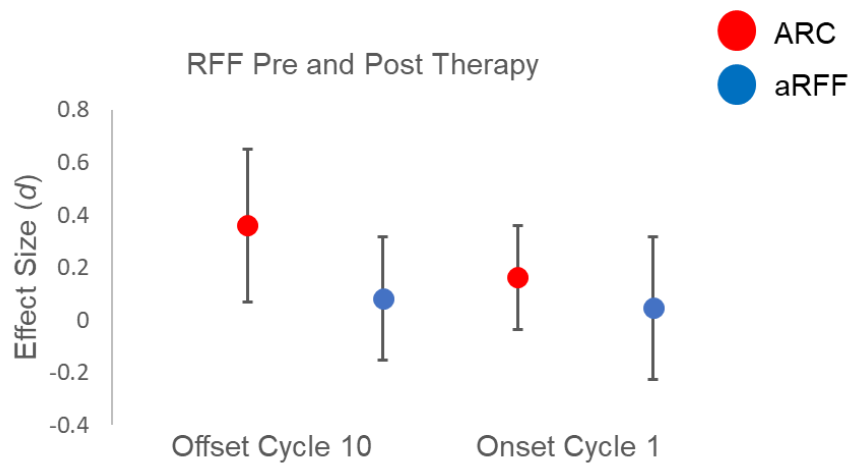


Figure 3. The effect sizes for Offset Cycle 10 and Onset Cycle 1 between ARC and aRFF for the Therapy Group. Error bars indicate 95% confidence intervals.

Quantitative Discussion

There were no statistically significant differences found between the effect sizes for RFF when calculated via the semi-automated method versus ARC in either group (i.e., effort or therapy). This suggests that RFF estimation via ARC is as accurate as the

current semi-automated estimation method. Because ARC is more efficient in terms of training, accessibility, and time, the implication of this finding is that ARC is sufficiently accurate and more suitable for clinical use compared to semi-automated methods (Gill et al., 2023; Vojtech et al., 2019). Based on our results, RFF calculated via ARC is an efficacious tool for assessing strain/vocal effort. However, an important deciding factor of whether RFF is a useful clinical measure is if clinicians think it is a useful measure and it is feasible to obtain during a voice evaluation.

Qualitative Methods

To address the second set of research questions, a qualitative study was performed. Although ARC addresses several of the drawbacks of the semi-automated estimation method (i.e., training, accessibility, and time), a clinician's personal judgment is the most important factor in determining what kinds of analysis are performed in a voice evaluation. In order to understand how clinicians currently view the evaluation of strain and vocal effort and the use of acoustics more generally, qualitative data on their opinions were gathered. These data were used to contextualize the clinical measurement of RFF and determine whether clinicians view ARC as a beneficial clinical tool. Clinicians' opinions about ARC were documented, which may provide guidance of future updates for ARC that will increase its clinical utility.

Participants

Participants were 10 SLPs (nine cisgender females and one cisgender male) who specialize in voice disorders. The participants practiced at different voice centers around the United States and had at least 3 years of experience treating a caseload that consists primarily of patients with voice and upper airway disorders, consistent with our previous studies (Dahl et al., 2021). Participants ranged in age from 28–61 years ($M = 43.9$ years, $SD = 12.06$ years). Their years of experience specializing in voice disorders ranged from 3–34 years ($M = 12.9$ years, $SD = 9.68$ years). The participants practice in a variety of states representing: Connecticut, Illinois, Massachusetts, Michigan, New York, Pennsylvania, Utah, and Washington. Many were trained in different states during their clinical fellowship including: Alabama, Georgia, Iowa, Maine, Massachusetts, Michigan, New York, and Wisconsin.

Protocol

Ten SLPs were recruited via email to participate in the study via online videoconferencing. We conducted a semi-structured interview querying their current practice patterns related to voice evaluations. At the start of each interview, the researcher briefly introduced himself as a graduate student working on his master's thesis project to earn his master's degree in speech-language pathology. Next, the researcher consented the participant. The researcher 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 MTD, 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

researcher noted which aspects of a voice evaluation were present. The form shown in Figure 4 was used to document the participants' responses. Any aspect of the participants' voice evaluation process that was not already listed on the form was added. For those aspects of the voice evaluation process that were not mentioned, the researcher explicitly asked questions to elicit a response following the participant's complete response (e.g., if the participant did not address laryngeal imaging, the researcher asked, "Do you perform any laryngeal imaging during your voice evaluation?"). Further specific questions and details about stimuli were noted using the form in Figure 5.

Part of Evaluation	(check)	Circle and/or List	Notes
Case History			
Patient-Reported Outcomes		VRQOL, VHI, VHI-10, SVHI-10, Omni Vocal effort, vocal fatigue index, Reflux (RSI)	
Auditory-Perceptual Evaluation		CAPE-V, GRBAS	
Laryngeal Imaging (videostroboscopy)		Rigid, flex By SLP or laryngologist	
Acoustics		f_0 , f_0 range, SPL, SPL range, CPP, CSID, AVQI, HNR, L/H Ratio, spectral tilt, jitter, shimmer, pitch strength, acoustic breathiness index	
Aerodynamics		MPT (max phonation time; non-instrumental) s-z ratio (non-instrumental) Subglottal pressure Mean airflow Vocal efficiency Laryngeal resistance	
Hearing Screening			
Oral Mechanism			

Figure 4. Voice evaluation checklist used while conducting qualitative interviews.

Part of Evaluation	Specific questions	Stimuli
Auditory-Perceptual Evaluation	Do you perform these judgments live or from a recording? Live/Recording	Sustained vowels /a/ /i/ /u/, CAPE-V sentences, Rainbow Passage, Spontaneous speech (list any prompts)
Acoustics	What instrumentation/software do you use?	Praat, CSL/ADSV, Visi-pitch, Custom script
	f_0 , f_0 range	Pitch glides
	SPL, SPL range	
	CPP	Sustained vowel (/a/ /i/ /u/), CAPE-V sentences, Rainbow passage (part of the rainbow passage)
	CSID	Sustained vowel (/a/ /i/ /u/), CAPE-V sentences, Rainbow passage (part of the rainbow passage)
	AVQI	Sustained vowel (/a/ /i/ /u/), CAPE-V sentences, Rainbow passage
	HNR L/H Ratio Jitter Shimmer Pitch Strength	Sustained vowels only
	Acoustic Breathiness Index	Sustained vowels, Sentences, Phrases, Rainbow passage
	Anything else not mentioned above (LDDK?)	
Aerodynamics	MPT (max phonation time; non-instrumental)	Which vowel do you use? Subglottal pressure Mean airflow Vocal efficiency Laryngeal resistance
	Subglottal pressure Airflow (VE, LR are just ratios of subglottal pressure and airflow)	Pae/pa/pi
Hearing Screening		
Oral Mechanism	LDDK	

Figure 5. Voice evaluation checklist for specification used while conducting qualitative interviews.

The researcher queried the reasons for decisions on collecting acoustic or aerodynamic data. Participants were asked what benefits or barriers there are (if any) to implementing acoustics into their evaluation. Following this line of inquiry, the researcher 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” for the purposes of this study (Hunter et al., 2020). The participants were also asked about how the COVID-19 pandemic has influenced their voice evaluation process. Finally, the participants were asked if they had any more opinions that they would like to share about acoustics or aerodynamics.

Following the semi-structured interview, the researcher presented participants with the background and relevance of RFF using a *Microsoft PowerPoint* presentation. Next, the researcher provided participants with printed instructions and demonstrated how to use ARC. The instructor then gathered feedback from participants, specifically asking whether they could envision ARC being implemented into their clinical practice. This was done using a questionnaire composed of seven open-ended questions. These questions included: (1) Do you feel that relative fundamental frequency could be a clinically useful measure? (2) What do you like most about ARC? (3) What constructive feedback can you give us on ARC? (4) How could ARC be improved for telehealth use? (5) If ARC was commercialized, would you be willing to pay to use it? If so, would you be more open to a subscription model or a flat-fee model? If not, why not? (6) Would you be willing to use ARC? If so, how would you integrate this software into your workflow?

If not, why not? And (7) What are some features that you feel address or are affected by your previous answers about the use of acoustics in your clinical practice?

The interviews were conducted via online teleconference platform, and all interviews were recorded using two *Zoom H4n Pro* portable handheld detail-audio recorders, with the second recorder serving as a backup in the event of malfunction or human error. The only individuals present during the interview were the researcher and the participant. After the interview, the participant was compensated \$30/hr. The mean interview time was approximately 1hr 23min, ranging from 1hr 10min to 1hr 45min. Following data collection, the audio from each interview was played from the *Zoom H4n Pro* handheld recorder into a microphone connected to a *Lenovo ThinkCentre* PC. The researcher used the “Dictate” function in *Microsoft Word* to generate a transcript for each interview. After the transcript had been generated, the researcher read through the transcript to correct any spelling or grammatical errors.

Analysis

First, the transcripts were reviewed for meaning units, a word, phrase, sentence, or paragraph that describes a specific phenomenon. After this process a code book was developed by labeling the meaning units. The code book was then used as a guide to code the remaining transcripts, with revisions as needed. Initial review for meaning units as well as coding was performed independently by the primary researcher and a research advisor for the first two transcripts. After individual independent analysis of the remaining transcripts by the primary researcher, two transcripts were selected at random for coding by the research advisor. Following this process, the researchers compared the

codes and reached agreement on which were the most accurate for the given transcripts. The primary researcher and research advisor 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. This thematic analysis was performed specifically in the context of (1) the benefits and barriers acoustics and (2) the assessment of strain and vocal effort in the voice evaluation. Meaning units and codes were reviewed and compared across participants for the portions of the interviews covering the evaluation structure as well as the participants' opinions on ARC and RFF. Because the latter was presented as a series of distinct questions, each question was analyzed in turn, examining all unique answers, and determining which ideas were most common across participants.

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 chose suitable participants, cited quotations that accurately reflect the data, and selected relevant meaning units using four main strategies: Firstly, voice clinicians were defined as suitable participants for a study of acoustic analysis in voice evaluations. Secondly, the participants' responses were clarified during the interview to ensure that the correct interpretation was being drawn. Thirdly, each transcript was read separately by two researchers and coded by the primary researcher; the research advisor coded four

transcripts, after which the two coders discussed their findings until agreement was reached. Finally, common themes and patterns in the participants' opinions about (1) the use of acoustics and aerodynamics in voice evaluations, (2) RFF, and (3) ARC 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). 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. Additionally, any notes and forms generated from the interviews were securely kept, and any proper noun was assigned a pseudonym in order to maintain anonymity for future reference.

Qualitative Results

Evaluation Structure

All participants interviewed in the present study reported that they take a case history when the patient arrives to the clinic, using case history forms, interviewing, or a combination of the two. The case history questions primarily pertain to the patient's voice complaint, such as the onset and progression of their voice problem. Participants also commonly reported probing the patient's social and medical history as appropriate, with common questions referencing smoking and reflux. The extent of the case history portion of the evaluation was largely influenced by the length of the evaluation, which varied across participants. The average durations of voice evaluation reported by each

participant are as follows: 15 minutes, 15 minutes, 60 minutes, 90 minutes, 60 minutes, 45 minutes, 30 minutes, 120 minutes, 30 minutes, and 60 minutes ($M = 52.2$ minutes, $SD = 33.35$).

All participants reported collecting patient-reported outcome measures (PROMs), though the number, variety, and timing of when the patient completed them varied across participants. In total, the following PROMs were reported by at least one participant: the Voice Handicap Index-10 (VHI-10; Rosen et al., 2004), Voice Handicap Index-30 (VHI-30; Jacobson et al., 1997), Singing Voice Handicap Index (S-VHI; Cohen et al., 2007), Reflux Severity Index (RSI; Belafsky et al., 2002), OMNI-Vocal Effort Scale (OMNI-VES; Robertson et al., 2003), Cough Severity Index (CSI; Shembel et al., 2013), Dyspnea Index (DI; Gartner-Schmidt et al., 2014), Eating Assessment Tool (EAT-10; Belafsky et al., 2008), Communication Participation Item Bank (CPIB; Baylor et al., 2013), Glottal Function Index (GFI; Bach et al., 2005), Voice-Related Quality of Life (V-RQOL; Hogikyan & Sethuraman, 1999), and the Leicester Cough Questionnaire (LCQ; Birring et al., 2003). It was commonly noted that although general voice questionnaires such as the VHI were given to every patient, questionnaires targeting specific profiles such as the CSI, EAT-10, S-VHI were generally administered only when relevant to a specific patient concern. The two commonly reported ways of administering PROMs were either remotely prior to the evaluation or in person when the patient arrived at the clinic.

All participants reported performing an auditory-perceptual evaluation within their general voice evaluation process. Of the participants included in this study, 50%

reported using the GRBAS for their auditory-perceptual evaluation, whereas the other 50% reported using the CAPE-V. Two participants from the latter group explicitly stated that they do not use the original CAPE-V per the published protocol (Kempster et al., 2009). These participants noted that although they use the general structure of the CAPE-V, they make modifications, such as not physically ticking the line and then measuring it with a ruler, opting instead to assign ranges of numbers “mild,” “moderate,” and “severe” and selecting a number within a range when appropriate. Although the other three participants did not explicitly report any modifications to the CAPE-V, it was not verified whether they were using the published protocol. Nine participants reported making their auditory-perceptual judgments in real time during the appointment, with one participant noting that they return to the recording after the evaluation to make their judgments on voice quality.

In terms of laryngeal imaging, seven participants reported that a flexible endoscope is used more frequently than a rigid endoscope at their practice. The three remaining participants reported that (1) rigid endoscopes are used most often, (2) flexible and rigid endoscopes are both used, and (3) the use of a flexible or rigid endoscope is largely dependent on the voice quality and patient history alluding to a particular etiology (flexible scope for a breathy voice quality and rigid scope for a rough voice quality). Of the 10 participants, six reported that the laryngologist or laryngology fellow operated the endoscope all or most of the time. Five of these participants are present in the room (if possible) for the scoping, whereas one participant is not present. Three participants consistently perform the endoscopy, whereas the remaining participant operates the scope

if a patient is referred from an outside Ear Nose and Throat (ENT) clinic, but they observe the laryngologist in joint visits.

Out of the 10 participants in the study, only one reported not including an oral mechanism examination in their voice evaluation. Four participants reported that an oral mechanism examination is not a standard aspect of their voice evaluation, but they perform one as needed, most commonly for suspicion of a neurological disorder (e.g., dysarthria). The other five participants reported that oral mechanisms examinations are a standard component of their assessment process. One participant reported that they perform a hearing screening as a part of their standard voice evaluation. The other nine participants reported that they do not perform hearing screenings during their evaluations. One participant reported that all new patients are referred to their clinic's audiologists and thus they do not need to perform a screening. Another participant reported that they will ask questions pertaining to hearing during the case history portion of the evaluation and refer to their clinic's audiologist if there was a concern.

The two most reported systems for measuring acoustic parameters were the Computerized Speech Lab (CSL; Elemetrics, 1994) and Praat (Boersma, 2001). Of the 10 participants, 50% reported using CSL or its compatible software, 10% reported using only Praat, 10% reported using Praat and CSL-compatible software, and 30% reported using neither software. Across interviews, the following CSL-compatible software were mentioned: Analysis of Dysphonia in Speech and Voice (ADSV; model 5109 version 3.4.2; KayPENTAX, Montvale, NJ), Multidimensional Voice Program (MDVP; Elemetrics, 1993), and Real-Time Pitch (Elemetrics, 1994). One participant mentioned

using Phonanium (Maryn, 2017). Each participant that reported using either CSL or Praat used a unique combination of both programs in their acoustic analysis (e.g., ADSV, CSL, and Praat).

The participants also differed in the number and kind of acoustic and aerodynamic measures they take during their evaluation. The measures that each participant records during the evaluation are summarized in Table 5 below. A table of the tasks used to obtain these measures can be found in the appendix.

Table 5. Summary of the Acoustic and Aerodynamic Measures in Each Participant’s Voice Evaluation

Participant	Acoustic Measures	Aerodynamic Measures	Software/Equipment
01	Mean f_o f_o Range CPP HNR Cepstral Spectral Index of Dysphonia (CSID)	Phonatory Threshold Pressure Mean Airflow SPL Range	ADSV MDVP Phonatory Aerodynamic System (PAS)
02	Mean f_o f_o Range CPP Jitter	Phonation Threshold Pressure Mean Airflow SPL Range Mean Peak Air Pressure Maximum Phonation Time Aerodynamic Resistance Tasks	MDVP PAS
03	Mean f_o CPP RFF	Subglottal Pressure Transglottal Airflow Mean SPL Maximum Phonation Time Phonatory Efficiency	CSL PAS Pneumotachograph
04	Mean f_o f_o Range f_o Standard Deviation Smoothened CPP Acoustic Voice Quality Index (AVQI)	Subglottal Pressure Average Airflow Sustained Airflow Mean SPL SPL Range SPL Standard Deviation Maximum Phonation Time	Praat (Phonanium) PAS

05	Mean f_o f_o Range	Maximum Phonation Time	Smartphone Applications: Voice Analyst Pano Tuner
06	ADSV MDVP AVQI (occasionally)	Peak Expiratory Flow Maximum Phonation Time S/Z Ratio	ADSV MDVP Praat Peak Flow Meter
07	None	Maximum Phonation Time* *If concern for paralysis/paresis	None
08	Mean f_o f_o Range CSID	Peak Airflow Mean SPL Maximum Phonation Time S/Z Ratio	ADSV MDVP Real-Time Pitch Spirometer
09	None	Maximum Phonation Time S/Z Ratio	None
10	Mean f_o f_o Range CPP	Mean SPL SLP Range Maximum Phonation Time S/Z Ratio	MDVP

The most commonly reported lasting changes due to COVID-19 concern the Phonatory Aerodynamic System (PAS), a computer-based hardware and software system used to measure several aerodynamic voice characteristics. Of the 10 participants, 50% reported a change to their PAS protocol since the pandemic. Primarily these changes were related to the cleaning of the PAS. Several participants noted that prior to the pandemic, they would wipe down the mask and dispose of the intraoral pressure straw. They reported that now either the entire machine must be taken apart and cleaned after each patient or they must use 100% disposable parts that are disposed of after a single use. Two participants shared that they have discontinued using the PAS entirely to take aerodynamic measures due to the change in maintenance. Other lasting changes

mentioned include the option to perform evaluations via telehealth, asking case history questions specifically related to COVID-19, discontinuing laryngeal palpation to assess tension, placing less value on acoustic and aerodynamic measures, and continuing to wear masks. One participant reported that their voice evaluation has returned entirely to pre-pandemic protocols.

Benefits and Barriers to Acoustics

In the following subsections, results of the thematic analysis are described. A summary of the themes and subthemes along with their definitions are provided below in Tables 6 and 7.

Table 6. Themes and Sub-themes Describing the Benefits and Barriers of Performing Acoustics

Themes	Sub-themes	Definitions
Acoustics take time	Time for clinician	Evaluation length and added analysis Patient fatigue and frustration
	Time for patient	
Acoustics do not inform therapy patterns	Acoustics do not reflect perception	Incongruence between what someone is hearing and what the values are reflecting
	Acoustic tasks do not capture the problem	Many acoustic tasks are not reflective of everyday speaking conditions
Acoustics allow for the most accurate comparison	Recordings	Audio recordings can be used to compare voices at different time points Acoustic data can be used for progress tracking
	Objective progress	
Acoustics supplement patient-centered care	Satisfaction, validation, and buy-in	Patients like seeing acoustic data Clinicians use acoustics to educate patients
	Education	

Theme 1: Acoustics takes time.

Participants described how the inclusion of acoustics into their voice evaluation process adds extra time that negatively impacts both the clinician and patient.

Time for clinician. Several participants reported that the inclusion of acoustics into their evaluation process does or would add time that would negatively impact other aspects of their practice pattern. REACT01 reported that this extra time mainly comes from having to move the patient into a different room as well as clean the microphones and other items the patient came into contact with during this time rather than the acoustic task performance itself. REACT05 expressed a similar sentiment that the retrieving of the equipment and the set-up required outweighed the benefit of gathering the measures. They noted that although their clinic does have the CSL equipment, they only use it when they have time, which they reported to be infrequent. An important recurring trade-off observed across interviews was the weighing of time and clinical benefit. In reference to not using the CSL equipment, REACT05 stated that “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.” REACT07 also 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 aerodynamics and acoustics.” This view was similarly expressed by REACT09 who emphasized that acoustics 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 10 minutes, that's an extra 100 minutes added on to my day just to do this.”

Time for patient. Several participants described how the additional time added to the evaluation for acoustics may negatively impact the patient’s experience at the clinic. REACT09 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 what their pathology is. 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.’” REACT08 shared a similar thought, discussing how acoustics extend the length of the evaluation which can lead to patient fatigue, potentially resulting in an inaccurate depiction of their voice at that time. Although patient frustration due to added time from acoustic assessment was not a common concern across the 10 participants, together with the *Time for clinician* subtheme, its presence was important to note.

Theme 2: Acoustics do not inform therapy patterns.

Acoustics do not reflect perception. A common complaint across participants was that there is sometimes an incongruence between the participant’s perception of a given vocal quality (e.g., roughness, breathiness, etc.) and the acoustic correlate. REACT01 gave the example that “you don't hear roughness, you hear a little bit of breathiness, but when you do some acoustics you're getting like 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. REACT10 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.” REACT05 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 same issue holds true for patient perception as well. REACT03 explained that there are times when a patient arrives with complaints of vocal fatigue, discomfort, and/or pain but will generate acoustic and aerodynamic values that are within typical limits.

Acoustic tasks do not capture the problem. Another common barrier that participants reported was that the tasks themselves are far removed from typical voice use in which the problem is usually encountered. REACT03 shared that they often encounter this issue with singers. They explained that 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. “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.” REACT07 reported that even reading can change the quality of a patient's voice: “As soon as you put a script in front of somebody, in my opinion, they end up sometimes changing the way they use their voice,” which has the potential to affect the acoustic data. REACT10 explained that although some measures such as CPP address this issue, 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, use their speech functionally.”

Clinician perception is enough. This subtheme was commonly composed of two fundamental opinions: (1) clinicians can perceive what needs to be addressed in therapy without objective data and (2) patient perception is most important. Several participants expressed that, without gathering acoustic measures, they can form appropriate therapy treatments and targets. One participant described their view as follows:

I don't think there's anything that would be like ‘Oh, your numbers say this, so I should do this type of therapy with you.’ Like we already know you're not exhaling air... or you're holding back your airflow when you talk, so I already know that I'm going to work on that with you.

A similar line of thinking was reported by several other participants. One participant reported that some acoustic measures that were once upheld but have now proven inefficacious have led to a hesitancy to accept newer measures: “I'm jaded after jitter, shimmer, noise to harmonics ratio, relative average perturbation. You know I've been doing this for a long time, and I've seen the parade of numbers come through...And so I feel like as a field, we've been chasing this dream of being able to say, ‘Here's a number that tells us— that replaces our ear and tells us how dysphonic this person was,’ and I don't think we've achieved it.” REACT01 expanded on this by paraphrasing a quote from another voice researcher, saying: “the gold standard for all these acoustics that

we're going for and that we're trying to create is to match the ear.”

Several participants expressed that therapy targets are ultimately the decision of the patient. REACT01 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 REACT10 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.” REACT01 expressed this same view saying, “The really good voice clinician is the ‘I'm not fixing what I want to fix; I'm fixing, you know, what *you* want to fix.’” This participant added that they have had patients who leave therapy with an improved but still dysphonic voice, and as long as the patient is satisfied with their progress and voice, the participant does not insist on continuing therapy.

Theme 3: Acoustics allow for the most accurate comparison.

General recordings. Although one of the primary purposes of this study was to investigate the benefits of specific acoustics measures, several participants talked about the value of a general audio recording. Although general audio recordings do not qualify in themselves as an “acoustic measure,” they are typically obtained in the process of gathering the measures and can be viewed as a form of acoustic data, generally. Participants reported that saving a general recording for later reference was helpful for several different aspects of therapy. Several participants discussed that recordings are important to take in order to compare voice quality after a patient undergoes treatment or

a procedure (e.g., vocal polyp excision) to their baseline prior to treatment or surgery.

REACT10 explained that these comparison recordings allow both them and the patient to more accurately assess progress compared to relying on their memory, which may not be accurate. 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 I saw them, how different do they sound from before they started therapy. And often the patient, because they're living with their voice, they don't remember.”

Objective comparison and progress tracking. The most frequently reported benefit of taking acoustic data during voice evaluations is to gather data midway or at the end of therapy for objective progress tracking and comparison. Similar to the way in which recordings offer a more standardized way of perceptually tracking progress, participants reported that measuring acoustic data is important for showing objective change and ensuring measurable outcomes. REACT04 offered a specific example, “For example, CPP is very nice for showing any changes in the signal part of the recording. So, 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.

Theme 4: Acoustics supplement patient-centered care.

Satisfaction, validation, and buy-in. One concept that arose both while specifically discussing the benefits and barriers to acoustics, as well as throughout the

interviews at large, was that using acoustic measures positively influences therapy buy-in and patient satisfaction. Participants noted that patients enjoy seeing acoustic data when presented and explained correctly. According to several participants, much of this enjoyment comes from the validation that patients experience by seeing objective data related to their problem. REACT02 explained that the idea that “a computer hears what you hear or what you’re feeling” can be reassuring to patients. REACT08 specifically noted their experience with patients who want to know if their voice problem is “normal” or not: “I think that it's beneficial for patients to know, ‘Okay, this is within normal limits; you're OK,’ and saying that to them and validating what's going on because I've seen a lot of patients say, ‘Is that okay? Is that normal?’” REACT09 further specified that the graphs and charts generated by acoustic software are a large component of buy-in. Conversely, one participant noted that acoustics could be a detriment specifically because they 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. REACT10 gave a general insight that interpretation of acoustic data can be confusing for those who lack the training, including both other healthcare professionals and patients, emphasizing the importance of accurate and easy-to-understand explanations. REACT03 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.”

Education and understanding. Many participants reported that acoustics measures are great tools for patient education. REACT09 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 purely perceptual descriptions to describe the patient's voice. 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 data produced by software can be used as biofeedback for patients, helping them to understand how to adjust their voice.

Other Benefits and Barriers.

Despite not forming coherent themes, participants reported several other benefits and barriers to including acoustic measure in their voice evaluations. One commonly reported non-clinical benefit was the use of the data for research. Both participants who do and do not find acoustic data to be clinically valuable endorsed the usefulness of the data in research. Another benefit that some participants reported was that acoustic measures were useful for fleshing out a full characterization of a patient's voice disorder. Even for participants who do not find acoustics informative for voice therapy, some recognize a sentiment expressed by REACT10: "We want to have as much objective data as we possibly can...I mean, we do evidence-based practice, so we want to have as much evidence as we can." Another commonly reported benefit is that acoustic data provide the participants with validation of their auditory-perceptual evaluation. Similarly, one participant reported that acoustic measures allow them to support their opinions with quantitative and visual rationales. Additionally, several participants reported that the acoustic data guide their therapy decisions, such as what techniques to use and if the patient is ready to be discharged. One participant noted that acoustics are helpful because they can detect important diagnostic markers that clinicians can sometimes not perceive.

REACT04 gave the following example: “Well, one thing we've shown is that you may not hear vocal tremor, but acoustic measures are diagnostic of it, even if you can't hear it. And so that's a value. And if somebody does have vocal tremor, we'll measure the rate and extent of it.” This participant noted that this kind of acoustic data guides their therapy techniques as well, designing treatment targets based on atypical acoustic values even when they cannot be perceived.

Several additional barriers were also presented across the interviews. One participant focused on patient-oriented barriers related to performing acoustic tasks, such as individuals who are illiterate or uncomfortable reading aloud. Although this should not inherently affect the ability to obtain acoustic measures, it does cause a deviation from the clinic's general protocol. For example, instead of reading the Rainbow Passage to collect acoustic data as is done by other patients, the patient instead produces spontaneous speech about their plans for the day. Several participants reported that they face challenges standardizing an environment in which to record acoustic data. One participant reported that they do not have a sound-treated room and are thus exposed to environmental noise that competes with the recording. Another participant reported that the room in which acoustics are recorded varies day-to-day due to accommodating different schedules. Although they do not face this barrier themselves, REACT04 noted that some SLPs may not have the necessary equipment or a HIPAA-compliant method of recording a patient for acoustic analysis. Two participants reported that they experience no barriers in using acoustics but shared their least favorite aspect about their use. One participant described software like KayPENTAX as a “black box.” They expressed their

dislike for the opaqueness of how the values are achieved, compared to Praat, which they better able to modify. The other participant wanted to emphasize that acoustic values are only a part of the clinical picture and must be interpreted with reference to the patient. They found that barriers only begin to arise if a clinician or patient begins to focus solely on the numbers rather than the needs of the patient holistically.

Evaluation of Strain and Vocal Effort

Table 7. Themes and Sub-themes Describing the Evaluation of Strain and Vocal Effort

Themes	Sub-themes	Definitions
Clinicians lack consensus on objective measures of strain	-	Different 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 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

For the purposes of this study, strain was defined for the participants as “a listener’s auditory-perceptual judgment of vocal effort” and vocal effort was defined as “self-perceived exertion during vocalization by the speaker” based on Hunter et al. (2020).

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

Most clinicians (70%) did not report obtaining an objective correlate of vocal strain in their voice evaluation. Of the three participants who reported using an objective correlate to measure strain, one participant reported using subglottal pressure as well as the phonatory efficiency measures. Another participant reported expressed that phonatory efficiency has “a great correspondence with strain.” Additionally, the third participant reported referencing CPP values as well as the definition of harmonics in the signal in the spectrogram to assess strain. Aside from these five objective correlates reported by three participants, all other participants did not report the use of an objective measure of strain.

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

As mentioned previously, half of the participants reported using the CAPE-V whereas the other half reported using the GRBAS to measure strain, as both of these includes percepts of strained vocal quality. In filling out these scales, the participants reported using a variety of auditory, visual, tactile, and synesthetic information to inform their evaluation of strain.

Auditory Perception. Some participants decided to describe what they are listening for when rating strain using their perceptual scale. These data are best described using quotations from the participants to capture their exact wording due to the abstract nature of describing auditory-perceptual analysis (Table 8).

Table 8. Participants’ Descriptions of their Auditory-Perceptual Evaluation of Strain

Participant	Description of Strain
REACT02	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.
REACT03	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.
REACT05	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.
REACT08	Something that sounds like or feels like a pushing.
REACT09	Perception of effort or work essentially, or does it sound easy or does it sound hard.

The two major patterns that emerge above are the perception of strain as a “pushed” quality or as a loss of resonance.

Visual and Tactile Perception. Although strain is defined as an auditory-perceptual phenomenon, several participants reported pairing a visual-perceptual assessment to form their clinical picture of strain. Two participants reported that they observe patient posture and muscle tension as components of strain. In addition to observing muscle tension, laryngeal palpation was reported as a part of the clinical picture of strain as well. The most commonly reported visual-perceptual assessment of strain was endoscopy. Specifically, several 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, the patient not taking an adequate number of breaths while speaking is indicative of strain.

Synesthetic Perception. Several participants reported using their own body's response to auditory perception as a method of assessing strain. REACT03 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.” REACT05 reported a similar experience, which they described as, “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, REACT07 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.

According to the definition of vocal effort specified prior to the questioning, vocal effort is strictly a self-reported phenomenon experienced by the speaker. Therefore, only measures that assess patient-reported experience qualify. However, some participants also reported perceptual measures that they take note of to inform their clinical picture.

Patient reports. The most commonly reported method of assessing vocal effort was via open-ended patient explanation. This mainly consists of asking the patient for a verbal or written explanation of their vocal effort. One participant reported probing vocal effort via analogy: “I asked them to describe 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.” Rating scales were also used 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. Thirty percent of participants reported using the OMNI-VES (Robertson et al., 2003; shown in Figure 6) a standardized Borg scale on which 0 represents that 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. Fifty 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. One participant reported using specific questions on the VHI (a Likert-scale; Jacobson et al., 1997) to assess vocal effort, 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 half.”

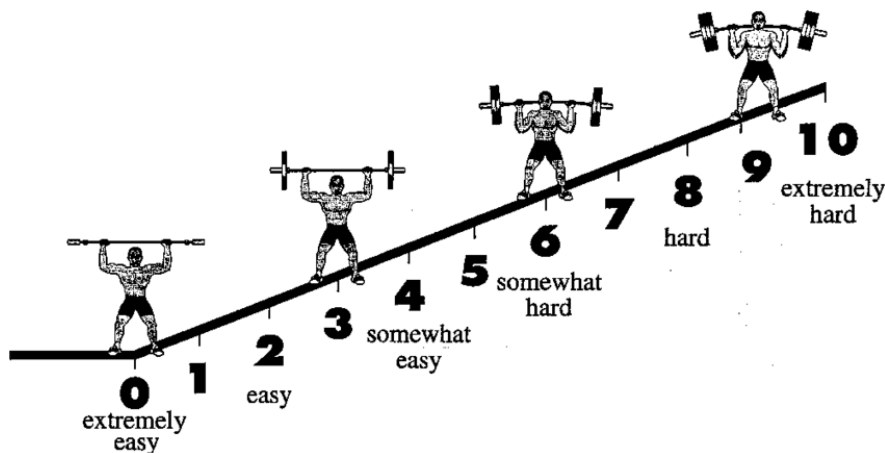


Figure 6. The OMNI-Vocal Effort Scale (Robertson et al., 2003).

Correlates. Although most participants stuck to the definition of vocal effort as self-perceived patient exertion while speaking, a minority mentioned the use of perceptual and objective measures for vocal effort. Two participants mentioned that visible tension around the patient's larynx adds to their clinical picture of effort. REACT05 described what they are looking for in this way: "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." One participant noted that their observation of the patient's respiration informs their perception of vocal effort. This includes gasping for breath or running out of breath while speaking. Two participants reported that auditory perception is something that they consider as well, specifically listening for stridor or "perceptually listening for the sound of somebody like lifting a piano." Additionally, another 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 phonatory efficiency).

Opinions on RFF and ARC

To assess whether and how the participants saw ARC being implemented into their practice, they were probed using a questionnaire composed of seven open-ended questions. Each item is addressed individually below. Because some questions prompted comments on the appearance of the interface of ARC, images of the software are

included in Figures 7–9.

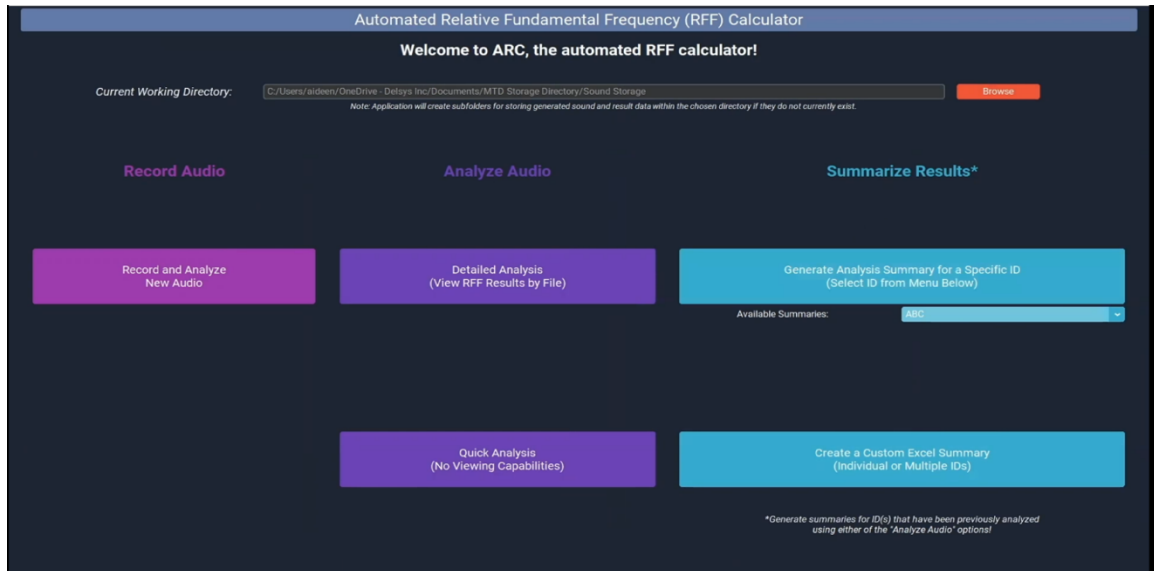


Figure 7. ARC Dashboard

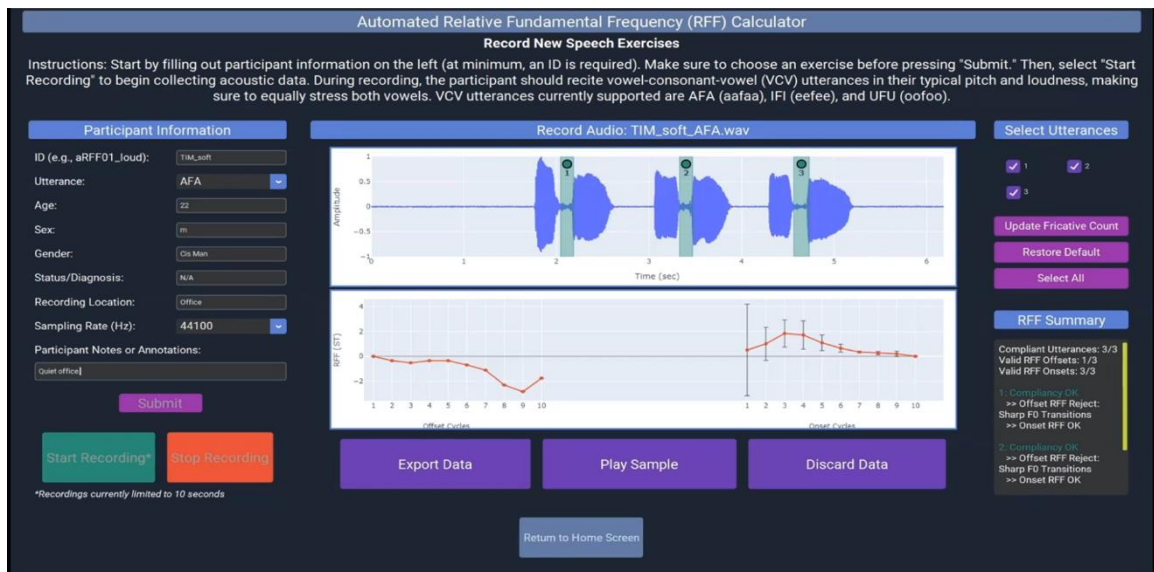


Figure 8. Waveform and RFF Graph of an Analyzed Sample in ARC

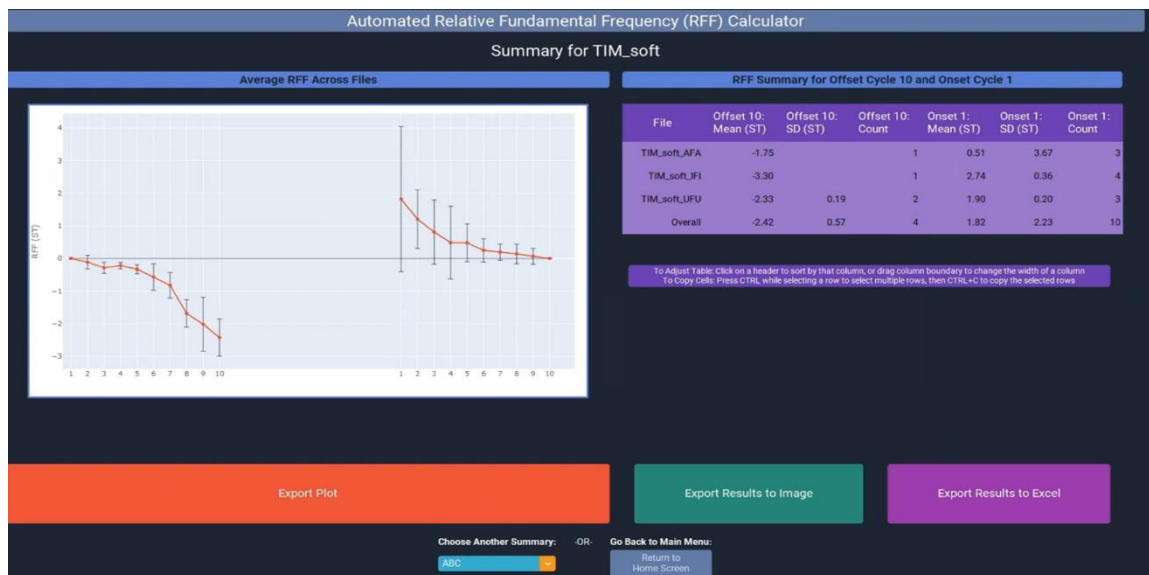


Figure 9. RFF Graph and Value Table of an Analyzed Sample in ARC

Question 1: How can you see relative fundamental frequency helping you clinically?

The most common answer to this question was that participants viewed RFF as a helpful tool for comparison. More than half of the participants reported that they would find it clinically useful to be able to objectively track changes in strain. Some participants discussed using RFF as a pre-/post-outcome measure in therapy, whereas others expressed interest in using it as a baseline measure at the beginning of each session and comparing it to RFF values following manual therapies, such as before and after laryngeal massage. Additionally, another participant discussed how they would use the change in RFF values to improve patient buy-in. Relatedly, the second-most reported response to how RFF could be clinically useful was as a biofeedback tool. Several participants noted the importance of showing the patient close-to-real-time objective data to aid in improvement of voice. One participant mentioned that they would use

biofeedback via RFF for patients to check if they were adequately warmed up prior to beginning other therapy exercises. It is unclear if the participant was referring to singers or nonoccupational voice users. In either case, RFF has not been studied in terms of short-term changes in vocal fold physiology vocal warmups are used to provide.

Additionally, some participants found value in RFF as a way to validate and supplement their own auditory perception of strain. Firstly, several participants noted that it is helpful to have an objective value that reflects an auditory-perceptual quality so that they can feel more confident in expressing their auditory-perceptual judgment. As one participant said, “Knowing that there's some correlation between RFF and strain... I'm not just leaning on my ear; there's some numbers that I can lean on.” Another participant noted that RFF would be clinically useful as a supplementary diagnostic tool in times when they suspect a certain disorder (e.g., MTD), but it is not clear. Along this same line of thinking, other participants endorsed RFF as potentially useful in measuring a patient's voice complaint when their voice is perceptually typical, or the participant's judgment differs from what the patient is experiencing.

Question 2: What do you like most about ARC?

A common response to this prompt, and one that was explicitly stated by 80% of participants, was that ARC is easy to use. Many participants overtly used descriptors such as “user-friendly,” “self-explanatory,” and “straightforward.” One participant who has an interest in technology noted that “I can imagine sitting down with it with very little instruction and still being able to like poke around and kind of figure out what it does.” Some participants contrasted the simplicity of ARC to other acoustic software such as

ADSV and Praat, which they perceived as potentially overwhelming and much less intuitive. The second-most frequent answer was that they liked ARC's visual appearance and organization. Some specific aesthetic features that the participants highlighted as positive were the large buttons and that the graph and table visuals were included on the same page, in addition to general positive comments on the color scheme and design of the program. In addition to these two broad categories, participants also noted specific features that they enjoyed about ARC. One of these features was the ability of ARC to analyze recordings in batches rather than solely one at a time. Other participants noted that the visuals generated by the analysis (i.e., the table and graph) are easily embeddable into their clinical reports. Several participants appreciated the integration of ARC with *Microsoft Excel* for visualizing data. Additionally, ARC's ability to detect errors and select what to include in the final analysis were reported by two participants as positive features. Finally, the ability to analyze both live and pre-recorded samples was seen as a positive feature of ARC.

Question 3: What constructive feedback can you give us on ARC?

This prompt resulted in the most diverse and numerous responses of all the questions included in this study. The most common response, given by 60% of respondents was that ARC should incorporate RFF normative values. Participants suggested the inclusion of error bars or shading on the graph or normative values in the data table that reflected when an utterance is within typical limits or not. Similarly, some participants reported that they would like a biofeedback component to be added to ARC that could show, for example, a value that the patient is attempting to achieve and where

their current production falls relative to that goal. One participant suggested an overlay feature in which a patient's RFF values can be viewed on the same graph to visualize change more easily.

Another common point of constructive feedback was the inclusion of an instructional page within the software. Although many participants reported the software itself as "straightforward," they thought the instructions for how to instruct the utterances were unclear. These participants had several ideas as to what they would like to be on such an instructional page if it were to be added to the software. These included (1) discrete instructions on what utterance should be used and why, (2) at what rate, frequency, and loudness the tokens should be produced, (3) the minimum number of tokens that must be gathered, (4) a standardized script to ensure that clinicians are instructing properly, and (5) an explanation of what can cause a failed analysis and how to instruct the patient after one has occurred. In addition to these stimuli-specific instructions, the participants also reported that it would be helpful to include additional information within the software as well, such as on which operating systems ARC can be used.

Some participants suggested that adding additional export options would be beneficial. Despite integration with *Microsoft Excel* being seen as a positive feature by some participants, others found it limiting as the only option. One participant reported that although they do have access to *Microsoft Excel*, they do not enjoy using it in their clinical work. Another participant noted concern that many clinicians would not have the *Microsoft* package which includes *Microsoft Excel*. Therefore, they suggested that ARC

should prompt the user with several different outputs upon exporting the data. This participant also suggested that a database be built into ARC to keep track of all patients, eliminating the need for an additional software. This connects directly to another aspect of constructive feedback given by other participants which is ARC can be made more efficient by eliminating the transition to *Microsoft Excel*:

It did seem like there was a lot of clicking and going back and forth from the program to the Excel to wherever it's saved and a lot of crossover. And I'm just trying to think— and like, there was a lot of extracting and analyzing and thinking about it and different ways, which I think is helpful, but to be able to do all of that and then all that cross-checking. I just honestly—I think it takes too much time.

This opinion of eliminating additional software was shared by several other participants, some of whom believed that ARC itself should be integrated into a pre-existing software to minimize the number of programs a clinician is expected to open during a voice evaluation. One participant drew a comparison to Phonanium, a plug-in with scripts that can be integrated into Praat.

There were several unique instances of constructive feedback that are useful in guiding future development of ARC. One participant suggested that a visual timer to be added so that a given patient can see how long they have to produce the required utterances, rather than the current countdown timer. This participant also thought inclusion of the live, generating waveform as the patient produces the utterances could be used as a form of biofeedback and patient education while also being visually

stimulating. Another suggestion was to include an alert that advises clinicians to check for errors in the analysis prior to fully analyzing the data. Other constructive feedback included using more patient-friendly labels (e.g., “Offset of Phonation Mean” rather than “Offset 10 Mean”), improving ARC’s background noise reduction ability, decreasing the likelihood of rejections, converting ARC to a smartphone application, and generally improving its aesthetics so it does not look like a prototype. One participant’s feedback connects directly with a primary barrier to acoustics as seen previously: they suggested that ARC be compatible with sentence-level analysis to more closely reflect daily speaking behaviors.

Question 4: How could ARC be improved for telehealth use?

Due to the popularization of telehealth as a result of the COVID-19 pandemic, optimization of ARC for remote use is a crucial component to target for future development. For this question, three categories of answers emerged, which include (1) how patients and clinicians could use ARC remotely, (2) what features could be added to ARC that would aid/supplement its remote use, and (3) general concerns for remote acoustic analysis. Of the participants who discussed how they saw ARC being used remotely, most suggested that the software be accessible by the patient, who could then synchronously record themselves while being instructed over telehealth and subsequently send the recording and results to the clinician via a HIPAA-compliant platform. Participants also expressed interest in using ARC as a semi-ambulatory monitoring device. These participants suggested synchronously instructing patients on how to operate the software so that they could asynchronously collect data throughout the day to

then send to the clinician.

There were several specific features that individual participants reported would be of benefit when using ARC remotely. Similar to the general constructive feedback, one participant suggested a visual timer as well as a live visual of the waveform during the production to increase engagement and feedback. Specific to telehealth sessions, one participant suggested that ARC be able to use several audio file types, such as WAV and MP3, or have a file conversion feature to ensure that whatever audio file type sent by the patient is compatible with the software. Additionally, because telehealth is relatively new, one participant suggested that ARC be as accessible as possible, particularly for patients who are less familiar with technology. To prevent errors caused by the telehealth format, some participants suggested that ARC have a feature to adjust for the specific carrier signal being used when sending the recording and data to the clinician. Another participant suggested requiring a microphone test prior to the official recording to ensure that the environment was adequate for acoustic recording.

Finally, the most common response was a general concern for the fidelity of acoustic recordings via telehealth. The three most frequently expressed concerns were: (1) equipment, (2) environment, and (3) protocol. In terms of equipment, participants were unsure that the quality of whatever microphone the patient was using would be adequate for a recording that was to be acoustically analyzed. A primary environmental concern was the inability to control for background noise, which may affect the accuracy of the acoustic data. Finally, there was a concern for the lack of standardization across participants, such as the inability to verify that the patient's microphone is positioned a

standard length away from their mouth.

Question 5: If ARC were commercialized, would you be willing to pay to use it? If so, would you be more open to a subscription model or a flat-fee model? If not, why not?

This question was designed to address the cost of acoustic analysis. In total, 20% of participants responded that they would not be willing to pay for ARC at their current practice, whereas 80% said that they would be willing to pay for it, in some cases if certain conditions were met. The 20% of participants who reported that they would not be willing to pay for ARC were also the participants who reported that they currently do not take any formal acoustic measures at their practice.

Between flat-fee and subscription models, participants most often reported that they would prefer a flat-fee model for a variety of reasons. The most reported reason for choosing a flat-fee option over a subscription was that it would be easier to secure approval of a one-time purchase than a recurring payment, for which one must repeatedly prove the value of the software. One participant advocated for a flat-fee model, noting that for medically based voice clinics, the price of acoustic software pales in comparison to other larger medical purchases. Another reported reason was a flat-fee model is more compatible for departments that receive either grants or patient donations. In the case of the former, if grant funds are put toward the flat-fee payment of a software, the clinic would own the software indefinitely. On the other hand, if grant funds were put toward the subscription payment of a software, at some point in time, the grant funds will end, and the clinic would lose ownership of the software. In the case of the latter, one participant reported that a donor is more likely to contribute if their funds secure the

clinic with a permanent product (i.e., a flat-fee purchase). Another rationale presented for a flat-fee model was that a flat fee is the standard for other acoustic software (MDVP, ADSV, etc.). Finally, one participant discussed “subscription fatigue,” a phenomenon that describes the frustration consumers feel in response to the high prevalence of subscriptions.

Question 6: Would you be willing to use ARC? If so, how would you integrate this software into your workflow? If not, why not?

In total, 20% of participants responded that they would not be willing to use ARC at their current practice, whereas 80% said that they would be willing to use it. The 20% of participants who reported that they would not be willing to use it at their current practice were the same participants who reported that they would not be willing to purchase the software since they do not include any formal acoustic measures as a section in their voice evaluation protocol.

Of the 80% of participants who reported they would integrate ARC, five said they would have their patient perform a live recording using the software during the acoustics and aerodynamic portion of their evaluation. One additional participant stated that they would integrate ARC into their evaluation if it was made into a smartphone application, as that is how they perform their other acoustics assessments during the voice evaluation. Another participant reported already collecting RFF stimuli during a voice evaluation, but they do not report RFF in their clinical report. With the introduction of ARC, this participant reported that they could envision embedding the graph and data table into their clinical report and providing a brief narrative analysis.

Two participants reported that they would likely integrate ARC outside of the voice evaluation process. For example, they would use ARC as a progress-tracking tool, comparing RFF values from two or more time points. Integration of ARC into therapy sessions as a means of biofeedback for the patient was another reported use. The two participants who reported that they would most likely not integrate ARC into their practice pattern had different reasons for doing so. One participant expressed that they needed to see more data and examples of how RFF reflects voice characteristics and severities and determine whether that would affect their therapy techniques. The other participant noted that the additional time it takes to gather, analyze, and report ARC outweighs the benefits of using it.

Question 7: What are some features that you feel address or are affected by your previous answers about the use of acoustic in your clinical practice?

Participants were split between ARC positively addressing their previous opinions on acoustics and negatively addressing their previous opinions on acoustics as well as reporting no changes. Two participants who noted that the small amount of time it takes to gather RFF values from ARC addressed their previously expressed belief that time is a major barrier to acoustics. These participants who reported that ARC would add an additional value to their data collection also reported that the speed in taking the measure would not add a significant amount of time to the evaluation. Three participants noted that the visuals generated in ARC align with the previously described benefit of having an easy way to educate the patient, increasing their satisfaction, and ultimately improving buy-in. One participant reported that ARC data logging features directly connected to

their view of acoustics as beneficial data to track improvement over the course of therapy. Returning to the concept of time, one participant expressed that exporting RFF values to *Microsoft Excel* adds time to their evaluation. They reported that the potential benefits of using ARC would not justify the added time. As previously reported, this participant explained, “It would take a lot a lot of time...the clicking and the going back and forth and just cross-checking and having all these different options.”

Qualitative Discussion

The purpose of this study was to assess the accuracy of ARC in measuring RFF values in typical speakers and speakers with MTD compared to current semi-automated RFF estimation algorithms. Additionally, we sought clinician feedback to survey what could be improved in future iterations of the software to increase its clinical utility. To gain a better understanding of how ARC might fit into clinical voice evaluations, we aimed to assess the general structure of voice evaluations across clinicians, probing specifically about their assessment of strain and vocal effort, as well as their opinions on the benefits and barriers of using acoustic measures in their voice evaluations.

Evaluation Structure

To better understand how RFF and ARC may play a role in voice clinics in the future, it was critical to assess the general structure of the voice evaluation. Examining the similarities and differences that exist between various voice centers sets the stage for further discussion about the use of acoustics broadly, and RFF more specifically. We

hypothesized that there would be variations in the way clinicians structure their voice evaluations in the clinic. Although there were general structural similarities, we found differences in almost every component of the evaluation. One of the primary distinctions was the duration of the evaluation. With times ranging from 15 minutes to two hours, time constraint plays a fundamental role in determining what measures can be made within a timeframe. These ranges are consistent with the literature Estes and Johnson (2023). Despite all participants reporting performing a case history, the depth and number of the questions varied significantly, most likely due to the time constraint faced by the participants. For those with 90-minute evaluation blocks, more questions could be asked. Similarly, for participants with shorter evaluation blocks, strategies such as remote completion of PROMs prior to the appointment were reported, which save time and allow for focus on other aspects of the evaluation.

The greatest variability between the voice evaluations of each participant is found in the acoustic and aerodynamic measurement portions. Although many of the participants reported collecting acoustic data in a manner consistent with recent clinical voice recommendations from 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. Of relevance to this study is the finding that eight of the 10 participants reported using acoustic assessment during their evaluation. Seven of these participants reported using computer software similar to ARC to gather these acoustic values. Based on our results from 10 participants, those who already perform acoustic measures have existing acoustic measurement in their evaluations that ARC could easily fit into.

The COVID-19 pandemic had a notable impact across several domains of everyday life and societal functioning. Throughout the pandemic, healthcare workers continued to provide care, but they had to do so under modified conditions due to a nationwide lockdown. Although COVID-19 remains present, many of the initial mandatory restrictions and regulations have since been relinquished. Despite this, the pandemic seems to have had varying and lasting impacts on the voice evaluation process in clinics across the United States.

Benefits and Barriers to Acoustics

Across interviews, the participants reported 11 benefits and 13 barriers to performing acoustic assessment during a voice evaluation. Although opinions on many of the benefits and barriers were held by two or more participants, two benefits and two barriers emerged as the most prominent. These four items formed the four major themes for this section: (1) *Acoustics allow for accurate comparison* and (2) *Acoustics supplement patient-centered care* for the benefits and (3) *Acoustics take time* and (4) *Acoustics do not inform therapy patterns* for the barriers. Based on these data, when developing new acoustic software or investigating novel acoustic measures, a primary focus should be ensuring that the software and/or measure have these same benefits while also addressing or minimizing these perceived barriers. Examining RFF and its research base, the measure has been shown to reflect change pre- and post-therapy, as reported by Stepp et al. (2011) as well as differentiate maximal effort from typical effort (Lien et al., 2015). Given that ARC estimated RFF with equal accuracy to the semi-automated algorithm, RFF via ARC aligns well with the perceived benefit that *Acoustics allow for*

the most accurate comparison.

RFF via ARC does not lend itself well to the second perceived benefit of *Acoustics supplement patient-centered care*. RFF falls short in this respect in that normative values have yet to be established. A component of both subthemes, *Satisfaction, validation, and buy-in* and *Education*, was the ability to share whether a patient had typical or atypical values in reference to a given measure. Participants reported that normative values allow them to give feedback to the patient on the day of the evaluation, which results in patient satisfaction, validation, and buy-in to therapy. Normative values have been established for many acoustic (Awan et al., 2010; Baken, 1987; Casper & Leonard, 2006; Maturo et al., 2012) and aerodynamic (Weinrich et al., 2013; Zraick et al., 2012) measures. Because RFF has primarily been studied as a comparative outcome measure, it does not have the same clinical utility as other normed measures in this one way. On one hand, this directs future research to focus on studying and creating normative values for RFF. On the other hand, this poses a question to the field of voice therapy about whether normative values are necessary for all acoustics measures or if there is value in a measure purely used for comparison. The attention given to normative values throughout and across interviews serves also as a reminder to clinicians to consider what it means for an acoustic measure to have “normative values.” Just as an acoustic measure is only as good as the sample from which it is calculated, normative values are only as good as the sample from which they are derived. For example, the AVQI is a normed measure, however, these values come from a homogenous sample of 30 people (Maryn et al., 2010).

Addressing the reported barriers to acoustics, RFF estimation via ARC is quicker than the current semi-automated algorithm. However, because most participants do not already gather RFF during their voice evaluation, it would add time. Depending on whether a participant already conducts an acoustic portion in their voice evaluation, some participants perceived the few added minutes as too much to justify adding RFF via ARC to their protocol, whereas others perceived it as minimally affecting the duration of their evaluation. Interestingly, those participants with a shorter evaluation block (e.g., 15 minutes) who already collect acoustic measures reported being more willing to implement ARC than participants who had longer evaluation blocks (e.g., 45 minutes) who do not collect acoustic measures. Estes and Johnson (2023) similarly found time to be a commonly reported barrier to vocal function assessments. They discuss that time constraints may result in lack of standardization between clinicians. However, they also suggest solutions, such as using an assistant or student to conduct analyses using a standardized, scripted approach. One potentially useful direction for future research may be to investigate how much time it takes to gather RFF values in a clinical setting. Because time is one of the most commonly reported barriers to acoustics reported by clinicians within this study, it would be useful to know on average how much longer a voice evaluation (and subsequent documentation) becomes upon implementation of ARC. This way, clinicians know exactly how much time they are adding to their protocols when adopting the measure. The second thematic barrier does not apply to RFF and ARC directly. *Acoustics do not inform therapy patterns* is a theme that reflects measures for which normative values exist. As it stands, RFF values gathered on the day of the

evaluation cannot be compared to normative values, and thus may serve only as a baseline for within-person comparison.

Within this discussion on the barriers and benefits of acoustics, there were several instances in which what one participant reported as a benefit, another reported as a barrier. For example, it was commonly reported that acoustic evaluations result in improved buy-in with patients because it is satisfying and validating for them to see objective data that support their complaints. Another participant gave nuance to this view, however, stating that acoustics could negatively influence patient buy-in in some cases. They argued that because acoustic assessment can be difficult to understand and involves vocal tasks perceived as unusual by patients, patients might not understand the connection to therapy, hampering buy-in. Another commonly reported barrier was that acoustic results sometimes do not reflect what a clinician perceives. However, another participant interpreted this as a benefit, saying that in these cases, acoustic measures can detect elements of the voice and provide insights that clinicians may be missing. Many researchers have sought to document the correlation between perceptual and acoustic assessment, and they have found varying degrees of correlation for a given measure and perceived vocal quality (Narasimhan & Rashmi, 2022; Park et al., 2019; v. Latoszek et al., 2018). This poses a larger question of what is most important for evaluating a patient's voice. Is it clinician perception, patient perception, or an objective measure? If a patient is still dysphonic at the end of therapy but is satisfied with their voice, should the clinician suggest continued therapy? If the acoustic data are outside of normal limits on a specific measure, but the clinician doesn't perceive it, should it be a target in therapy?

These questions warrant further investigation.

Evaluation of Strain and Vocal Effort

There was little consensus among participants about what quantitative correlates they use to assess strain. In this study, subglottal pressure, phonatory efficiency, CPP, and spectrographic data were reported as correlates 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). Participants also reported their perception of strain in various ways: some as a shift in resonance, some as a loss of resonance, some as a general perception of sounding effortful, and some using their own synesthetic awareness of tightness in their throat when listening to the patient speak. Additionally, some reported observing other physiology such as breathing or neck tension to inform their perception of strain. The same variability was reported for the assessment of vocal effort. Four different scales were reported across the interviews along with four different perceptual observations for evaluating vocal effort. Furthermore, only one participant noted objective correlates to vocal effort (i.e., subglottal pressure and phonatory 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 objective and subjective assessment of strain and vocal effort and the prevalence of those symptoms as patient complaints, there is a need to identify standard measures that correlate with perception of strain and vocal effort. Given the sensitivity to changes in vocal strain and effort that RFF has demonstrated in pre-/post-therapy outcomes (Stepp et al., 2011) and voluntary effort

manipulation (Lien et al., 2015) as well as its predictive value of laryngeal stiffness (McKenna et al., 2016), its estimation is promising in filling this gap in acoustic measurement.

Opinions on RFF and ARC

Generally, the participants' reception of RFF and ARC was positive, especially for those who currently use acoustic data during their voice evaluations. Although most of the questions were focused on ARC, the participants expressed a strong interest in the RFF. They gave useful feedback about ARC, but many ultimately expressed that the software was only as good as the measure it calculated. Many of the participants noted that RFF had the potential to help them clinically by providing many of the benefits they attribute to acoustics at large, such as providing biofeedback, a method of progress tracking, and validating their own perception of the patient's voice. Some participants discussed expressed interest in using it as a point of comparison before and after manual therapies, such as pre and post laryngeal massage. Despite this latter use having not been investigated empirically and published, it is nonetheless an interesting point of consideration and revealing in terms of the types of tools clinicians would like to use.

Naturally, there was some hesitance in accepting RFF as an acoustic correlate of strain and vocal effort. The participants in this study work at some of the premier voice centers in the United States and have been practicing clinically without RFF to guide their treatment practices, some for as long as 34 years. For this reason, several participants expressed that they would be willing to incorporate RFF measures into their practice *if* it is a reliable and accurate measure. Regarding these concerns, the results

presented in the quantitative portion of this paper in addition to the body of literature discussed demonstrate the accuracy of RFF. Additionally, RFF has been shown to have comparable reliability to conventional measures of voice, such as vocal fundamental frequency, smoothed cepstral peak prominence, shimmer, HNR, and mean airflow rate (Park & Stepp, 2019). As more acoustic measures have been discounted over the years (e.g., jitter and shimmer), there is a general sense that participants would like to assess the robustness of RFF prior to implementing it into their protocol. This indicates that the evidence behind RFF is not the limiting factor but rather the lack of exposure of RFF to clinicians. Additionally, most of the published research on RFF has come from the STEPP Lab for Sensorimotor Rehabilitation Engineering in Boston, Massachusetts. Additional confirmatory evidence of RFF as a robust correlate of vocal strain and effort from researchers in different laboratories would ensure a reduction in bias for the measure.

Furthermore, in addition to the discussions about RFF as an acoustic measure, the participants offered several opinions about ARC that could be used to direct future development of the software for the clinic setting. The participants' ideas were clinic-oriented, and they brought to light concerns and suggestions that only those with clinical experience would have the knowledge to provide. Suggestions included integrating ARC or RFF into pre-existing acoustic software, offering more export options, and including an overlay feature to aid in data comparison. Collectively, they provided 14 actionable ideas to consider regarding the advancement of ARC, as well as a consensus on a flat-fee payment model with experience-based rationale. One aspect of improvement that was

suggested was the ability to perform sentence-level analysis to reflect a patient's daily voicing more accurately. Although ARC currently supports /afa/, /ifi/, and /ufu/, previous research, such as that of Lien et al. (2014), has been conducted using Praat and the semi-automated algorithms to analyze RFF gestures that are embedded within sentences. For example, the word *beautiful* contains the VCV sequence /ɪfə/ which meets the criteria for RFF stimuli.

Another common piece of feedback focused on the RFF and was that that ARC should incorporate RFF normative values through error bars, shading on the graph, in the data table. As discussed previously, although normative values do not currently exist for RFF, this information is telling of the kinds of data clinicians find most useful in the clinic. Additionally, participants expressed interest in using ARC for ambulatory voice monitoring device, suggesting that patients asynchronously collect data throughout the day to then send to the clinician following synchronous instruction on the software. This use of ARC would require improved noise reduction features, but otherwise would function well as a software on a patient's laptop. For true ambulatory monitoring, however, compatibility as a smartphone application is essential. This has not yet been implemented but could be a future direction for the software.

Limitations and Future Study Directions

One limitation of this study is that participants did not have hands-on access to ARC. Prior to answering questions about the software, their exposure was limited to a 15-minute narrated video about how to use the software. The limited access to the software

may have affected their ability to answer some of the proposed questions. For example, some participants were hesitant when answering the question, “How could ARC be improved for telehealth?” because they reported they needed to first use the software in person before they could determine how it could be improved for remote use. A future direction for this research would be to include a hands-on training portion and a trial period in which the participants are given the software to use for a period of time. During this time, participants could practice using the software and become more familiar with its features. Their increased familiarity and exposure with the software would likely result in more detailed and specific feedback.

Another limitation is the small sample size included in the study. The goal of any qualitative research is to achieve content saturation. Content saturation refers to the concept of gathering data until no novel ideas or opinions are presented. This is a sign that a researcher has collected all possible major concepts relevant to their research question. Although several ideas recurred across interviews and were sufficient to form distinct themes and categories, content saturation was not achieved. For example, REACT10, the final interview conducted for the study, offered a novel barrier to acoustics that no other participant had previously mentioned. This pattern occurred to some degree across all questions in the study. Similarly, there were ideas and opinions offered by participants in the initial interviews that were not mentioned by other participants later on. The result is that many questions in the study have responses backed up by only one participant. Although some proposed constructive feedback about ARC and barriers/benefits of acoustic measures were not mentioned by more than one

participant, the ideas and opinions are still valid and useful to consider. In future studies investigating this topic, the sample size should be increased to allow for content saturation.

It is important to note that 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. If presented with an additional fourth benefit, they might have agreed that they too perceived that as a benefit, despite not having thought of 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 pull up evaluation forms and protocols to reference, but many described the evaluation structure and acoustic measures from memory. Several participants at one point in the interview 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 accurate description of their evaluation process. In future related work, direct observations of voice evaluation practices may elucidate more accurate accounts of voice evaluation practices.

An additional limitation in this study was the structure and content of some of the questions asked to participants. There were a few questions included in this study that elicited different types of responses across participants, which implies that some

questions may have been ambiguous. One prompt for which this was the case, was “Tell me specifically about how you evaluate strain using perceptual measures and/or quantitative measures if any.” The intended purpose of the “perceptual” piece of this prompt was to elicit any forms or scales (e.g., GRBAS, CAPE-V, etc.) that a participant uses when assessing strain via perceptual measures. Although many participants interpreted the question this way, there were a few who attempted to describe what they were perceiving when listening to a strained voice. Although interesting, the latter interpretation is much more abstract and less relevant for the purposes of this study. In future studies, questions should be better assessed for ambiguity. Additionally, the final question, “What are some features [of ARC] that you feel are affected by your answers about the use of acoustics in your clinical practice” resulted in different types of answers as well. This question was immediately preceded by a summary of the participants’ previous answers to assist the participant in remembering their specific opinions, which added to the length of the question. Below is a typical prompt for this specific question:

And then just one final question here referring back to your previous answers about the use of acoustics in your evaluation. So, you said some benefits included using acoustics gives the client a jumping-off point, maybe helps with the buy-in, and shows that behavioral changes actually work in changing their voice; and then some barriers would be that sometimes the acoustic data isn't actually really reflective of the problems they're experiencing. So, what are some features of ARC that you feel address or are affected by these previous answers about the use of acoustics in your clinical practice?

The prompt that was used was long and increased the cognitive load on the participants at the end of a long interview, and some participants expressed confusion or difficulty in interpreting the question. For future studies, a visual could be used to summarize their previously given answers about the benefits and barriers of acoustics to reduce cognitive burden. A better way to word the question would be “Does ARC address any of these opinions?” Rewording the question may result in more accurate answers by the participants.

Conclusion

This study found that there were no statistically significant differences between RFF effect sizes when calculated via semi-automated methods versus ARC for either group. This suggests that RFF estimation performed using ARC is as accurate as current semi-automated estimation methods, and thus is an advancement of RFF estimation due to its less costly and user-friendly, automated nature. Furthermore, this study offers a detailed survey of voice-specializing clinicians’ opinions on (1) the importance of acoustics measurements, (2) methods of strain and vocal effort assessment, and (3) the adoption of ARC and RFF in their evaluation protocol. The results show that diverse opinions exist regarding the importance of acoustic measures and to what extent they should be used when making therapeutic decisions and making comparisons across therapy. For those who collect acoustic measures, there is no consensus about the acoustic measures that may be correlated with strain or vocal. The participants expressed an openness to using ARC to calculate RFF but expressed the need for greater research

and hands-on familiarity with both before implementing them within their voice evaluation protocol. Additionally, participants provided insights into what features of ARC could be modified for added to aid in its usefulness in the clinical setting.

Future studies examining perceived benefits and barriers to acoustic measures, methods of evaluating strain and vocal effort, and improvements to RFF and ARC would benefit from an increased sample size of participants from diverse training backgrounds, settings, and years of experience. The participants included in this study all currently practice at large voice centers. Voice clinicians who work in private practice or primarily telepractice settings were not surveyed but may offer unique opinions on the use of ARC and RFF in their respective settings. Future work may address the use of ARC across settings. Hands-on experience with the software may foster invaluable feedback on the practical application of the software in the clinic for both in-person and remote voice evaluations.

APPENDIX

Participant	Acoustic Measures	Aerodynamic Measures
REACT01	Mean f_0 – “we were away a year ago” f_0 Range – pitch glides CPP – /a/, /i/, “we were away a year ago” NHR – /a/, “we were away a year ago” CSID – /a/, /i/, “we were away a year ago”	Phonation Threshold Pressure – “pa pa pa pa pa” Mean Airflow – “pa pa pa pa pa” SPL Range – “we were away a year ago”
REACT02	Mean f_0 – rainbow passage f_0 Range – pitch glides CPP – rainbow passage, /a/ Jitter – /a/	Phonation Threshold Pressure – “pi pi pi pi” Mean Airflow – “pa pa pa pa pa” SPL Range – /a/ Mean Peak Air Pressure – “pa pa pa pa pa” Maximum Phonation Time – /a/ Aerodynamic Resistance Tasks – “pa pa pa pa pa”
REACT03	Mean f_0 – rainbow passage f_0 Range – pitch glides CPP – rainbow passage, /a/ RFF – VCV, sentences	Subglottal Pressure Transglottal Airflow SPL Range Mean SPL Maximum Phonation Time Phonatory Efficiency – “pi pi pi pi pi”
REACT04	Mean f_0 – connected speech, /a/, /i/ f_0 Range – pitch glides f_0 Standard Deviation – connected speech Smoothened CPP – /a/, /i/ AVQI – /a/, rainbow passage (2 nd and 3 rd sentences)	Subglottal Pressure – “pi pi pi pi pi” Average Airflow – “pi pi pi pi pi” Sustained Airflow Mean SPL – connected speech SPL Range – /a/, /i/ SPL Standard Deviation – connected speech Maximum Phonation Time – /a/
REACT05	Mean f_0 – connected speech f_0 Range – pitch glides, connected speech	Maximum Phonation Time – /a/
REACT06	ADSV – /a/, rainbow passage (2 nd and 3 rd sentences) MDVP – /a/ AVQI (occasionally)	Peak Expiratory Flow Maximum Phonation Time S/Z Ratio
REACT07	None	Maximum Phonation Time* *If concern for paralysis/paresis

REACT08	Mean f_0 f_0 Range CSID – “we were away a year ago”	Peak Airflow Mean SPL Maximum Phonation Time – /a/ S/Z Ratio
REACT09	None	Maximum Phonation Time S/Z Ratio
REACT10	Mean f_0 – rainbow passage f_0 Range – rainbow passage CPP – /a/, rainbow passage	Mean SPL – rainbow passage SLP Range – rainbow passage Maximum Phonation Time – S/Z Ratio

BIBLIOGRAPHY

- Anand, S., Kopf, L. M., Shrivastav, R., & Eddins, D. A. (2019). Objective indices of perceived vocal strain. *Journal of Voice*, 33(6), 838–845.
- Awan, S. N., Roy, N., Jetté, 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.
- 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.
- Baken, R. J. (1987). *Clinical measurement of speech and voice*. Boston: College-Hill Press.
- 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.
- 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\(2012/12-0140\)](https://doi.org/10.1044/1092-4388(2012/12-0140))
- 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 Otolaryngology, Rhinology & Laryngology*, 117(12), 919–924.
- 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.
- 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.
- 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.
- Birring, S., Prudon, B., Carr, A., Singh, S., Morgan, M., & Pavord, I. (2003). Development of a symptom specific health status measure for patients with chronic cough: Leicester Cough Questionnaire (LCQ). *Thorax*, 58(4), 339–343.

- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*, 5(9–10), 341–345.
- Brandt, J. F., Ruder, K. F., & Shipp Jr, T. (1969). Vocal loudness and effort in continuous speech. *The Journal of the Acoustical Society of America*, 46(6B), 1543–1548.
- 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.
- Camacho, A. (2012). On the use of auditory models' elements to enhance a sawtooth waveform inspired pitch estimator on telephone-quality signals. *2012 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA)*, pp. 1080–1085.
<https://doi.org/10.1109/ISSPA.2012.6310450>
- Casper, J. K., & Leonard, R. (2006). *Understanding voice problems: A physiological perspective for diagnosis and treatment*. Lippincott Williams & Wilkins.
- 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.
- Coffey, A., & Atkinson, P. (1996). *Making sense of qualitative data: Complementary research strategies*. Sage Publications, Inc.
- 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.
- Dahl, K. L., Weerathunge, H. R., Buckley, D. P., Dolling, A. S., Díaz-Cádiz, M., Tracy, L. F., & Stepp, C. E. (2021). Reliability and accuracy of expert auditory-perceptual evaluation of voice via telepractice platforms. *American Journal of Speech-Language Pathology*, 30(6), 2446–2455.
- Dworkin, J. P., Meleca, R. J., & Abkarian, G. (2000). Muscle tension dysphonia. *Current Opinion in Otolaryngology & Head and Neck Surgery*, 8(3), 169–173.
- Dworkin, J. P., Meleca, R. J., Simpson, M. L., & Garfield, I. (2000). Use of topical lidocaine in the treatment of muscle tension dysphonia. *Journal of Voice*, 14(4), 567–574.
- Elemetrics, K. (1993). *Multi-Dimensional Voice Program (MDVP)*. [Computer program.]. Pine Brook, NJ: Author.

- Elemetrics, K. (1994). CSL Computerized Speech Lab: Model 4300B, Software ver, 5. X. In: Lincoln Park, NJ: Kay Elemetrics Corporation.
- 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.
- 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.
- Gill, A., Raiff, L., Kirchgessner, E., Stepp, C. E., Kline, J. C., & Vojtech, J. M. (2023). Validation of an automated relative fundamental frequency analysis for clinical voice evaluation. [poster presentation]. *15th International Conference on Advances in Quantitative Laryngology, Voice and Speech Research*, Phoenix, AZ, USA.
- 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. e131–131. e134.
- Heman-Ackah, Y. D., Michael, D. D., & Goding Jr, G. S. (2002). The relationship between cepstral peak prominence and selected parameters of dysphonia. *Journal of Voice*, 16(1), 20–27.
- Hillman, R. E., Stepp, C. E., Van Stan, J. H., Zañartu, M., & Mehta, D. D. (2020). An updated theoretical framework for vocal hyperfunction. *American Journal of Speech-Language Pathology*, 29(4), 2254–2260. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8740570/pdf/AJSLP-29-2254.pdf>
- Hirano, M. (1981). Clinical examination of voice. *Disorders of Human Communication*, 5, 1–99.
- Hirano, M. (1981). Psycho-acoustic evaluation of voice. *Clinical Examination of Voice*, 81–84.
- 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.

- 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.
- 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.
- 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.
- Kempster, G. B., Gerratt, B. R., Abbott, K. V., 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. [https://doi.org/10.1044/1058-0360\(2008/08-0017\)](https://doi.org/10.1044/1058-0360(2008/08-0017))
- Kopf, L. M., Jackson-Menaldi, C., Rubin, A. D., Skeffington, J., Hunter, E. J., Skowronski, M. D., & Shrivastav, R. (2017). Pitch strength as an outcome measure for treatment of dysphonia. *Journal of Voice*, *31*(6), 691–696.
- Koufman, J. A., & Blalock, P. D. (1991). Functional voice disorders. *Otolaryngologic Clinics of North America*, *24*(5), 1059–1073.
- Krom, G. d. (1995). Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments. *Journal of Speech, Language, and Hearing Research*, *38*(4), 794–811.
- 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.
- Lien, Y.-A. S., Gattuccio, C. I., & Stepp, C. E. (2014). Effects of phonetic context on relative fundamental frequency. *Journal of Speech, Language, and Hearing Research*, *57*(4), 1259–1267.
- Lien, Y.-A. S., Heller Murray, E. S., Calabrese, C. R., Michener, C. M., Van Stan, J. H., Mehta, D. D., Hillman, R. E., Noordzij, J. P., & Stepp, C. E. (2017). Validation of an algorithm for semi-automated estimation of voice relative fundamental frequency. *Annals of Otology, Rhinology & Laryngology*, *126*(10), 712–716.
- Lien, Y.-A. S., Michener, C. M., Eadie, T. L., & Stepp, C. E. (2015). Individual monitoring of vocal effort with relative fundamental frequency: Relationships with aerodynamics and listener perception. *Journal of Speech, Language, and Hearing Research*, *58*(3), 566–575.

- Lien, Y.-A. S., & Stepp, C. E. (2013). Automated estimation of relative fundamental frequency. *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2136–2139. <https://doi.org/10.1109/EMBC.2013.6609956>
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage.
- 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 Otolaryngology, Rhinology & Laryngology*, *121*(8), 539–548.
- 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.
- Martins, R. H. G., do Amaral, H. A., Tavares, E. L. M., Martins, M. G., Gonçalves, T. M., & Dias, N. H. (2016). Voice disorders: etiology and diagnosis. *Journal of Voice*, *30*(6), 761. e761–761. e769.
- Maryn, Y. (2017). Practical acoustics in clinical voice assessment: A Praat primer. *Perspectives of the ASHA Special Interest Groups*, *2*(3), 14–32.
- Maryn, Y., Corthals, P., Van Cauwenberge, P., Roy, N., & De Bodt, M. (2010). Toward improved ecological validity in the acoustic measurement of overall voice quality: combining continuous speech and sustained vowels. *Journal of Voice*, *24*(5), 540–555.
- 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*(10), 956–961.
- 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.
- McKenna, V. S., Heller Murray, E. S., Lien, Y.-A. S., & Stepp, C. E. (2016). The relationship between relative fundamental frequency and a kinematic estimate of laryngeal stiffness in healthy adults. *Journal of Speech, Language, and Hearing Research*, *59*(6), 1283–1294.
- McKenna, V. S., & Stepp, C. E. (2018). The relationship between acoustical and perceptual measures of vocal effort. *The Journal of the Acoustical Society of America*, *144*(3), 1643–1658.

- McKenna, V. S., Vojtech, J. M., Previtera, M., Kendall, C. L., & Carraro, K. E. (2022). A Scoping Literature Review of Relative Fundamental Frequency (RFF) in Individuals with and without Voice Disorders. *Applied Sciences*, *12*(16), 8121.
- Morrison, M. D., Rammage, L. A., Belisle, G. M., Pullan, C. B., & Nichol, H. (1983). Muscular tension dysphonia. *The Journal of Otolaryngology*, *12*(5), 302–306.
- Morse, J. M. (2015). Critical analysis of strategies for determining rigor in qualitative inquiry. *Qualitative Health Research*, *25*(9), 1212–1222.
- Nagle, K. F. (2022). Clinical use of the CAPE-V scales: Agreement, reliability and notes on voice quality. *Journal of Voice*. <https://doi.org/10.1016/j.jvoice.2022.11.014>
- Narasimhan, S., & Rashmi, R. (2022). Multiparameter voice assessment in dysphonics: correlation between objective and perceptual parameters. *Journal of Voice*, *36*(3), 335–343.
- Paes, S. M., & Behlau, M. (2017). Dosage dependent effect of high-resistance straw exercise in dysphonic and non-dysphonic women. *CoDAS*, *29*(1), e20160048. <https://doi.org/10.1590/2317-1782/20172016048>
- Park, J. W., Kim, B., Oh, J. H., Kang, T. K., Kim, D. Y., & Woo, J. H. (2019). Study for correlation between objective and subjective voice parameters in patients with dysphonia. *Journal of The Korean Society of Laryngology, Phoniatrics and Logopedics*, *30*(2), 118–123.
- Park, Y., Cádiz, M. D., Nagle, K. F., & Stepp, C. E. (2020). Perceptual and acoustic assessment of strain using synthetically modified voice samples. *Journal of Speech, Language, and Hearing Research*, *63*(12), 3897–3908.
- Park, Y., & Stepp, C. E. (2019). Test–retest reliability of relative fundamental frequency and conventional acoustic, aerodynamic, and perceptual measures in individuals with healthy voices. *Journal of Speech, Language, and Hearing Research*, *62*(6), 1707–1718.
- 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.
- Robb, M. P., & Smith, A. B. (2002). Fundamental frequency onset and offset behavior. *Journal of Speech, Language, and Hearing Research*, *45*(3), 446–456. [https://doi.org/10.1044/1092-4388\(2002/035\)](https://doi.org/10.1044/1092-4388(2002/035))

- Robertson, R. J., Goss, F. L., Rutkowski, J., Lenz, B., Dixon, C., Timmer, J., Frazee, K., Dube, J., & Andreacci, J. (2003). Concurrent validation of the OMNI perceived exertion scale for resistance exercise. *Medicine & Science in Sports & Exercise*, 35(2), 333–341.
- 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.
- Rosenthal, A. L., Lowell, S. Y., & Colton, R. H. (2014). Aerodynamic and acoustic features of vocal effort. *Journal of Voice*, 28(2), 144–153.
- Roy, N., Fetrow, R. A., Merrill, R. M., & Dromey, C. (2016). Exploring the clinical utility of relative fundamental frequency as an objective measure of vocal hyperfunction. *Journal of Speech, Language, and Hearing Research*, 59(5), 1002–1017.
- Roy, N., Ford, C. N., & Bless, D. M. (1996). Muscle tension dysphonia and spasmodic dysphonia: the role of manual laryngeal tension reduction in diagnosis and management. *Annals of Otology, Rhinology & Laryngology*, 105(11), 851–856.
- 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.
- Stepp, C. E., Hillman, R. E., & Heaton, J. T. (2010). The impact of vocal hyperfunction on relative fundamental frequency during voicing offset and onset. *Journal of Speech, Language, and Hearing Research*, 53(5), 1220–1226. [https://doi.org/10.1044/1092-4388\(2010/09-0234\)](https://doi.org/10.1044/1092-4388(2010/09-0234))
- Stepp, C. E., Merchant, G. R., Heaton, J. T., & Hillman, R. E. (2011). Effects of voice therapy on relative fundamental frequency during voicing offset and onset in patients with vocal hyperfunction. *Journal of Speech, Language, and Hearing Research*, 54(5), 1260–1266. [https://doi.org/10.1044/1092-4388\(2011/10-0274\)](https://doi.org/10.1044/1092-4388(2011/10-0274))
- v. Latoszek, B. B., Maryn, Y., Gerrits, E., & De Bodt, M. (2018). A meta-analysis: Acoustic measurement of roughness and breathiness. *Journal of Speech, Language, and Hearing Research*, 61(2), 298–323.
- van Leer, E., & van Mersbergen, M. (2017). Using the Borg CR10 physical exertion scale to measure patient-perceived vocal effort pre and post treatment. *Journal of Voice*, 31(3), 389. e319–389. e325.
- 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.

- van Mersbergen, M., & Delany, M. (2014). Vocal responses to emotional picture viewing. *Logopedics Phoniatrics Vocology, 39*(3), 99–107.
- van Mersbergen, M., Lyons, P., & Riegler, D. (2017). Vocal responses in heightened states of arousal. *Journal of Voice, 31*(1), 127. e113–127. e119.
- Van Mersbergen, M., Patrick, C., & Glaze, L. (2008). Functional dysphonia during mental imagery: Testing the trait theory of voice disorders. *Journal of Speech, Language, and Hearing Research, 51*(6), 1405–1423.
[https://doi.org/10.1044/1092-4388\(2008/06-0216\)](https://doi.org/10.1044/1092-4388(2008/06-0216))
- van Mersbergen, M., Vinney, L. A., & Payne, A. E. (2020). Cognitive influences on perceived phonatory exertion using the Borg CR10. *Logopedics Phoniatrics Vocology, 45*(3), 123–133.
- Verdolini, K., Titze, I. R., & Fennell, A. (1994). Dependence of phonatory effort on hydration level. *Journal of Speech, Language, and Hearing Research, 37*(5), 1001–1007.
- Vojtech, J., & Heller Murray, E. (2019). Tutorial for manual relative fundamental frequency (RFF) estimation using Praat. In: Boston University, Boston, MA.
- Vojtech, J. M., Cilento, D. D., Luong, A. T., Noordzij Jr, J. P., Diaz-Cadiz, M., Groll, M. D., Buckley, D. P., McKenna, V. S., Noordzij, J. P., & Stepp, C. E. (2021). Acoustic identification of the voicing boundary during intervocalic offsets and onsets based on vocal fold vibratory measures. *Applied Sciences, 11*(9), 3816.
- Vojtech, J. M., Segina, R. K., Buckley, D. P., Kolin, K. R., Tardif, M. C., Noordzij, J. P., & Stepp, C. E. (2019). Refining algorithmic estimation of relative fundamental frequency: Accounting for sample characteristics and fundamental frequency estimation method. *The Journal of the Acoustical Society of America, 146*(5), 3184–3202.
- Watson, B. C. (1998). Fundamental frequency during phonetically governed devoicing in normal young and aged speakers. *The Journal of the Acoustical Society of America, 103*(6), 3642–3647.
- Webb, A., Carding, P., 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*, 429–434.
- 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.

- Yumoto, E., Gould, W. J., & Baer, T. (1982). Harmonics-to-noise ratio as an index of the degree of hoarseness. *The Journal of the Acoustical Society of America*, *71*(6), 1544–1550.
- 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\(2010/09-0105\)](https://doi.org/10.1044/1058-0360(2010/09-0105))
- 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.

CURRICULUM VITAE

