Data Collection Methods for the Voice Range Profile: A Systematic Review

by

Grace McKinnon Cutchin

A thesis submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Auburn, Alabama
May 3, 2020

Keywords: voice range profile, VRP, phonetogram, data collection,
f_0 range, SPL range

Approved by

Mary J. Sandage, Chair, Associate Professor of Communication Science Disorders
Laura W. Plexico, Professor of Communication Science Disorders, Department Chair
Aurora Weaver, Assistant Professor of Communication Science Disorders

Abstract

Purpose: To assess data collection variability in the Voice Range Profile (VRP) across clinicians and researchers, a systematic review was conducted to evaluate the extent of variability of specific data collection points that affect the determination of frequency range and sound level and determine next steps in standardization of a VRP protocol.

Method: A systematic review was conducted using the Preferred Reporting Items for Systematic Review and Meta-Analysis checklist (PRISMA). Full text journal articles were identified through PubMed, Web of Science, Psych Info, ProQuest Dissertations and Theses Global, Google Scholar and hand searching of journals.

Results: A total of 1134 articles were retrieved from the search; of these 463 were duplicates. Titles and abstracts of 671 articles were screened, with 203 selected for full-text review. Fifty-four articles were considered eligible for inclusion. The information extracted from these articles revealed the methodology used to derive the VRP was extremely variable across the data points selected. Additionally, there were 8 common acoustic measures used for statistical analysis described in included studies that were added as a data point.

Conclusions: The data collection methods for the VRP varied considerably. Standardization of procedures were recommended for clinicians and researchers.

Acknowledgments

I would first like to thank Dr. Sandage, my committee chair, for her diligence, patience, guidance, and encouragement throughout this process. Not only have you impacted my understanding of research and writing, but you have also significantly contributed to my own personal growth and character throughout this process. I would also like to acknowledge the other members of my thesis committee, Dr. Weaver and Dr. Plexico: thank you for the time, the support, and the wisdom you so willingly shared in contribution to this project. Additionally, I would like to thank Adelia Grabowsky, the Auburn University Health Sciences Librarian, for your insight whenever questions arose regarding the process of the systematic review. Lastly, I would like to thank my parents, Barry and Martha McKinnon, for instilling within me a love of learning and to my husband, Brian Cutchin, for being my constant encourager for the duration of this project.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

The Voice Range Profile (VRP), or phonetogram, quantifies laryngeal function in regards to the fundamental frequency (f_0) and sound level (dB SPL). When referring to amplitude, both SPL and intensity have been used in recent literature. However, the emergence of interdisciplinary studies of voice function have determined intensity nomenclature should be updated to the term sound level to avoid confusion with any references to exercise intensity (Hoch & Sandage, 2017). Refer to Figure 1 for an example of a VRP.

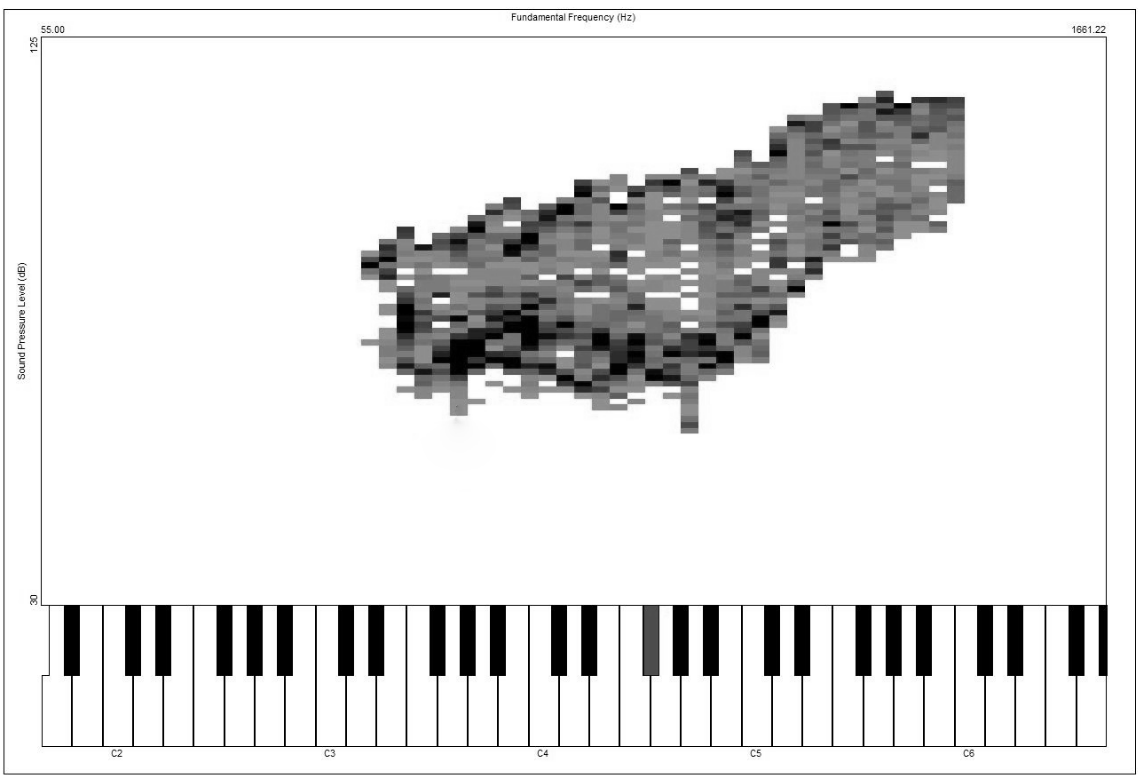

Figure 1. Example VRP created with Voice Range Profile (CSL, PENTAX Medical, Tokyo, Japan).

The x-axis demonstrates the lowest and highest frequencies the vocal folds can produce and the y-axis shows the maximum and minimum sound level (Schutte & Seidner, 1983). Before the term "voice range profile" was officially proposed by the Voice Committee of the International Association of Logopedics and Phoniatrics (Bless et al., 1992), there were many terms used including: phonetogram (Airainer &Klingholz, 1993; Gramming & Sundberg, 1988), phonetography (Heylen, Wuyts, Mertens, & Pattyn, 1996), voice profile (Böhme & Stuchlik, 1995), phonational profiles (Brown, Morris, Hicks, & Howell, 1993), and voice area (Ma & Yiu, 2011; Schutte & Seidner, 1983; Titze, 1992).

History of the Voice Range Profile

Interest in the relationship between f_0 and sound level has been discussed since the 1930s (Stout, 1938; Wolf, Stanley, & Sette, 1935). Wolfe, et al. (1935), presented a graph plotting vocal power, or sound level, as a function of pitch range on the four vowels, "ah," "ay," "oo", and "ee." These authors compared the sound level of classical singers against non-singers, the variance in medium and wide vibratos, and also included a discussion about the change in voice quality throughout the range of both women and men (Wolf et al., 1935). Stout (1938) evaluated the independent effect in changes of pitch and sound level using three vowels, "ah," "oo," and "ee." He created a graph of a single subject's functional singing area for each vowel and included graphs showing changes in the harmonic structure for each pitch and vowel. Stout (1938) described the acoustic differences of the vowels, noting the distinct differences above and below 1800 Hz. Damsté (1970) popularized the VRP in his publication "The Phonetogram." In this article, Damsté argued that description of the laryngoscopic image could not be used to describe the properties of the voice and provided the procedures of his phonetogram to be used to evaluate voice quality (Damsté, 1970). The use of the VRP was continued into the 1980's as the Union of

European Phoniatricians (UEP) presented a standardized protocol to be used for the VRP (Schutte & Seidner, 1983) and Gramming and Sundberg (1988) evaluated spectrum factors that should be taken into consideration when using the VRP.

Determining the Voice Range Profile

Extreme sound pressure level (SPL) values produced by the voice may reflect certain aspects of vocal fold vibration and can potentially reveal relevant aspects of laryngeal function (Gramming & Sundberg, 1988). For instance, an individual's VRP may reveal areas of sound pressure level (SPL) or f_0 discontinuities, where the upper and/or lower contours experience a notch or a spike. This typically means there is a lack of control, like a register shift, or there is a physical problem, like nodules (Coleman, 1993). Because this tool is focused on finding the extremes of vocal function, the minimum and maximums of the voice should be quantified by producing a pitch in any manner without regard to mode (register), sample type (vowel, CVC), or other specific constraints that may be placed on the laryngeal, respiratory, and supralaryngeal systems (Coleman, 1993). Originally, VRPs were collected with a simple sound level measuring device and tone generator as the patient would hold the tone for 2-3 seconds. The first computerized VRP technology was commercialized in 1983 (Printz, Rosenberg, Godballe, Dyrvig, & Grøntved, 2018) and was later published in *The Journal of Hearing Research*, thus providing a new method to measure VRP (Sulter, Wit, Schutte, & Miller, 1994). The computerized VRP technology has developed over the last few decades, allowing for faster sampling and more comprehensive mapping of an individual's entire range of frequency and SPL function (Coleman, 1993).

VRP's have a distinct and recognizable shape (Figure 1). Typically, there is a gradual upward tilt in the upper and lower contours, indicating there may be a systematic increase in the average sound level with f_0 (Titze, 1992). Variations in shape are expected between men and women and due to vowel choice. It is difficult to have independent control over f_0 and intensity: they tend to covary in speech and singing. In speech, syllable stress is often accompanied by an f_0 rise. In singing, it is difficult to sing a high note softly and to sing a low note loudly. When singing at a higher frequency, the harmonics add almost no power to the f_0 because they are above the first and second formant, which would indicate there is no reinforcement from the vocal tract. Therefore, the amplitude of higher harmonics is determined almost solely by the f_0. There are two reasons why the slope of the contour increases about 8-9 dB per octave. First, subglottal pressure needs to increase with the f_0 to stay above the phonation threshold pressure. Second, the vocal tract radiates power more effectively as the f_0 is raised (Titze, 1992).

Clinical Use

Despite its exclusion from recommended basic instrumental protocols for assessment in vocal function (Patel et al., 2018), the VRP remains a compelling measure for the assessment of voice. It is considered a useful tool in the evaluation of therapy effects (Speyer, 2008) and pre/post-surgical vocal function (Rendón et al., 2018; Salmen et al., 2018). Specifically, VRPs may be used to assess the vocal performance of the normal voice (Schutte, 1980), the potential of a singer's voice (Seidner, Wendler, Wagner, & Rauhut, 1981), and as a diagnostic outcome for voice therapy and treatment (Schutte, 1980).

Use of the VRP in clinical voice assessment can identify the extent of functional impairment of vocal ability. Physiologically, an individual with respiratory muscle weakness and excellent laryngeal function would not build the same VRP profile as an individual with excellent pulmonary and laryngeal function. It is probable that the former individual would experience difficulty producing sufficient subglottal pressure required to produce adequate functional loudness, despite having excellent laryngeal function. Conversely, we would anticipate a trained classical singer to have greater facility in producing high notes softly because of the specific training that the individual received.

VRP may also be valuable as a means to track vocal changes in function during maturation, as a means to differentiate trained voices from untrained voices, or changes in the pathological conditions of the system as a whole (Coleman, 1993). It provides performers with a tool to assess the risk associated with a specific f_0 and dB (SPL) range of a particular opera role relative to their own voice ability. For example, when a tenor is considering an operatic role, he might compare the range as well as the tessitura of the role with his own VRP. If the tessitura of the role sits comfortably within his vocal limits, then this role would most likely be appropriate. However, if the vocal requirements of the role extended beyond the outer limits of the tenors range, then he might consider turning down the role as he would be at a greater risk for developing a voice disorder.

For the VRP to be useful as a reliable, repeatable measure of voice function, the methodology must be consistently applied and efficient to avoid fatiguing the individual. Researchers and clinicians who use standardized methodology will have greater confidence when interpreting VRP outcomes. Currently, the literature describes wide variations for the procedures used for collecting VRPs (Awan, 1991; Coleman, 1993; Lycke & Siupsinskiene, 2016; Ma & Yiu, 2011; Pabon & Plomp, 1988; Schutte & Seidner, 1983; Speyer, Wieneke, Wijck-Warnaar, & Dejonkere, 2003; Van Mersbergen, Verdolini, & Titze, 1999). Variance in procedural methodology used to determine the VRP makes it difficult for clinicians and researchers to determine whether VRP outcomes described across studies or clinics occurred as a result of a medical, behavioral, or research intervention or as a result of procedural differences (Coleman, 1993).

Review of Procedural Differences

Methodological differences abound for the collection of the VRP for research and clinical application. These procedural variances contribute to the outcomes of the extremes of the participant's frequency range, or the extremes of the participant's sound level, or both. They are related to the wide technological resources available to collect VRPs, differing opinions of experts in the field, whether the end goal of the VRP is physiological or functional, and consistency between clinicians. Standardizing procedures for VRPs would improve reliability within and between clinicians and researchers. To advance evidence-based practice, research using VRP as an outcomes measure requires consistent methods for reliable pre and post testing. Consistent research methods within and between researchers creates opportunities for meta-analyses of data across investigations.

Procedural variance related to technological resources. Variance among VRPs may be attributed to the change from the conventional means of collecting a VRP, using a keyboard and sound decibel meter, to the automated computer program collection. The traditional means requires the participant to be able to match a specific pitch as the clinician is taking the information and necessitates more coaching to be provided by the clinician. In the traditional approach, the clinician graphs the points in whatever manner he or she feels is appropriate, but in many cases the recommended procedures of the Union of European Phoniatricians (Schutte & Seidner, 1983) are used. The UEP recommends collecting data points which are located at intervals of 10% throughout the participants range and then the points are connected to create the full VRP (Schutte & Seidner, 1983).

Automated VRP Determination. While the shift to computer-based VRP systems removed the necessity of hand-graphing the data point, it introduced systematic differences between platforms and equipment used. There are many platforms available that facilitate the collection of a VRP which include the following; Phog (Hitech Development AB, Täby, Sweden), Voice Profiler (Alphatron Medical Systems, Rotterdam, The Netherlands), Voice Range Profile (CSL, PENTAX Medical, Tokyo, Japan), Dr. Speech: Phonetogram (Tiger Drs Inc., Seattle, Washington, USA), lingWAVES Voice Diagnostic Center (WEVOSYS Medical Technology GmbH, Baunach, Germany), and DiVAS software (XION GmbH, Berlin, Germany).

In addition to platform differences there are also differences observed with regard to the equipment used to collect the data. For example, microphone differences are often observed in regards to the specifications associated with each microphone and the position of the microphone, e.g., headset or handheld. Calibration procedures and methods to account for ambient noise also differ between those platforms listed above. Because of the differences in computer systems used and the variations in the additional equipment needed, the VRP may derive outcomes that differ markedly from those derived with the traditional handheld equipment.

Microphone. There are many factors related to the microphone used for collection of VRP that provide a great deal of variance. It is important to realize that in research for speech and voice, the microphone is used to convert sound pressure signals to an electronic signal with the same characteristics. Most microphones are not developed for these purposes, but are developed for recording or broadcasting (Svec & Granqvist, 2010). When choosing a microphone the following needs to be considered; frequency and range of microphone, directionality, the dynamic range, the transducer type used, the microphone preamplifier used (Svec & Granqvist, 2010) and the distance from the mouth to the microphone (Coleman, 1993). In ASHA's 2010 recommendations for microphone, clear specifications are provided. These recommendations include information regarding the lower and upper dynamic limits of the microphone, the lower and upper frequency limits of the microphone, and frequency response details (Svec & Granqvist, 2010). Because of the many specifications related to microphones and the cost of high-quality microphones, they may cause considerable variation in the data collection of VRPs.

Variability in suggested graphing forms. When creating the VRP, there are many recommendations in which one can graph data points. Schutte and Seidner (1983) recommended a very specific graph with 10dB increments going from 40 dBA SPL to 120 dBA SPL on the y axis. This varies from the graph suggested by Pabon and Plomp (1988), as they used a SPL range from 30-110 and a frequency range of 1-1400 Hz. There are also specific graphs used for each of the computer systems that maintain a level of variance from one another. The differences among graphs make it difficult to compare VRPs across studies or clinics unless the graphs are all converted to the same protocol.

Graphing is complicated further when one begins to look at the multiple ways in which data points may be plotted and collected. Some VRPs are plotted using 10% increments within the participants range and connecting the points (Coleman, 1993; Damsté, 1970; Schutte & Seidner, 1983) and some VRPs are collected by having the participant sing every semitone within their range at the softest and loudest sound level (Heylen, Wuyts, Mertens, De Bodt, & Van de Heyning, 2002; Ma & Yiu, 2011). Others have the participant do a glissandi or glide from the lowest to the highest point of their range as loud as they can and then as soft as they can and accept this production as the full VRP (Speyer et al., 2003).

This difference in graphs used, VRP points taken, and technology used may contribute to VRP data variance. These methodological differences hinder use of the VRP for reliable comparison across studies or even between different clinicians. There have been studies which used a method initially created for the traditional measure of VRP and modified them to be appropriate measures for a computer VRP system. For example, in her study on the VRPs of choral teachers, Schwartz (2009) used the methods laid out by Awan (1991) and formatted the computer program to generate the output in the same format. This provided evidence that the traditional method of creating a VRP and the computerized method can be aligned to have comparable methodology to create the VRP.

Procedural variance related to extremes of frequency range.

Vowel. The vowel used to create a VRP has remained inconsistent across methodologies from the VRP's inception and continues to be a point of discussion. It is widely accepted practice to use a consistent vowel throughout the VRP as was recommended by the UEP (Schutte & Siedner, 1983) and Gramming and Sundberg (1988). The use of a consistent vowel is considered an important feature as it maintains the same formant frequencies throughout the participant's range and extremes. In general, the lower contour of the VRP is dependent on sound level of the f_0, while the upper contour depends largely on the harmonics associated with the f_0. The individual characteristics of certain vowels or a speaker's articulatory patterns can complicate these relationships (Gramming & Sundberg, 1988).

The most popular vowel to use in ascertaining the VRP is the /a/ vowel, because the high first formant avoids interaction with the f_0 on the majority of phonations in the VRP (Schwartz, 2009). The UEP suggests using the vowels /a/, /i/, and /u/ to create multiple color-coded VRP plots in the same graph, asserting that this juxtaposition of differing vowels offers more comprehensive acoustic data, especially in the case of singers (Schutte & Seidner, 1983). Gramming and Sundberg (1988) recommended the use of only the /a/ vowel and suggested that the vowels /i/ and /u/ do not offer additional acoustic information. Upon comparison of the /i/ and /a/ phonetograms taken by Gramming and Sundberg (1988), Titze (1992) found there was a systematic increase in the average intensity with the f_0 in the graphs of the /a/, but with the vowel /i/ there was a restricted intensity range for the low frequencies and less of a tilt in the oval shape as compared to the /a/. This is relevant when one considers most classical singers have vowels they prefer to use when singing through a particular region of their own registration. These vowel preferences differ across individuals for the same pitch class promoting the assumption that a singer may be able to sing louder or softer on their "favorite vowel" as opposed to a different vowel.

Physiological versus functional. The end goal of the VRP is a contributing factor in the data collection methods of a researcher or clinician. If the clinician is trying to determine the physiological limits of the voice, then every sound the participant can produce will be considered a data point. This means a "squeak" at the uppermost outer limit of the range will be included, even if they cannot hold out the frequency for a second or more. However, if the clinician wants to create a functional VRP, then only the frequencies that can be produced for a full second or more will be included as data points.

Registration. Singers who have learned manipulation of their vocal mechanism will know how to navigate the three modes of laryngeal function (pulse, modal, and falsetto) with more skill than a participant who has no vocal training. To complete the VRP, the voice data collected does not have to be beautiful. Instead, “unsingable” or “ugly” sounds should still be recorded as a means of quantifying functional voice ability in a general sense, but not necessarily in a performance sense. In this instance, it will be important to coach both the trained and untrained voice user to use their extreme limits and not to control their register breaks. Lycke and Siupsinskiene (2016) considered the break as a parameter in their study comparing singers to non-singers, but few studies control this variable among participants (Coleman, 1993).

Range. The way in which the participant’s range is determined may be done with the use of a glissandi, or glide, or with a steady state production. Glissandi often yield a significantly larger semitone range by about 2 to 3 semitones (Reich, Frederickson, Mason, & Schlauch, 1990). This can be problematic if the procedure requires the participant to produce sustained productions when facilitating the reading of the sound level meter. In doing so, the physiological limits of the voice versus the functional limits of the voice contrast, meaning there is a difference between what is vocally possible and what is vocally probable. Coleman (1993) suggests using discrete steps in the downward direction followed by a glissando down. The subject is then asked to produce the lowest pitch three times to determine if it is replicable and controllable. This same procedure is repeated to find the highest note in the participants range.

Another procedure that does not use the glissandi to determine the outliers of the range was described by Ma and Yiu (2011). This method begins with the habitual pitch level progressing down by half step. The client is asked to match the pitch at a comfortable level and then get as soft as possible. The participant repeats the same protocol for the upper range. After all of the softest sounds are collected, then these procedures for the lower and upper range are repeated to obtain the loudest sound levels. Clinicians use different yet effective methods to collect a participant's range, but there is a probability of variance in the measures taken simply because each method is different from the other.

Duration of target tone. There is also a great deal of variance among studies about the duration of the target tone. In the conventional means of collecting the VRP, it is important for the production at the target frequency to last for a few seconds, usually two or three, for the clinician to accurately read the sound level meter. Coleman et al. (1977; 1978) required that productions be sustained for at least 2 seconds and Pabon and Plomp (1988) required participants to sustain the tone for several seconds. However, other researchers accepted short phonation times of less than a second when using computer systems (Pabon, 1991; Speyer et al., 2003). Therefore, the conventional means of collecting VRPs will only create a functional VRP, but the computer systems could create a physiological VRP or a functional VRP depending on the procedures outlined by the clinician or researcher.

Vibrato versus steady-state production. Vibrato is a normal occurrence in singing, so it seems only natural that trained singers would use it when creating a VRP. However, when using vibrato, a participant is violating the basic requirement that a single target frequency is produced. Vibrato usually varies in both frequency, from .5 to 2 semi-tones, and sound level, from 2 dB to 10 dB (Rothman, Nielsen, & Hicks, 1979). This adds another layer of variability, as this can often be hard to control among participants depending on their background (Coleman, 1993).

Warm-up. Another factor that can influence the extremes of the vocal range, is warm-up. Several studies included warm-up as a part of the initial training process. The investigator would have the client glide up and down as they were learning the procedure (Coleman, 1993; Ma & Yiu, 2011), or have the participant sustain an /a/ at their habitual pitch and explore their voice down and up (Heylen et al., 2002). While this helps warm up the range of the participant to prepare for an accurate VRP, it also gives the clinician an idea of the expected range in regards to the physiological and functional outer limits.

Time of Day. Many singers and voice pedagogues attest to the fact that it is often difficult to sing in the morning. Because of this, there has been speculation about the time of day the VRP is taken having an effect on the voice and the extreme range parameters. However, there is not enough evidence to support this assertion. A study conducted by Van Mersbergen, Verdonlini, and Titze (1999) compared evening and morning VRPs within subjects and showed only minimal systematic changes to the participant's VRPs. Despite the limited evidence, Ma and Yiu (2011) recommend taking pre and post treatment samples at the same time of day.

Procedural variance related to extremes of sound level.

Repeated versus single productions. It is unlikely a participant will produce the softest or loudest tone in their initial effort at producing the target tone. It has been suggested that it is helpful to have the participant do the softest and loudest productions three times or more to obtain the most accurate dB readings of the loudest and softest productions (Coleman, 1993; Ma & Yiu, 2011). Other studies have had the participant produce only a single production of the softest sound on the target pitch followed by a single production of the loudest sound on the same target pitch (Lycke & Siupsinskiene, 2016; Van Mersbergen et al., 1999). It should be noted that Lycke's (2012, 2013; 2016) studies were performed only on trained singers.

dBA versus dBC. There have been several recommendations about the scale used for the SPL. Both the UEP (Schutte & Seidner, 1983) and Gramming and Sundberg (1988) recommended measuring the SPL with the A-weighted curve (dBA). This scale is often used because the sound is filtered in a way that models the filter of the human ear. This is done by attenuating the low frequency sounds. However, Coleman (1993) argues that dBA is "biased" against frequencies of less than 200 Hz and influences the SPL values obtained. He recommends using a dB(LIN), like that of the dBC scale, because it produces higher SPL values than the dBA scale. Use of the dBC scale is recommended in standard protocols for instrumental assessment of voice function (Patel et al., 2018) because it would provide uniform measurement of the frequency range and would not discriminate against the lower frequencies often found in speech and singing. Many studies have adopted the use of the dBA scale (Camarrone, Ivanova, Decoster, De Jong, & Van Hulle, 2015; Lycke, Decoster, Ivanova, Van Hulle, & De Jong, 2012; Lycke, Ivanova, Van Hulle, Decoster & DeJong, 2013; Lycke & Siupsinskiene, 2016), but there are several studies that do not state the scale used (Ohlsson et al., 2018; Rendón et al., 2018).

Mouth opening and breathy tone. Trained singers versus participants who have no vocal training may demonstrate a difference in the amount of mouth opening and the amount of breathy versus non-breathy tone. Regarding mouth opening, more constriction means there will be less vocal output. Trained singers may modify the mouth opening, which can change the sound level up to 30 dB, but untrained singers will have little awareness of the shape of their mouth (Sundberg, 1987). A breathy tone and non-breathy tone can also affect the overall sound level outcome of a VRP. A tone with less noise will produce a greater sound level, meaning that a non-breathy tone will be a louder sound level than a breathy tone. Comparison of the same pitch with and without noise reveals a 15dB difference across VRPs (Coleman, 1993).

Procedural variance related to extremes of frequency range and sound level.

Training or coaching provided by clinician. The amount of training and coaching provided by the researcher or clinician, as well as the subjective judgment as to whether what has been gathered is a representation of the participant's vocal function will also contribute to variation (Coleman, 1993). In both the conventional and computerized methods, it is important to explain the equipment and what is expected to the participant before they begin the trial. The explanation of the equipment and data collection methods likely lacks standardization between clinicians and researchers.

When collecting the conventional VRP, the clinician has to be in the room to measure the sounds demonstrated and then the clinician is able to determine whether the measurements collected are perceived to be accurate. Because the clinician collecting this data is familiar with the task and has an expected value for the softest and loudest phonations, they might then encourage the participant with gestures, facial expressions, and other body language to elicit the best results. In situations with an untrained singer there may be complications with matching the appropriate pitch. In these instances, the clinician is expected to know how to troubleshoot the situation and collect the most representative data.

Matching the target tone might be circumvented to a degree with a computer system because there is visual feedback provided for the participant. However, they may still require coaching from the clinician to understand and produce the target tone and modulate sound level. Even trained singers may require coaching to elicit the softest and loudest or the highest and lowest pitches because most singers do not want to produce a sound they would not sing in public (Coleman, 1993). Many studies using computer automated systems still have a researcher coach and train the participants with the unfamiliar task of collecting the VRP (Lycke & Siupsinskiene, 2016; Ma & Yiu, 2011; Speyer et al., 2003), but because the computer systems do not require a clinician to be present, there have been studies that do not have a coach present during collection of the VRP (Van Mersbergen et al., 1999).

Procedural variables that require consideration.

Room acoustics. The room acoustics of the recording environment has been discussed at length. Ma and Yiu (2011) state that poor acoustics can affect the validity of the sound levels measured. The UEP (Schutte & Seidner, 1983) recommends using a room with living room acoustics is optimal, but Coleman (1993) argues that the size and absorption characteristic of the room environment do not limit recording. Both Coleman (1993) and Ma and Yiu (2011) recommend the environment should be quiet, with 40dB SPL or less ambient noise. The best way to control this environment is to record in a sound treated booth (Coleman, 1993; Lycke & Siupsinskiene, 2016; Van Mersbergen et al., 1999) or to monitor the environment with a sound level meter as the recording is taking place (Schwartz, 2009).

Justification

A systematic comparison of the methods affecting the determination of the most representative profile of the voice should be conducted for three primary reasons: 1) the methodology for collection of the VRP varies; 2) it is a valuable clinical and research assessment of vocal physiology and function; and 3) it serves as a vocal function outcome measure that is useful after both surgical and behavioral intervention. This review will focus on outcomes based literature including only studies that measure pre/post VRP outcomes in therapy or treatment intervention and studies comparing VRPs across participants. Because a VRP for research purposes and a VRP for clinical measures can be very different, this will allow similar studies to be compared providing useful information for speech pathologists in clinical practice. Extremes of the participant's range might be affected by one or more of the following; the vowel used to create the VRP, the way in which the initial range is established, the amount of time each pitch is held, sung or unsung sustained phonation, use of vibrato or steady-state production, and the use of warm-up. Inclusion or exclusion of coaching from the clinician varies widely across studies. This component factors into finding the true extremities of the range, but also plays a factor in determining whether a sound is the loudest or softest production. The loudest and softest productions might also be affected by; the number of trials used to attain the SL, the equipment used, a scale of dBA or dBC, and the microphone to mouth distance. The following are hypothesized for the variables chosen for review:

- /a/ will be the most common vowel used;
- A glide will be the most common way the range is established;
- Pitches will be held for one to two seconds;
- Phonations will be unsung and steady-state;
- Warm-up will not be included as a part of the collection procedure;
- There will be little coaching by the clinician reported;
- There will be variance in the numbers of trials used to attain the sound level;
- There will be variance in the equipment used;
- dBA will be used more often than dBC; and,
- There will be variance in the distance from the microphone to the mouth.

To the author's knowledge, one other systematic review has been performed regarding the VRP evaluating the reproducibility of the automated VRP (Printz et al., 2018). Study findings will be useful to standardize data collection practices to improve test-retest reliability.

Chapter 2. Manuscript

Introduction

The voice range profile (VRP), or phonetogram, quantifies laryngeal function in regards to the fundamental frequency (f_0) and sound level (dB SPL). Before the term "voice range profile" was officially proposed by the Voice Committee of the International Association of Logopedics and Phoniatrics (Bless et al., 1992), there were many terms used including; phonetogram (Airainer &Klingholz, 1993; Gramming & Sundberg, 1988), phonetography (Heylen et al., 1996), voice profile (Böhme & Stuchlik, 1995), phonational profiles (Brown et al., 1993), and voice area (Ma & Yiu, 2011; Schutte & Seidner, 1983; Titze, 1992). The x-axis demonstrates the lowest and highest frequencies the vocal folds can produce and the y-axis shows the maximum and minimum sound level (Schutte & Seidner, 1983). VRP's have a distinct and recognizable shape as shown in Figure 1. Variations in shape are expected between men and women and due to vowel choice. Typically, there is a gradual upward tilt in the upper and lower contours, indicating there may be a systematic increase in the average sound level with f_0 (Titze, 1992).

History and Determination of the Voice Range Profile

Interest in the relationship between f_0 and sound level has been documented since the 1930s (Stout, 1938; Wolf et al., 1935); however; Damsté (1970) popularized the VRP in his publication "The Phonetogram." In this article, Damsté (1970) argued that description of the laryngoscopic image could not be used to describe the properties of the voice and provided the procedures of his phonetogram to be used to evaluate voice quality. The use of the VRP was continued into the 1980's as the Union of European Phoniatricians (UEP) presented a standardized protocol to be used for the VRP (Schutte &Seidner, 1983) and Gramming and Sundberg (1988) evaluated spectrum factors that should be taken into consideration when using the VRP.

Originally, VRPs were collected with a simple sound level measuring device and tone generator as the individual held the tone for 2-3 seconds. The first computerized VRP technology was commercialized in 1983 (Printz et al., 2018) and was later published, thus providing a new method to measure VRP (Sulter et al., 1994). The computerized VRP technology has evolved over the last few decades, allowing for faster sampling and more comprehensive mapping of an individual's entire range of frequency and SPL function (Coleman, 1993).

Research and Clinical Use

Despite its exclusion from recommended basic instrumental protocols for assessment in vocal function (Patel et al., 2018), the VRP remains a compelling measure for the assessment of voice. It is considered a useful tool in the evaluation of therapy effects (Speyer, 2008), pre/post-surgical vocal function (Rendón et al., 2018; Salmen et al., 2018), as a diagnostic outcome for voice therapy and treatment (Schutte, 1980), and to track changes in the pathological conditions of the system as a whole (Coleman, 1993). Specifically, VRPs may be used to assess the vocal performance of the normal voice (Schutte, 1980), the potential of a singer's voice (Seidner et al., 1981), to track vocal changes in function during maturation, as a means to differentiate trained voices from untrained voices, and as a tool to assess the risk associated with a specific f_0 and dB(SPL) range of a particular opera role relative to a singer's voice ability.

For the VRP to be useful as a reliable, repeatable measure of voice function, the methodology must be consistently applied and efficient to avoid fatiguing the individual. Researchers and clinicians who use standardized methodology will have greater confidence when interpreting VRP outcomes. Currently, the literature describes wide variations for the procedures used for collecting VRPs (Awan, 1991; Coleman, 1993; Lycke & Siupsinskiene, 2016; Ma & Yiu, 2011; Pabon & Plomp, 1988; Schutte & Seidner, 1983; Speyer et al., 2003; Van Mersbergen et al., 1999). Variance in procedural methodology used to determine the VRP makes it difficult for clinicians and researchers to determine whether VRP outcomes described across studies or clinics occurred as a result of a medical, behavioral, or research intervention or as a result of procedural differences (Coleman, 1993).

Review of Procedural Differences

Methodological differences abound for the collection of the VRP for research and clinical application. These procedural variances contribute to the outcomes of the determination of the participant's frequency range, or the determination of the participant's sound level, or both. They are related to the wide technological resources available to collect VRPs, differing opinions of experts in the field, whether the end goal of the VRP is physiological or functional, and consistency between clinicians and researchers. Standardizing procedures for VRPs would improve reliability within and between clinicians and researchers. To advance evidence-based practice, research using VRP as an outcomes measure requires consistent methods for reliable pre and post testing. Consistent research methods within and between researchers creates opportunities for meta-analyses of data across investigations.

Procedural variance related to technological resources. Variance among VRPs may be attributed to the change from the conventional means of collecting a VRP, using a keyboard and sound decibel meter, to the automated computer program collection. These differences include the platforms and equipment used to collect the VRP (Camarrone et al., 2015; Coleman, Mabis, & Hinson, 1977; Heylen et al., 2002; Lycke et al., 2012; Ohlsson et al., 2018), the selection and specifications of the microphone used (Švec & Granqvist, 2010), the distance from the mouth to the microphone (Coleman, 1993), the graph used to demonstrate data points (Pabon & Plomb, 1988; Schutte & Seidner, 1983), and the multiple ways in which data points may be plotted and collected (Coleman, 1993; Damsté, 1970; Heylen et al., 2002; Ma & Yiu, 2011; Schutte & Seidner, 1983; Speyer et al., 2003). There have been studies which used a method initially created for the traditional measure of VRP and modified them to be appropriate measures for a computer VRP system (Awan, 1991; Schwartz, 2009). This provides evidence that the traditional method of creating a VRP and the computerized method can be aligned to have comparable methodology to create the VRP.

Procedural variance related to determination of frequency range. Many variables related to the determination of the participant's frequency range may also contribute to procedural variance. The vowel used to create a VRP has remained inconsistent across methodologies from the VRP's inception and continues to be a point of discussion. It is widely accepted practice to use a consistent vowel throughout the VRP as was recommended by the UEP (Schutte & Seidner, 1983) and Gramming and Sundberg (1988). Popular suggestions include using only /a/, which is most popular (Gramming & Sundberg, 1988; Schwartz, 2009) and using the vowels /a/, /i/, and /u/ to create multiple color-coded VRP plots in the same graph (Schutte & Seidner, 1983).

Other factors related to the determination of the participant's frequency range include whether the end goal of the VRP is physiological or functional, manipulation of registration and a singer's aversion to making "ugly sounds" (Coleman, 1993), whether range is determined using a glide or with a steady state production (Coleman, 1993; Reich et al., 1990), the duration of the target tone (Coleman, 1993; Coleman, et al., 1977; Pabon & Plomb, 1988; Pabon, 1991), whether the target tone is produced with vibrato or as a steady-state production (Coleman, 1993; Rothman et al., 1979), incorporation of warm-up (Coleman, 1993; Ma & Yiu, 2011; Heylen et al., 2002), and the time of day the VRP data is collected (Coleman, 1993; Ma & Yiu, 2011; Van Mersbergen et al., 1999).

Procedural variance related to determination of sound level. Several variables affect the determination of the participant's sound level. The number of productions elicited at each target tone varies among the literature. Some suggestions have the participant do the softest and loudest production three times or more (Coleman, 1993; Ma & Yiu, 2011) and others have the participant produce only a single production of each (Lycke & Siupsinskeine, 2016; Van Mersbergen et al., 1999). There have also been several recommendations about the scale used for the SPL. Some suggest using the A-weighted curve (dBA) (Gramming & Sundberg, 1988; Schutte & Seidner, 1983), however, Coleman (1993) recommends using the dBC scale because dBA is "biased" against frequencies of less than 200 Hz. Other studies do not state a scale used (Ohlsson et al., 2018; Rendón et al., 2018). Lastly, the degree of mouth opening and degree of breathy tone affect the determination of the participant's sound level (Coleman, 1993).

Procedural variance related to determination of frequency range and sound level. The amount of training and coaching provided by the researcher or clinician, as well as the subjective judgment as to whether what has been gathered is a representation of the participant's vocal function will also contribute to variation. In both the conventional and computerized methods, it is important to explain the equipment and what is expected to the participant before they begin the trial. Many studies using computer automated systems still have a researcher coach and train the participants with the unfamiliar task of collecting the VRP (Lycke & Siupsinskiene, 2016; Ma & Yiu, 2011; Speyer et al., 2003), but because the computer systems do not require a clinician to be present, there have been studies that do not have someone present to provide coaching during collection of the VRP (Van Mersbergen et al., 1999).

Procedural variables that require consideration. The room acoustics of the recording environment has been discussed at length. Ma and Yiu (2011) state that poor acoustics can affect the validity of the sound levels measured. The UEP (Schutte & Seidner, 1983) recommends using a room with living room acoustics as optimal, but Coleman (1993) argues that the size and absorption characteristic of the room environment do not limit recording. Both Coleman (1993) and Ma and Yiu (2011) recommend the environment should be quiet, with 40dB SPL or less ambient noise. The best way to control this environment is to record in a sound treated booth (Coleman, 1993; Lycke & Siupsinskiene, 2016; Van Mersbergen et al., 1999) or to monitor the environment with a sound level meter as the recording is taking place (Schwartz, 2009).

Justification and Hypotheses

A systematic comparison of the methods affecting the determination of the most representative profile of the voice should be conducted for three primary reasons: 1) the methodology for collection of the VRP varies; 2) it is a valuable clinical and research assessment of vocal physiology and function; and 3) it serves as a vocal function outcome measure that is useful after both surgical and behavioral intervention. This review focused on outcomes based literature including only studies that measure pre/post VRP outcomes in therapy or treatment intervention and studies comparing VRPs across participants. Because a VRP for research purposes and a VRP for clinical measures can vary, this allowed similar studies to be compared providing useful information for speech pathologists in clinical practice. Determination of the participant's range might be affected by one or more of the following; the vowel used to create the VRP, the way in which the initial range was established, the amount of time each pitch was held, sung or unsung sustained phonation, use of vibrato or steady-state production, and the use of warm-up. Inclusion or exclusion of coaching from the clinician varies widely across studies. Coaching influences identification of the true extremities of the range, but also plays a factor in determining whether a sound is the loudest or softest production. The loudest and softest productions might also be affected by the following; the number of trials used to attain the SL, the equipment used, a scale of dBA or dBC, and the microphone to mouth distance. We hypothesized the following in regards to variables chosen for review:

- /a/ will be the most common vowel used;
- A glide will be the most common way the range is established;
- Pitches will be held for one to two seconds;
- Phonations will be unsung and steady-state;
- Warm-up will not be included as a part of the collection procedure;
- There will be little coaching by the clinician reported;
- There will be variance in the numbers of trials used to attain the sound level;
- There will be variance in the equipment used;
- dBA will be used more often than dBC; and,
- There will be variance in the distance from the microphone to the mouth.

To the author's knowledge, one other systematic review has been published regarding the VRP evaluating the reproducibility of the automated VRP (Printz et al., 2018).

Methods

Protocol and Data Management

A systematic review was conducted using the Preferred Reporting Items for Systematic Review and Meta-Analysis checklist (PRISMA) (Moher, Liberati, Tetzlaff, & Altman, 2009). The systematic literature search used PubMed, Web of Science, Psych Info, ProQuest Dissertations and Theses Global, and Google Scholar. The search in Google Scholar was limited to the first 100 citations with a filter for the English language and elimination of published patents. The searches in the four databases were limited to only the English language, but no other filters, including time constraints, were used to ensure the majority of relevant studies using the conventional and automated means of collecting the VRP were included. The electronic search strategy that was used for these databases is provided in Table 1. The search results from the electronic databases were imported into Endnote X6 (Thomson Reuters, New York, NY). Following the full search of the databases, duplicates were removed.

Table 1. Search Strategy

Database	**Search Terms**
PubMed, Web of Science, Psych Info, and ProQuest Dissertations and Theses Global	phonetogram OR phonetography OR “voice profile” OR “voice profiles” OR “phonational profiles” OR “voice area” OR “voice areas” OR “voice range profile” OR “voice range profiles” OR “vocal range” OR “vocal ranges” OR “voice frequency range”
Google Scholar	phonetogram | phonetography | voice profile | phonational profile | voice area | voice range profile | vocal ranges | voice frequency range

Study Selection

The titles and abstracts of studies were be screened for inclusion by two independent raters (GC and MS) using predefined inclusion and exclusion criteria. Exclusion criteria included:

- The study did not use VRP as an outcome measure; and/or
- The study was a paper based in theory and with no clinical or research outcomes; and/or
- Children 18 years and younger were included in the participant pool; and/or
- The methods were not described in enough detail (i.e., meeting abstract); and/or
- Inability to gain access; and/or
- Articles written in foreign languages.

Inclusion criteria met both of the following requirements:

- The study was written in English; and
- The study used the VRP as an outcomes measure;

In addition, each study had one of the following:

- vowel used [which vowel(s)]; and/or,
- how the initial range was established (glide or stair step); and/or,
- amount of time each pitch was held (number of seconds); and/or
- sung or unsung target (S/U); and/or
- vibrato or steady-state production (V/S); and/or
- inclusion of warm-up (yes/no); and/or
- coaching provided by the researcher (yes/no); and/or,
- the number of trials used to establish extreme SL (number used); and/or,
- the equipment used; and/or
- scale used (dBA or dBC); and/or
- the microphone to mouth distance

If an article did not meet exclusion criteria at the title and abstract level or if the raters disagreed on whether exclusion criteria was met at this level, then the study was included in the evaluation of full texts. The full texts were reviewed by two independent raters (GC and MS) with a third rater (LP) available to determine the inclusion of an article if a disagreement arose between the two raters. The final included studies were based on the agreement of the three reviewers with the use of a spreadsheet for subsequent tallying of findings. The reference sections of all articles were reviewed by the primary reader to determine if there are remaining articles that should be added. Figure 2 represents an overview of the study selection process.

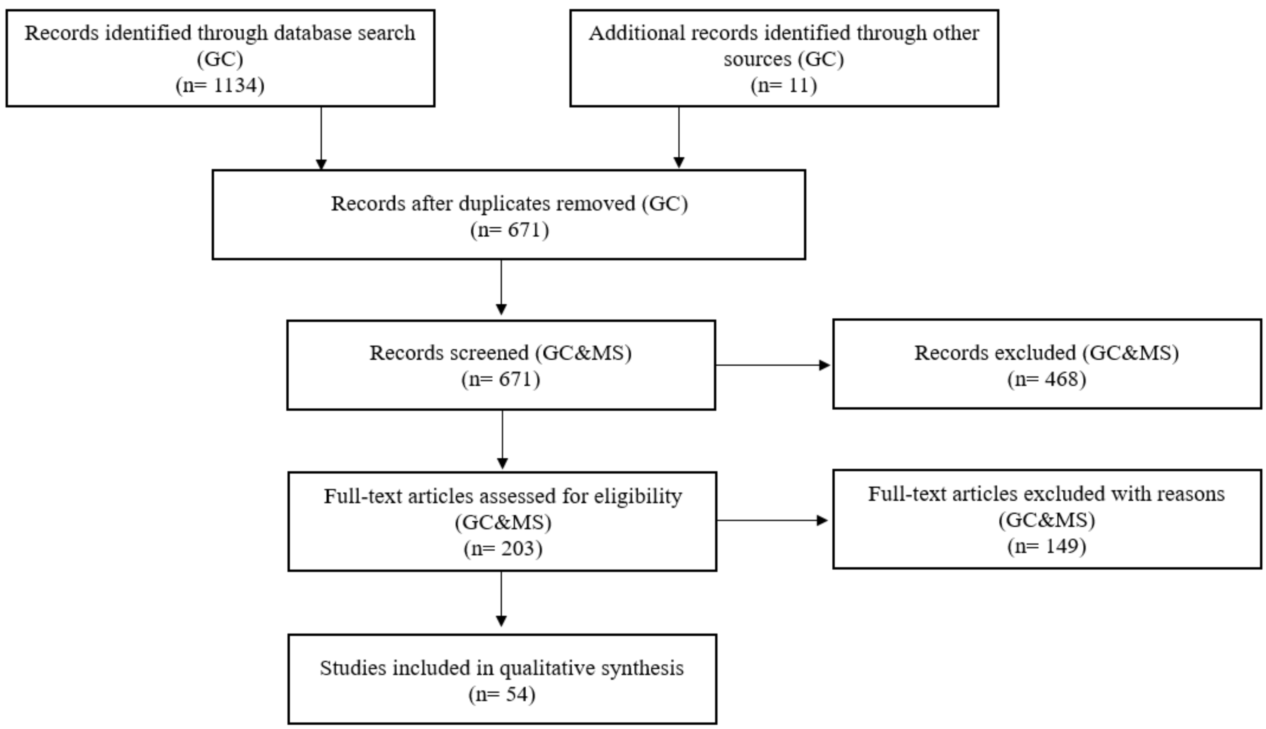

Figure 2. References included in the search and selection process. Flow diagram adapted from PRISMA (Moher et al., 2009) .

Data Extraction

Data was extracted by an independent author (GC) from each included study. A table was created with the descriptive data regarding the vowels used, how initial range was established, the amount of time each pitch was held, sung or unsung target, vibrato or steady-state production, whether warm-up was reported, whether coaching was provided by the researcher, the number of trials used to establish the SL, the equipment used, the scale used, and the microphone to mouth distance. This table also included columns determining the level of evidence for each study, the participant's sex, participant's age range, whether the study was performed on singers or non-singers, why the study used VRP, whether VRP was used as a functional or physiological outcome measure, and which acoustic measures were used for analysis. Data extraction was repeated by an independent author (MS) for 10% of the included articles to establish inter-rater reliability.

Risk of Bias

Use of the Critical Appraisal Skills Programme ("CASP checklists," 2014) checklist for each individual study was deferred due to this investigation's interest in the data collection methods and not the participant selection (other than age) or the results of any intervention. To circumvent publication bias, Google Scholar and ProQuest Dissertations and Theses Global were used for the review search.

Results

Search and Selection

The database search and reference list review resulted in 1145 titles and abstracts (Figure 2); 463 were duplicates. Titles and abstracts of 671 articles were screened, with 203 selected for full-text review. Using the exclusion and inclusion criteria listed, 54 articles were considered eligible for inclusion in the systematic review. Characteristics of included studies are listed in Table 2.

Table 2. Demographics of articles included in systematic review

Author(s)	Year	Title	CASP Identification
Ahlander, V.L., Rydell, R., & Löfqvist, A.	2012	How do teachers with self-reported voice problems differ from their peers with self-reported voice health?	Randomized Controlled Trial
Åkerlund, L.	1993	Averages of sound pressure levels and mean fundamental frequencies of speech in relation to phonetograms: Comparison of nonorganic dysphonia patients before and after therapy	Cohort Study
Åkerlund, L., Gramming, P., & Sundberg, J.	1992	Phonetogram and averages of sound pressure levels and fundamental frequencies of speech: Comparison between female singers and nonsingers	Case Control
Anderson, J.A.	1999	Vocal function in subjects with compensated unilateral vocal fold paralysis pre and post medialization thyroplasty	Cohort Study
Barrier, J.T.	1993	The development of criteria for the selection of age-appropriate literature for the senescent voice	Systematic Review
Camarrone, F., Ivanova, A., Decoster, W., De Jong, F., & Van Hulle, M.M.	2015	Stable voice clusters identified when using the maximum versus minimum intensity curve in the phonetogram	Cohort Study
Chen, S.H.	1996	Voice range profile of Taiwanese normal young adults: A preliminary study	Cohort Study
Chen, S.H.	2007	Sex differences in frequency and intensity in reading and voice range profiles for Taiwanese adult speakers	Cohort Study
Chen, S.H.	2008	Voice range profiles for tonal dialect of Min	Cohort Study
Coleman, R.F., Mabis, J.H., & Hinson, J.K.	1977	Fundamental frequency-sound pressure level profiles of adult male and female voices	Cohort Study
Gramming, P., & Åkerlund, L.	1988	Non-organic dysphonia II. Phonetograms for normal and pathological voices	Case Control
Gramming, P. & Sundberg, J.	1988	Spectrum factors relevant to phonetogram measurement	Cohort Study
Gramming, G., Sundberg, J., & Åkerlund, L.	1991	Variability of phonetograms	Case Control
Hallin, A.E., Fröst, K., Holmberg, E.B., & Södersten, M.	2012	Voice and speech range profiles and Voice Handicap Index for males – methodological issues and data	Case Control
Heylen, L, Wuyts, F.L., Mertens, F., De Bodt, M., & Van de Heyning, P.H.	2002	Normative voice range profiles of male and female professional voice users	Cohort Study
Holmberg, E.B., Ihre, E., & Södersten, M.	2007	Phonetograms as a tool in the voice clinic: Changes across voice therapy for patients with vocal fatigue	Case Control

Holmes-Bendixen, A.R.	2013	The influence of whistle register phonation exercises in conditioning the second passaggio of the female singing voice	Cohort Study
Hunter, E.J., Švec. J.G., & Titze, I.R.	2006	Comparison of the produced and perceived voice range profiles in untrained and trained classical singers	Case Control
Hunter, E.J., & Titze, I.R.	2005	Overlap of hearing and voice ranges in singing	Case Control
Johansson et. al.	2018	Assessment of voice, speech, and communication changes associated with cervical spinal cord injury	Case Control
Keilmann et al.	2010	Long-term functional outcome after unilateral cordectomy	Cohort Study
Kelly, V., Hertegård, S., Eriksson, J., Nygren, U., & Södersten, M.	2018	Effects of gender-confirming pitch-raising surgery in transgender women a long-term follow-up study of acoustic and patient-reported data	Cohort Study
Kolker, A.	2017	Practical elicitation methods for the voice range profile	Cohort Study
Lamarche, A., Ternström, S., & Pabon, P.	2010	The singer's voice range profile: Female professional opera soloists	Cohort Study
Lamesch, S., Doval, B., & Castellengo, M.	2012	Toward a more informative voice range profile: The role of laryngeal vibratory mechanisms on vowels dynamic range	Cohort Study
LeBorgne, W.D., & Weinrich, B.D.	2002	Phonetogram changes for trained singers over a nine-month period of vocal training	Cohort Study
Lee, H.Y., Lee, J., Dionigi, G., Bae, J.W., & Kim, H.Y.	2015	The efficacy of intraoperative neuromonitoring during robotic thyroidectomy: A prospective, randomized case-control evaluation	Randomized Controlled Trial
Ma et. al.	2007	Reliability of speaking and maximum voice range measures in screening for dysphonia	Case Control
Ma, E., & Yiu, E.	2006	Multiparametric evaluation of dysphonic severity	Case Control
Mailänder, E., Mühre, & Barsties, B.	2017	Lax Vox as a voice training program for teachers: A pilot study	Cohort Study
Marunick, M.T., & Menaldi, C.J.	2000	Maxillary dental arch form related to voice classification: A pilot study	Cohort Study
Neuschaefer-Rube, C., Šram, F., & Klajman, S.	1997	Three-dimensional phonetographic assessment of voice performance in professional and non-professional speakers	Case Control
Ohlsson et. al.	2018	Voice therapy outcome—a randomized clinical trial comparing individual voice therapy, therapy in group, and controls without therapy	Randomized Controlled Trial
Park, M.W., Baek, S., Park, E., & Jung, K	2018	Long-term voice outcome after thyroidectomy using energy based devices	Cohort Study
Pei et al.	2018	Voice range change after injection laryngoplasty for unilateral vocal fold paralysis	Cohort Study
Printz, T., Sorenson, J.R., Godballe, C., & Grentved, Å.M.	2018	Test-retest reliability of the dual-microphone voice range profile	Cohort Study

Ray, C., Trudeau, M.D., & McCoy, S.	2018	Effects of respiratory muscle strength training in classically trained singers	Cohort Study
Schneider, B., Bigenzahn, W., End, A., Denk, D., & Klepetko, W.	2003	External vocal fold medialization in patients with recurrent nerve paralysis following cardiothoracic surgery	Cohort Study
Schneider, B., Denk, D., & Bigensahn, W.	2003	Functional results after external vocal fold medialization thyroplasty with the titanium vocal fold medialization implant	Cohort Study
Schneider-Stickler, B., Knell, C., Aichstill, B., & Jocher, W.	2010	Biofeedback on voice use in call center agents in order to prevent occupational voice disorders	Case Control
Schönweiler, R., Wohlfarth, K., Dengler, R., & Ptok, M.	1998	Supraglottal injection of botulinum toxin type A in adductor type spasmodic dysphonia with both intrinsic and extrinsic hyperfunction	Cohort Study
Sihvo, M., Laippala, P., & Sala, E.	2000	A study of repeated measures of softest and loudest phonations	Cohort Study
Sihvo, M., & Sala, E.	1996	Sound level variation findings for pianissimo and fortissimo phonations in repeated measurements	Cohort Study
Šiupšinskienė, N., Adamonis, K., & Toohill, R.J.	2009	Usefulness of assessment of voice capabilities in female patients with reflux-related dysphonia	Case Control
Storck, C., Brockmann, M., Scnellmann, E., Stoecklie, S.J., & Schmid, S.	2007	Functional outcome of vocal fold medialization thyroplasty with a hydroxyapatite implant	Cohort Study
Tae et. al.	2012	Functional voice and swallowing outcomes after robotic thyroidectomy by a gasless unilateral axillo-breast approach: Comparison with open thyroidectomy	Cohort Study
Teles-Magalhães, L.C., Pegoraro-Krook, M.I.m & Pegoraro, R.	2000	Study of elderly females' voice by phonetography	Cohort Study
Titze., I.R., Wong, D., Milder, M.A., Hensley, S.R., & Ramig, L.O.	1995	Comparison between clinician-assisted and fully automated procedures for obtaining a voice range profile	Cohort Study
Tuomi et. al.	2017	Voice range profile and health-related quality of life measurements following voice rehabilitation after radiotherapy; a randomized controlled study	Randomized Controlled Trial
Van Gogh et. al.	2005	The efficacy of voice therapy in patients after treatment for early glottis carcinoma	Randomized Controlled Trial
Verdonck-de Leeuw, I.M., & Mahieu, H.F.	2004	Vocal aging and the impact on daily life: A longitudinal study	Cohort Study

Wingate, J.M., Brown, W.S., Shrivastav, R., Davenport, P., Sapienza, C.M.	2007	Treatment outcomes for professional voice users	Cohort Study
Yiu et. al.	2006	A randomized treatment-placebo study of the effectiveness of acupuncture for benign vocal pathologies	Randomized Controlled Trial
Yiu, E.M, & Chan, R.M.	2003	Effect of hydration and vocal rest on the vocal fatigue in amateur karaoke singers	Randomized Controlled Trial

Data Extraction

Data was extracted by an independent author (GC) and 10% of data was assessed for reliability by a second author (MS). The average interrater agreement across the study variables was 91%. The following data points were collected: the vowels used, how initial range was established, the amount of time each pitch was held, sung or unsung target, vibrato or steady-state production, whether warm-up was reported, whether coaching was provided by the researcher, the number of trials used to establish the SL, the equipment used, the scale used, and the microphone to mouth distance. Because the equipment used for each study was highly variable and not often described in detail, this data point was not included in table form. However, the acoustic parameters used for analysis were determined to be important for method variability and were added as a data point. Table 3 outlines the data points found in individual studies.

Table 3. Data extraction of articles included in systematic review

Author	Vowel	Range (Glide/Stepwise)	Seconds Pitch Held	Sung/Unsung	Vibrato	Warm-Up	Coaching	# of Trials for SL	SL Scale	Mic to Mouth Distance	Acoustic Parameters
Ahlander et. al. (2012)	/a/	Glide for whole VRP	NA	NR	NR	NR	Y	NR	NR	7 cm (corrected to 30 cm)	max. f_0, min. f_0, max. SPL, min. SPL, max. area
Åkerlund (1993)	/a/	NR	2 sec.	NR	NR	NR	NR	NR	NR	30 cm	f_0 range, max. & min SPL curves
Åkerlund et al. (1992)	/a/	Preset pitches, triads, and glides	2 sec. sustained; 1 sec. triad/glide	NR	NR	NR	NR	NR	NR	30 cm	f_0 range, max. & min. SPL curves
Anderson (1999)	/a/	Glide	3 sec.	NR	NR	Y	Y	3	dBC	30 cm (SLM)	max. area, f_0 range
Barrier (1993)	/a/	Stepwise	NR	Sung	NR	NR	Y	NR	NR	NR	max. area
Camarrone et al. (2015)	NR	NR	NR	Unsung	NR	NR	NR	NR	dBA	NR	max. & min SPL curves
Chen (1996)	/a/	stepwise	2 sec.	NR	NR	NR	Y	NR	NR	1 cm	max. f_0, min. f_0, f_0 range
Chen (2007)	/a/	stepwise	2 sec.	NR	NR	NR	Y	Several	dBA	1 cm	min. f_0, max. f_0, f_0 range, SPL min., SPL max, SPL range
Chen (2008)	/a/	stepwise	2 sec.	NR	NR	NR	Y	Min. 3	dBA	1 cm	min. f_0, max. f_0, f_0 range, SPL min., SPL max., SPL range
Coleman et al. (1977)	NR	Both	2 sec.	NR	NR	NR	NR	1 each	NR	6 in	F_0 range, SPL range, SPL

											min., SPL max.
Gramming & Åkerlund (1988)	/a/	Preset pitches	2 sec.	NR	NR	NR	NR	NR	linear	30 cm	SPL min., SPL max.
Gramming & Sundberg (1988)	/a/; /i/, /a/, /u/; /i/, /a/, /u//e/	Preset pitches	2 sec.	NR	NR	NR	NR	NR	dBA & dBC	30 cm	study specific measures
Gramming et al. (1991)	/a/	Preset pitches	2 sec.	NR	NR	NR	NR	NR	flat	30 cm	study specific measures
Hallin et al. (2012)	/a/	Unclear	NR	NR	NR	NR	Y	As many as needed	linear	15 cm; corrected to 30 cm	max.f_0, min. f_0, f_0 range, min. SPL, max. SPL, SPL range; max. area
Heylen et al. (2002)	/a/	Unclear	Unclear	NR	NR	Y	NR	NR	dBA	NR	max. & min SPL curves
Holmberg et al. (2007)	/a/	Glides for whole VRP	NA	NR	NR	NR	Y	several	linear	8 cm; corrected to 30 cm	max. SPL curve, max area
Holmes-Bendixen (2013)	/a/	stepwise	2 sec.	NR	NR	NR	NR	NR	dBC	NR (to SLM)	SPL range, f_0 range
Hunter et al. (2006).	/i/, /a/, /u/	stepwise	1.5 sec.	Sung	NR	NR	NR	NR	Linear; converted to dB & dBA	1 m; corrected to 30 cm	study specific measures
Hunter & Titze (2005)	/i/, /a/, /u/	stepwise	1.5 sec.	Sung	NR	NR	NR	unclear	Linear; converted to dB	1 m; corrected to 30 cm	study specific measures
Johansson et al. (2018)	/a/	Glides for whole VRP	NA	NR	NR	NR	Y	As many as needed	NR	15 cm	max. area, min. SPL, max. SPL

Keilmann et al. (2010)	NR	both	NR	NR	NR	NR	NR	NR	NR	NR	max. SPL, SPL range, f_0 range
Kelly et al. (2018)	/a/	both	NR	NR	NR	NR	Y	As many as needed	NR	15 cm; corrected to 30 cm	min. f_0, max. f_0, f_0 range
Kolker (2017)	/a/	Stepwise	2 sec.	NR	NR	Y	Y	Min. of 2 & max. of 6	dBA	30 cm	f_0 range, SPL range, max. & min SPL curves
Lamarche et al. (2010)	/a/	Glide; preset pitches; stepwise (triad)	NA; 2 sec.; NA	NR; Sung; Sung	NR; V; V	Y	Y	As many as needed; 1 messa di voce; 1 pp, mf, ff	linear	30 cm	min. f_0, max. f_0, f_0 range; min. SPL, max. SPL
Lamesch et al. (2012)	/i/, /a/, /o/	No range	NR	NR	NR	Y	Y	Unclear	NR	30 cm	min. SPL, max. SPL
LeBorgne & Weinrich (2002)	/a/	Glide; sustained 3 xs	3 sec.	Either	Either	NR	NR	1 for 3 sec.	NR	15 cm	f_0 range; min. SPL, max. SPL
Lee et al. (2015)	NR	glide	NR	NR	NR	NR	NR	NR	NR	NR	min. f_0, max. f_0, min. SPL, max. SPL
Ma et al. (2007)	/a/	stepwise	NR	NR	NR	Y	Y	3	dBA	5 cm; corrected to 30 cm	max. f_0, min. f_0, max. SPL, min. SPL
Ma & Yiu (2006)	/a/	NR	NR	NR	NR	NR	NR	NR	dBA	NR	min. f_0, max.. f_0, min. SPL, max. SPL, f_0 range; SPL range; max. area
Mailänder et al. (2017)	/a/	stepwise	2 sec.	NR	NR	NR	NR	NR	dBA	NR	min. f_0, max.. f_0, min. SPL,

											max. SPL, f_0 range; SPL range
Marunick et al. (2000)	/a/	both	NR	NR	NR	NR	NR	NR	NR	10 cm	f_0 range
Neuschaefer-Rube et al. (1997)	/a/	stepwise	As long as possible	NR	NR	NR	NR	NR	dBA	30 cm	min. f_0, max.. f_0, min. SPL, max. SPL
Ohlsson et al. (2018)	/a/	Glide for whole VRP	NA	NR	NR	NR	NR	As many as needed	NR	NR	max. area
Park et al. (2018)	/a/	NR	NR	NR	NR	NR	NR	NR	NR	NR	max. area
Pei et al. (2018)	/a/	stepwise	NR	NR	NR	NR	NR	NR	NR	NR	min. f_0, max.. f_0, min. SPL, max. SPL, f_0 range, SPL range, f_0 of max. SPL, f_0 of min. SPL
Printz et al. (2018)	NR	Both for whole VRP	NR	NR	NR	NR	NR	As many as needed	NR	2-3 cm; 30 cm	min. f_0, max.. f_0, min. SPL, max. SPL, f_0 range, SPL range, f_0, max. area
Ray (2018)	NA	stepwise	3-5 sec.	NR	NR	NR	NR	NR	NR	NA	f_0 range, SPL range
Schneider, Bigenzahn et al. (2003)	NR	Preset pitches	NR	NR	NR	NR	NR	NR	NR	NR	f_0 range
Schneider, Denk, & Bigensahn (2003)	NR	NR	NR	NR	NR	NR	NR	NR	dBA	NR	f_0 range

Schneider-Stickler et al. (2010)	NR	stepwise	NR	NR	NR	NR	NR	NR	NR	NR	f_0 range, min. f_0, max.. f_0
Schönweiler et al. (1998)	NR	NR	NR	Sung	NR	NR	NR	NR	NR	NR	min. f_0, max.. f_0, max. SPL, min. SPL
Sihvo et al. (2000)	/a/	Preset pitches	3 sec.	NR	NR	NR	NR	10	dBA	30 cm	min. SPL, max SPL
Sihvo et al. (1996)	/a/	Preset pitches	2 sec.	NR	NR	NR	NR	10	dBA	NR	min. SPL, max. SPL
Šiupšinskienė et al. (2009)	/a/	NR	2 sec.	NR	NR	NR	NR	2 or more within 3 dB	dBA	NR	max. f_0, f_0 range, min. SPL , SPL range; max. area
Storck et al. (2007)	/o/	NR	NR	NR	NR	NR	NR	NR	NR	NR	SPL range; f_0 range
Tae et al. (2012)	/a/	Glide for whole VRP	NA	NR	NR	NR	NR	NR	NR	15 cm	min. f_0, max.. f_0, f_0 range, SPL range
Teles-Magalhães et al. (2000)	/a/	Preset pitches	5 sec.	NR	NR	NR	NR	1 each	dBA	30 cm	min. f_0, max.. f_0, f_0 range, min. SPL, max. SPL, SPL range, max. area
Titze et al. (1995)	/a/	Stepwise; NR	NR; unclear	Unsung; NR	NR	NR	Y; N	3; as many as needed	Linear; NR	8 cm	max. area
Tuomi et al. (2017)	/a/	NR	NR	NR	NR	NR	NR	NR	NR	12 cm	max. area, min. f_0, max.. f_0, max. SPL, min. SPL
Van Gogh et al. (2005)	/a/	Glide	NR	NR	NR	NR	NR	NR	NR	NR	f_0 range, SPL range

Verdonck-de Leeuw & Mahieu (2004)	/a/	NR	NR	NR	NR	NR	NR	NR	NR	NR	f_0 range, SPL range
Wingate et al. (2007)	/a/	Glides for whole VRP	NR	NR	NR	NR	NR	Multiple times	NR	6 in	max. area, min SPL, max SPL, min. f_0, max. f_0
Yiu et al. (2006)	/a/	stepwise	NR	NR	NR	NR	NR	NR	dBA	5 cm; corrected to 30 cm	max. f_0, f_0 range, max. SPL; SPL range; max. area
Yiu & Chan (2003)	/a/	stepwise	As long as needed	NR	NR	NR	NR	NR	NR	10 cm	min. f_0, max.. f_0, f_0 range, min. SPL, max. SPL, SPL range, max. area

Note: NR represents data points that were not reported and NA represents data points that were not applicable based on the study.

Synthesis of Results

Data related to determination of frequency range. The most common vowel used for the VRP was /a/, which was used in 74% of included studies. However, studies varied in which vowel or vowel combinations were implemented for data collection. The way in which the range was established did not have a most commonly used method. The glide, pre-established pitches without prior range determination, and stepwise methodology together made up 69% of the methods in included studies. The time (s) pitches were held ranged from 1 second to 10 seconds. Twenty-eight percent of studies held the pitch for 2 seconds; however, 46% of studies did not report how long the pitch was held. Few studies reported whether a pitch was sung or unsung, whether participants used vibrato or a steady-state pitch, and if warm-up was a part of the data collection process. Table 4 describes the findings of data related to the determination of the participant's frequency range.

Table 4. Combined data related to determination of the participant's frequency range

Data Point	Percent of studies that used measure
Vowel Used	
/a/	74%
/o/	2%
/i/, /a/, /o/	2%
/i/, /a/, /u/	6%
/i/, /a/, /u//e/	2%
Not Reported	19%
Range Established	
Pre-established pitches (no range)	17%
Glide (includes VRP done on glide)	19%
Stepwise	33%
Glide with sustained outer limits	2%
Both/participant choice	9%
No range	2%
Not Reported/Unclear	24%
Seconds Held	
1 second	2%
1.5 seconds	4%
2 seconds	28%
3 seconds	6%
3 – 5 seconds	2%
5 seconds	2%
As long as needed/possible	4%
Not Reported/Unclear	46%
Not Applicable (glide VRPs)	13%
Sung or Unsung	
Sung	11%
Unsung	4%
Either	2%
Not Reported	89%
Vibrato or Steady-state	
Vibrato	4%
Either	1%
Not Reported	98%
Warm-Up	
Reported	11%
Not Reported	89%

Note: the variables were divided by the number of studies. Because some studies used multiple for comparison, the percentages don't add to 100%

Data related to the determination of sound level. The equipment used likely affects the determination of the participant's sound level. However, due to the number of platforms used, and the high variability of equipment, these data points were not compiled. The number of trials used for the sound level varied in number, but also in the way it was described. A study might have 1 trial for soft and 1 trial for loud or they may have the client do 1 trial for pianissimo, 1 trial for mezzo forte, and 1 trial for fortissimo. The sound level scale used was not reported for 56% of studies. The most commonly reported scale was dBA. The microphone distance from the mouth ranged from 1 cm to 30 cm or a distance corrected to 30 cm. The commonly found data point was the use of 30 cm or signal corrected to 30cm. Table 5 describes combined data related to the determination of the participant's sound level.

Data related to determination of frequency range and sound level. Coaching was reported in 30% of studies. Coaching was occasionally used as terminology in the literature, but coaching was often described as an explanation of the procedure, training, or help with motions and gestures during the procedure. Table 5 includes the findings related to coaching as a data point.

Table 5. Combined data related to the determination of the participant's sound level

Data Point	Percent of studies that used measure
Coaching*	
Reported	30%
Not Reported	70%
Number of SL Trials	
1 for each	4%
1 messa di voce	2%
1 for pp, mf, ff	2%
2 or more within 3dBA	2%
Min. of 2 and max of 6	2%
3 for each (soft/loud)	7%
Min. of 3	2%
10	4%
Several	6%
Not Reported/Unclear	63%
Sound Level Scale	
dBC	4%
dBA	28%
Linear	13%
Flat	2%
Both	2%
Not Reported	56%
Mic Distance	
1 cm	6%
2-3 cm	2%
8 cm	2%
10 cm	4%
12 cm	2%
15 cm	4%
6 in	5%
30 cm	24%
Corrected to 30 cm	15%
Not Reported	35%
Not Applicable	2%

Note: the variables were divided by the number of studies. Because some studies used multiple for comparison, the percentages don't add to 100%.
*Coaching affects both fundamental frequency and sound level, but was only reported in this table.

Acoustic outcome measures. Often studies used measures related to their specific participants and hypotheses. However, eight measures were described in over 10% of studies and used as a part of analysis in the literature. These included: f_0 range, maximum f_0, minimum f_0, SPL range, maximum SPL, minimum SPL, maximum area, and the maximum and minimum SPL curves. Table 6 shows the percentage of included studies that used these measures for analysis.

Table 6. Combined data related to acoustic measure

Measures Reported	Percent of studies that used measure
f_0 Range	56%
Max. f_0	44%
Min. f_0	41%
SPL Range	37%
Max. SPL	48%
Min. SPL	46%
Max. Area	44%
Max. and Min. SPL curves	11%

Note: the variables were divided by the number of studies. Because some studies used multiple for comparison, the percentages don't add to 100%. Additional study-specific acoustic measures were not included in this table.

Discussion

The purpose of this systematic review was to determine the degree of variability for data collection practices for the VRP. The findings support the hypothesis that variability exists in data collection methods for the VRP. However, the degree to which data collection methods varied and the high percentage of studies that did not report any of the inclusion variables was unexpected. Eight of the 10 data points queried were not reported in 35% or more of the included studies. These variables included several basic protocol points such as the duration of the sustained pitch, the number or trials used to determine sound level, and the microphone to mouth distance. The following hypotheses about individual data points were supported by the literature: (1) /a/ was the most common vowel used; (2) warm-up was not reported as a part of the collection procedure; (3) coaching by the clinician was not reported in the majority of the papers reviewed; (4) there was variance in the numbers of trials used to attain the sound level; (5) there was variance in the equipment used; (6) dBA was used more often than dBC; and, (7) there was variance in the distance from the microphone to the mouth across studies. The following hypotheses were not supported or could not be determined: (1) a glide was the most common way the range was established; (2) pitches were held for one to two seconds; and, (3) phonations was unsung and steady-state.

Methodology of the VRP

Variables related to task elicitation. While the UEP (Schutte & Seidner, 1983) recommended using /i/, /a/, and /u/ in overlapping graphs, the vowel most commonly used was /a/. This is likely because Gramming and Sundberg (1988) recommended using only /a/ to prevent formant interactions and suggested that the vowels /i/ and /u/ do not offer additional acoustic information. This is relevant when one considers most classical singers have vowels they prefer to use when singing through a particular region of their own registration. These vowel preferences differ across individuals for the same pitch class promoting the assumption that a singer may be able to sing louder or softer on their "favorite vowel" as opposed to a different vowel.

When using the VRP as a targeted outcome of therapy or medical intervention, it is important to document the patient's broadest and most dynamic vocal output. If only a single vowel is used, then the client may be limited and the VRP may not capture the degree to which they can produce higher or lower and softer or louder productions. Further, when obtaining a

VRP for a client with a disordered voice, the extremes of their range and sound level will likely be considerably affected. Using the vowel with which they are most comfortable could provide information about the voice production that would not otherwise be found with a single, predetermined vowel, resulting in a more ecologically valid assessment of vocal function. It is acknowledged that use of a predetermined vowel may be optimal for comparison of pre/post data; however, the limitation of a predetermined vowel should be understood.

The use of a glissandi, or glide, or steady state production for range determination was not accounted for in most of the studies evaluated. This may be due to conflicting information about which of these yields the largest semitone range. In a study performed by Reich et al. (1990), glissandi most consistently yielded a significantly larger semitone range by about 2 to 3 semitones, but a more recent study reported discrete half steps yielded a larger semitone range than a glissandi (Barrett, Lam, & Yiu, 2018). If a participant yields a larger semitone range through a glide, then a problem arises if the procedure requires the participant to produce sustained productions when facilitating the reading of the sound level meter.

The variability of how the range was established fell into two categories: (1) establishing the range prior to the collection of the VRP data points and (2) the data points themselves serving as the range of the participant. If range was established prior to the collection of the VRP data, then a glide was used to get the outer limits and these outer limits were used to determine that data points that would be collected for the VRP (Anderson, 1999; LeBorgne & Weinrich, 2002; Mailander, Muhre, & Barsties, 2017). There were several methodologies that did not have a procedure for collecting the range before data was taken. These included a full VRP using every pitch in the participant's range (Chen, 2008; Holmes-Benedixen, 2013; Hunter, Švec, & Titze, 2006, Kolker, 2017, Ma, Robertson, Radford, Vagne, El-Halabi, & Yiu, 2007; Yiu et al.,

2006) , glides for a set amount of time to collect the whole VRP (Ahlander, Rydell, & Lofqvist, 2012; Holmberg, Ihre, & Södersten, 2007; Johansson et al., 2018; Ohlsson et al., 2018), and the use of a modified VRP with pre-established pitches (Gramming & Akerlund, 1988; Marunick, & Menaldi, 2000; Teles-Magalhães, Pegoraro-Krook, & Pegoraro, 2000). For those that used pre-established pitches, it is difficult to determine whether the outermost frequencies were gathered since it may not be in the pre-determined pitch class, unless the study otherwise stated they confirmed the range using additional measures.

The initial data extraction included determination of whether the study collected the functional versus the physiological range in the VRP. However, this was difficult to ascertain for each study due to limited details in the methodological description. If the physiological range was the end goal, then every sound the participant could produce would be considered a data point. This means a "squeak" at the uppermost outer limit of the range would be included, even if the participant could not hold out the frequency for a second or more. A functional VRP would require only the frequencies that could be produced for a full second or more in the data points. Future studies should state this explicitly as it was difficult to determine descriptively for each study.

The number of seconds the target frequency was sustained was not always applicable. For instance, in several studies (Ahlander et al., 2012; Holmberg et al., 2007; Johansson et al., 2018; Ohlsson et al., 2018) where a glide was used to obtain all VRP, occurring in 11% of included studies, no target frequencies were sustained during data collection. In studies using sustained frequencies for data collection, 46% did not report the duration. This can largely be accounted for by the number of studies using computer-based programs like Phog (Hitech Development AB, Täby, Sweden), Voice Profiler (Alphatron Medical Systems, Rotterdam, The Netherlands), Voice Range Profile (CSL, PENTAX Medical, Tokyo, Japan) or other various platforms. These programs detect the frequency in milliseconds, allowing the protocol to easily omit or overlook this component. However, this step is important for reduplication and to ascertain that the SPL recorded is accurate. The length of time the pitch is held also influences whether the derived VRP is functional or physiological. If the pitch is sustained for a longer period of time, then the VRP would likely be more constricted yielding a functional VRP and if the pitch is sustained for a shorter period of time then the VRP be more in line with the physiological VRP.

The range for the duration of the sustained frequency varied from 1 second to 10 seconds. This is a considerable difference when one considers that voice disordered clients would not likely be able to hold it out for more than a few seconds. Erring of the side of a shorter duration would be more likely to capture as much of the frequency range and intensity variations as possible. Further, if the client is not a singer, they do not require sustained vowels for longer than a fraction of a second in connected speech, so there is no ecologically valid reason to sustain it for over 1 to 2 seconds (Smith & Sandage, 2017).

The number of trials required to determine the sound level was largely unreported or unclear in 63% of the included references, a subgroup of which (11%) used a glide for a pre-determined amount of time to determine the whole VRP. However, a large percentage of studies remained that used sustained target frequencies with no number of trials reported. Not only did the number of SPL trials vary, but the terminology used was considerably different. Ahlander et al. (2012) had participants perform 1 soft and 1 loud production and fill in the contour, another study had participants perform 1 messa di voce and compared with 1 trial for pianissimo, mezzo forte, and fortissimo (Lamarch, Ternström, & Pabon, 2010). While each of these used one trial, the reported terminology has different meanings for each participant and should therefore be streamlined with consistent terminology. This also raises the point that singers and non-singers would respond to this vocabulary in different ways. A trained singer would likely use their loudest and softest productions within only one production of a messa di voce, however, a recreational singer or non-singer would need several trials to accurately depict their loudest and softest productions.

Variables related to equipment and technology. The microphone to mouth distance was highly variable. This is partly accounted for because older literature used a 30 cm distance (Schutte & Seidner, 1983) before headsets were reliable and used with an electronic system. Eight percent of studies used a headset with a 1 cm to 3 cm distance, 15% of studies used a distance between 8 cm and 15 cm, 24% of studies used a 30 cm distance, 15% of studies used a preset distance and corrected the signal to 30 cm, and 35% of studies did not report the microphone distance. It is important in research for speech and voice that the microphone is used to convert sound pressure signals to an electronic signal with the same characteristics. Most microphones are not developed for these purposes, but are developed for recording or broadcasting (Svec & Granqvist, 2010). When choosing a microphone, the following needs to be considered; frequency and range of microphone, directionality, the dynamic range, the transducer type used, the microphone preamplifier used (Svec & Granqvist, 2010), and the distance from the mouth to the microphone (Coleman 1993). In ASHA's 2010 recommendations for microphone, clear specifications are provided for microphones (Patel et al., 2018). These recommendations include information regarding the lower and upper dynamic limits of the microphone, the lower and upper frequency limits of the microphone, and frequency response details (Svec & Granqvist, 2010). Microphone selection is of particular importance for the VRP due to the goal of recording the extremes of f_0 and dB SPL.

There have been several recommendations about the scale used for measurement of the sound level. Both the UEP (Schutte & Seidner, 1983) and Gramming and Sundberg (1988) recommended measuring the SPL with the A-weighted curve (dBA). Twenty-eight percent of included studies used dBA per these recommendations. Use of the dBC scale, which is considered a linear or flat frequency curve, is recommended in standard protocols for instrumental assessment of voice function (Patel et al, 2018) because it would provide uniform measurement of the frequency range and would not discriminate against the lower frequencies often found in speech and singing. Ternström et al. (2016) recommend using a flat frequency response if the environmental noise can be accounted for because the VRP should represent voice production and not voice perception. Nineteen percent of included studies reported use of the dBC scale, a linear scale, or a flat scale. Fifty-six percent of included studies did not report the sound level scale used. This is likely because computer programs attenuate the sound systematically and the scale used is often predetermined, unknown, or difficult to identify. Computer formatted programs should make this information easily accessible as the sound level scale used could influence a researcher or clinician's decision for purchase.

Variables that are largely unreported but were considered. Several data points are not part of the typical standard protocol, but were included in this study to see how often they were reported. Seventeen percent of included studies reported the VRP data could be sung, unsung, or either (Barrier, 1993; Camarrone et al., 2015; Hunter at al., 2006; Hunter & Titze, 2005, LeBorgne & Weinrich, 2002). Only 5% of included studies reported whether vibrato was used or allowed during data collection (Lamarche at al., 2010; LeBorgne & Weinrich, 2002). Studies reporting this information were largely related to singing, therefore differentiation of these points was more relevant. Most studies did not discuss whether data points were sung or if participants used vibrato and therefore fell into the "not reported" category.

Another factor that was largely unreported was the use of warm-up. Those studies that did include a warm-up often had the client glide up and down as they were learning the procedure (Coleman, 1003; Ma & Yiu, 2011) or the participant was asked to sustain an /a/ at their habitual pitch and explore their voice down and up (Heylen et al., 2002). While 89% of studies did not report the use of warm-up, it can be a helpful tool in the data collection process as it helps warm up the range of the participant to prepare for an accurate VRP and it gives the clinician an idea of the expected range in regards to the physiological and functional outer limits.

Lastly, coaching of the participant throughout the procedure was reported in only 30% of included studies. Studies that have used researcher or clinician coaching often use gestures, explanations, and demonstrations to help the participant to achieve the outer limits of their range and SPL (Coleman, 1993; Lycke & Siupsinskiene, 2016; Ma & Yiu, 2011; Speyer et al., 2003). Pabon and Plomb (1988) found that coaching the participant was important for the elicitation of optimal VRP results. Coaching should be reported more consistently because it may make a difference in the VRP outcomes.

Strengths and Limitations

Systematic review is the highest level of evidence available and to that end, this investigation provided valuable information regarding the variability of the ways in which the VRP is obtained. The systematic review approach allowed for identification of clinic and research practice patterns in a manner that is not possible via other methodology to provide a framework for standardization of use of the VRP in clinic and research. Despite the wide ranging use of VRP and prior published standards (Schutte and Seidner, 1983), the extent to which published methods varied could not be appreciated without this systematic review approach. Given the procedural variability identified and the importance of the VRP for both clinical and research purposes, procedural recommendations for standardization of the VRP are summarized in Table 7, which provides recommended protocols in detail within the methodological areas assessed in this systematic review. The table was crafted to allow for standardization considerations for both researchers and clinicians to be able to answer questions about vocal function for both physiological and functional abilities.

Table 7. Recommended procedures for functional and physiological VRP

Recommendation Area	Functional VRP	Physiological VRP
Microphone Specifications	Omnidirectional microphone positioned at a distance of 4-10 cm from the lips at an angle of 45 degrees – 90 degrees meeting the following specifications: (a) flat frequency response (b) noise level at least 10 dB lower than the sound level of the quietest vocal sound (c) upper limit of the dynamic range should be above the sound level of the loudest phonations.[1-3] To understand the basics of selecting microphones see Švec and Granqvist.[2]	
Environmental Noise	Document background noise levels for 5 seconds while the room is quiet. This should meet the following specifications **30 cm mic distance**: should be at least 10 dB weaker than the level of the quietest phonations in dBC and should be 25 dB weaker if using dBA.[1, 4] **Omnidirectional head mounted microphone:** < 35 dBA and < 48 dBC.[1, 4] If these qualifications cannot be met, then consider the use of a sound-treated environment.[1, 5] Unidirectional microphones may be used to reduce the impact of environmental noise, but SPL measures and spectral measures could be somewhat compromised because of the proximity effect.[2]	
Sound Scale	dBC[1, 6]	
Coaching/Instructions	Detailed instructions with ample coaching is important throughout the process. These include but are not limited to: (a) many examples before starting and during the procedure, (b) demonstrations, (c) a conversation about producing ugly sounds in the outermost limits of the voice, (d) trying to limit vibrato, and (e) gestures to elicit the extremes of the participant's vocal production.[5, 7-10]	

Warm-up	Should be included because it helps warm up the range of the participant to prepare for an accurate VRP and it gives the clinician an idea of the expected range in regards to the physiological and functional outer limits.[5, 8]	
Determination of Range	Go down as low as possible using stairsteps or glissandi[11, 12] and sustain lowest pitch for 2 seconds Repeat this 3 times for lower range[5] Repeat above procedure 3 times for the upper range	Go down as low as possible using stairsteps or glissandi.[11, 12] Repeat above procedure for the upper range *Note: This does not have to be sustained at the outer limits of the voice since it is physiological.
Determination of data points in the VRP	This can be done two different ways and is at the discretion of the clinician or researcher: (1) Determine ST range and divide by 11: this will include lowest, highest, and 10% intervals[13] (2) Have client produce each pitch in his or her entire range[8]	Since the outer limits of the voices production are difficult to reproduce and sustain, the following is recommended for the physiological VRP: (1) Determine ST range and divide by 11: this will include lowest, highest, and 10% intervals[13]
Trials and duration for each pitch	Match each frequency 1 time then produce 3 xs for loudest production and 3xs for the softest production with each production only being accepted if it is within 3 dB of other productions.[5, 14] *Note: Singers may require fewer trials Hold each pitch for 2 seconds *Note that highest frequency and lowest frequency for physiological VRP will likely not be sustained for sound level values	
Vowel Recommendation	Vowel of participants choice and may vary across the entire VRP[15]	

References: [1]Patel et al. 2018, [2]Švec & Granqvist, 2010, [3]Winholtz & Titze, 1997, [4]Sramkova, Granqvist, Herbst, & Švec, 2015, [5]Coleman, 1993, [6]Ternstrom et al., 2016, [7]Lycke & Siupsinskiene, 2016, [8]Ma & Yiu, 2011, [9]Speyer et al., 2003, [10]Pabon & Plomb, 1988, [11]Barrett et al., 2018, [12]Reich et al., 1990, [13]Orlikoff & Baken, 1993, [14]Siupsinskiene, Adamonis & Toohill, 2009, [15]Hoch et al., 2019

Despite the high level of evidence provided by a systematic review, limitations are acknowledged. Operationally defining adults as > 18 years of age resulted in the exclusion of many studies that described data collection methods in detail. Nineteen as the lowest age was determined for this systematic review to align with the World Health Organization (2013) given the inclusion of international publications. Sixty-four out of 203 full texts were excluded because they included participants 18 years or younger or they did not report the full age range of participants reported in the study. Inclusion of studies with 18 year olds may have influenced the outcomes of this systematic review.

The data points chosen for inclusion in this systematic review were not inclusive for all of the potential methodological aspects possible for collection of the VRP. The data points selected for inclusion were determined to be those that most contributed to and affected the determination of the outermost limits of a participant's range and sound level. Some data points that were initially collected were not easily synthesized, such as the functional versus physiological range and the technology used for determination of the VRP. The equipment used varied considerably as there were many programs reported and highly variable manual methods. Further, the technology was not always described in enough detail.

Conclusion

The findings of this study affirmed the hypothesis that the data collection methods for the VRP are highly variable and that a standard protocol would be of benefit to allow for comparison and analysis across research studies and clinical practice. Should the VRP be adopted as part of the suite of recommended acoustic measures considered for a standardized voice assessment, elimination of procedural variability will be of value when comparing data in a repeated measures design or for use of this measurement as an outcome measure.

References

Ahlander, V. L., Rydell, R., & Lofqvist, A. (2012). How do teachers with self-reported voice problems differ from their peers with self-reported voice health? *Journal of Voice, 26*(4), E149-E161.

Airainer, R., & Klingholz, F. (1993). Quantitative evaluation of phonetograms in the case of functional dysphonia. *Journal of Voice, 7*(2), 136-141.

Akerlund, L. (1993). Averages of sound pressure levels and mean fundamental frequencies of speech in relation to phonetograms: comparison of nonorganic dysphonia patients before and after therapy. *Acta Otolaryngologica, 113*(1), 102-108.

Akerlund, L., Gramming, P., & Sundberg, J. (1992). Phonetogram and averages of sound pressure levels and fundamental frequencies of speech - Comparison between female singers and nonsingers. *Journal of Voice, 6*(1), 55-63.

Anderson, J. A. (1999). *Vocal Function in Subjects with Compensated Unilateral Vocal Fold Paralysis Pre- and Post-Medialisation Thyroplasty.* (MQ46181 M.Sc.). University of Toronto (Canada), Ann Arbor. ProQuest Dissertations & Theses Global database.

Awan, S. N. (1991). Phonetographic profiles and F0-SPL characteristics of untrained versus traned vocal groups. *Journal of Voice, 5*(1), 41-50.

Barrett, E. A., Lam, W., & Yiu, E. M. L. (2018). Elicitation of minimum and maximum fundamental frequency and vocal intensity: Discrete half steps versus glissando. *Journal of Voice*.

Barrier, J. T. (1993). *The Development of Criteria for the Selection of Age-Appropriate Literature for the Senescent Voice.* (9322680 A.Mus.D.). The University of Arizona, Ann Arbor. ProQuest Dissertations & Theses Global database.

Bless, D. M., Baken, R. J., Hacki, T., Fritzell, B., Laver, J., Schutte, H., . . . Faure, M. (1992). International Association of Logopedics and Phoniatrics (IALP) Voice Committee discussion of assessment topics. *Journal of Voice, 6*(2), 194-210.

Böhme, G., & Stuchlik, G. (1995). Voice profiles and standard voice profile of untrained children. *Journal of Voice, 9*(3), 304-307.

Brown, W. S., Morris, R. J., Hicks, D. M., & Howell, E. (1993). Phonational profiles of female professional singers and nonsingers. *Journal of Voice, 7*(3), 219-226.

Camarrone, F., Ivanova, A., Decoster, W., De Jong, F., & Van Hulle, M. M. (2015). Stable voice clusters identified when using the maximum versus minimum intensity curve in the phonetogram. *Folia Phoniatrica et Logopaedica, 67*(5), 259-266.

CASP checklists. (2014). *Critical Appraisal Skills Programme (CASP): Making sense of evidence*.

Chen, S. H. (1996). Voice range profile of Taiwanese normal young adults: A preliminary study. *Zhonghua Yi Xue Za Zhi (Taipei), 58*(6), 414-420.

Chen, S. H. (2007). Sex differences in frequency and intensity in reading and voice range profiles for Taiwanese adult speakers. *Folia Phoniatrica et Logopaedica, 59*(1), 1-9.

Chen, S. H. (2008). Voice range profiles for tonal dialect of Min. *Folia Phoiatrica et Logopaedica, 60*(1), 4-10.

Coleman, R. F. (1993). Sources of variation in phonetograms. *Journal of Voice, 7*(1), 1-14.

Coleman, R. F., Mabis, J. H., & Hinson, J. K. (1977). Fundamental frequency-sound pressure level profiles of adult male and female voices. *Journal of Speech and Hearing Research, 20*(2), 197-204.

Coleman, R. F., & Mott, J. B. (1978). Fundamental frequency and sound pressure level profiles of young female singers. *Folia Phoniatrica et Logopaedica, 30*(2), 94-102.

Damsté, P. (1970). The phonetogram. *Practica oto-rhino-laryngologica, 32*(3), 185-187.

Gramming, P., & Akerlund, L. (1988). Non-organic dysphonia. II. Phonetograms for normal and pathological voices. *Acta Otolaryngologica, 106*(5-6), 468-476.

Gramming, P., & Sundberg, J. (1988). Spectrum factors relevant to phonetogram measurement. *The Journal of the Acoustical Society of America, 83*(6), 2352-2360.

Gramming, P., Sundberg, J., & Akerlund, L. (1991). Variability of phonetograms. *Folia Phoniatrica et Logopaedica, 43*(2), 79-92.

Hallin, A. E., Fröst, K., Holmberg, E. B., & Södersten, M. (2012). Voice and speech range profiles and Voice Handicap Index for males — methodological issues and data. *Logopedics Phoniatrics Vocology, 37*(2), 47-61.

Heylen, L., Wuyts, F., Mertens, F., De Bodt, M., & Van de Heyning, P. (2002). Normative voice range profiles of male and female professional voice users. *Journal of Voice, 16*(1), 1-7.

Heylen, L., Wuyts, F., Mertens, F., & Pattyn, J. (1996). Phonetography in voice diagnoses. *Acta oto-rhino-laryngologica Belgica, 50*(4), 299-308.

Hoch, M., & Sandage, M. J. (2017). Working toward a common vocabulary: Reconciling the terminology of teachers of singing, voice scientists, and speech-langauge pathologists. *Journal of Voice*, *31*(6), 647-648.

Holmberg, E. B., Ihre, E., & Södersten, M. (2007). Phonetograms as a tool in the voice clinic: Changes across voice therapy for patients with vocal fatigue. *Logopedics Phoniatrics Vocology, 32*(3), 113-127.

Holmes-Bendixen, A. R. (2013). *The Influence of Whistle Register Phonation Exercises in Conditioning the Second Passaggio of the Female Singing Voice.* (3595108 Ph.D.). The University of Iowa, Ann Arbor. ProQuest Dissertations & Theses Global database.

Hunter, E. J., Švec, J. G., & Titze, I. R. (2006). Comparison of the produced and perceived voice range profiles in untrained and trained classical singers. *Journal of Voice, 20*(4), 513-526.

Hunter, E. J., & Titze, I. R. (2005). Overlap of hearing and voicing ranges in singing. *Journal of Singing, 61*(4), 387-392.

Johansson, K., Seiger, A., Forsen, M., Nilsson, J. H., Hartelius, L., & Schalling, E. (2018). Assessment of voice, speech and communication changes associated with cervical spinal cord injury. *International Journal of Language & Communication Disorders, 53*(4), 761-775.

Keilmann, A., Napiontek, U., Engel, C., Nakarat, T., Schneider, A., & Mann, W. (2011). Long-term functional outcome after unilateral cordectomy. *Journal for Oto-rhino-laryngology and its Related Specialties, 73*(1), 38-46.

Kelly, V., Hertegard, S., Eriksson, J., Nygren, U., & Sodersten, M. (2018). Effects of gender-confirming pitch-raising surgery in transgender women a long-term follow-up study of acoustic and patient-reported data. *Journal of Voice*.

Kolker, A. (2017). *Practical Elicitation Methods for the Voice Range Profile.* (10256004 M.A.). Northern Illinois University, Ann Arbor. ProQuest Dissertations & Theses Global database.

Lamarche, A., Ternström, S., & Pabon, P. (2010). The singer's voice range profile: Female professional opera soloists. *Journal of Voice, 24*(4), 410-426.

Lamesch, S., Doval, B., & Castellengo, M. (2012). Toward a more informative voice range profile: the role of laryngeal vibratory mechanisms on vowels dynamic range. *Journal of Voice, 26*(5), 672 e679-618.

LeBorgne, W. D., & Weinrich, B. D. (2002). Phonetogram changes for trained singers over a nine-month period of vocal training. *Journal of Voice, 16*(1), 37-43.

Lee, H. Y., Lee, J. Y., Dionigi, G., Bae, J. W., & Kim, H. Y. (2015). The efficacy of intraoperative neuromonitoring during robotic thyroidectomy: A prospective, randomized case-control evaluation. *Journal of Laparoendoscopic and Advanced Surgical Techniques, 25*(11), 908-914.

Lycke, H., Decoster, W., Ivanova, A., Van Hulle, M., & De Jong, F. (2012). Discrimination of three basic female voice types in female singing students by voice range profile-derived parameters. *Folia Phoniatrica et Logopaedica, 64*(2), 80-86.

Lycke, H., Ivanova, A., Van Hulle, M. M., Decoster, W., & de Jong, F. (2013). Discrimination of three basic male voice types by voice range profile-derived parameters. *Folia Phoniatrica et Logopaedica, 65*(1), 20-24.

Lycke, H., & Siupsinskiene, N. (2016). Voice range profiles of singing students: the effects of training duration and institution. *Folia Phoniatrica et Logopaedica, 68*(2), 53-59.

Ma, E., Robertson, J., Radford, C., Vagne, S., El-Halabi, R., & Yiu, E. (2007). Reliability of speaking and maximum voice range measures in screening for dysphonia. *Journal of Voice, 21*(4), 397-406.

Ma, E. P. M., & Yiu, E. M. L. (2006). Multiparametric evaluation of dysphonic severity. *Journal of Voice, 20*(3), 380-390.

Ma, E. P. M., & Yiu, E. M. L. (2011). Voice range profile: Phog. *Handbook of voice assessments*, 253-267.

Mailander, E., Muhre, L., & Barsties, B. (2017). Lax Vox as a voice training program for teachers: A pilot study. *Journal of Voice, 31*(2), 262 e213-262 e222.

Marunick, M. T., & Menaldi, C. J. (2000). Maxillary dental arch form related to voice classification: a pilot study. *Journal of Voice, 14*(1), 82-91.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine, 151*(4), 264-269.

Neuschaefer-Rube, C., Sram, F., & Klajman, S. (1997). Three-dimensional phonetographic assessment of voice performance in professional and non-professional speakers. *Folia Phoniatrica et Logopaedica, 49*(2), 96-104.

Ohlsson, A. C., Dotevall, H., Gustavsson, I., Hofling, K., Wahle, U., & Osterlind, C. (2018). Voice therapy outcome-A randomized clinical trial comparing individual voice therapy, therapy in group, and controls without therapy. *Journal of Voice*.

Orlikoff, R. F., & Baken, R. J. (1993). Clinical application: The sound level meter, vocal intensity, and the voice range profile. In *Clinical Speech and Voice Measurement* (pp. 165-186). San Diego, California: Singular Publishing Group, Inc.

Pabon, J. (1991). Objective acoustic voice-quality parameters in the computer phonetogram. *Journal of Voice, 5*(3), 203-216.

Pabon, J., & Plomp, R. (1988). Automatic phonetogram recording supplemented with acoustical voice-quality parameters. *Journal of Speech, Language, and Hearing Research, 31*(4), 710-722.

Park, M. W., Baek, S. K., Park, E. H., & Jung, K. Y. (2018). Long-term voice outcome after thyroidectomy using energy based devices. *Auris Nasus Larynx, 45*(3), 527-532.

Patel, R. R., Awan, S. N., Barkmeier-Kraemer, J., Courey, M., Deliyski, D., Eadie, T., . . . 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.

Pei, Y. C., Chuang, H. F., Chang, C. F., Chang, T. L., Chiang, H. C., & Fang, T. J. (2018). Voice range change after injection laryngoplasty for unilateral vocal fold paralysis. *Journal of Voice, 32*(5), 625-632.

Printz, T., Rosenberg, T., Godballe, C., Dyrvig, A.K., & Grøntved, Å. M. (2018). Reproducibility of automated voice range profiles, a systematic literature review. *Journal of Voice, 32*(3), 273-280.

Printz, T., Sorensen, J. R., Godballe, C., & Grontved, A. M. (2018). Test-retest reliability of the dual-microphone voice range profile. *Journal of Voice, 32*(1), 32-37.

Raschle, N. M., Smith, S. A., Zuk, J., Dauvermann, M. R., Figuccio, M. J., & Gaab, N. (2014). Investigating the neural correlates of voice versus speech-sound directed information in pre-school children. *PLoS One, 9*(12), e115549.

Ray, C., Trudeau, M. D., & McCoy, S. (2018). Effects of respiratory muscle strength training in classically trained singers. *Journal of Voice, 32*(5), 644 e625-644 e634.

Reich, A. R., Frederickson, R. R., Mason, J. A., & Schlauch, R. S. (1990). Methodological variables affecting phonational frequency range in adults. *Journal of Speech and Hearing Disorders, 55*(1), 124-131.

Rendón, M. M. R., Ermakova, T., Freymann, M. L., Ruschin, A., Nawka, T., & Caffier, P. P. (2018). Efficacy of phonosurgery, logopedic voice treatment and vocal pedagogy in common voice problems of singers. *Advances in Therapy, 35*(7), 1069-1086.

Rothman, H., Nielsen, K., & Hicks, J. (1979). *Perceptual Classification of Voice Movements.* Paper presented at the Transcripts of the Eighth Symposium: Care of the Professional Voice at The Juilliard School. New York: The Voice Foundation.

Salmen, T., Ermakova, T., Schindler, A., KO, S.-R., Göktas, Ö., Gross, M., . . . Caffier, P. (2018). Efficacy of microsurgery in Reinke's oedema evaluated by traditional voice assessment integrated with the Vocal Extent Measure (VEM). *Acta Otorhinolaryngologica Italica, 38*(3), 194.

Schneider-Stickler, B., Knell, C., Aichstill, B., & Jocher, W. (2012). Biofeedback on voice use in call center agents in order to prevent occupational voice disorders. *Journal of Voice, 26*(1), 51-62.

Schneider, B., Bigenzahn, W., End, A., Denk, D. M., & Klepetko, W. (2003). External vocal fold medialization in patients with recurrent nerve paralysis following cardiothoracic surgery. *European Journal of Cardio-thoracic Surgery, 23*(4), 477-483.

Schneider, B., Denk, D. M., & Bigenzahn, W. (2003). Functional results after external vocal fold medialization thyroplasty with the titanium vocal fold medialization implant. *Laryngoscope, 113*(4), 628-634.

Schonweiler, R., Wohlfarth, K., Dengler, R., & Ptok, M. (1998). Supraglottal injection of botulinum toxin type A in adductor type spasmodic dysphonia with both intrinsic and extrinsic hyperfunction. *Laryngoscope, 108*, 55-63.

Schutte, H. (1980). Untersuchungen von stimmqualitäten durch phonetographie. *HNO-Praxis, 5*, 132-139.

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

Schwartz, S. M. (2009). Voice range profiles of middle school and high school choral directors. *Journal of Research in Music Education, 56*(4), 293-309.

Seidner, W., Wendler, J., Wagner, H., & Rauhut, A. (1981). Spektrales Stimmfeld. *HNO-Praxis, 6*, 187-191.

Sihvo, M., Laippala, P., & Sala, E. (2000). A study of repeated measures of softest and loudest phonations. *Journal of Voice, 14*(2), 161-169.

Sihvo, M., & Sala, E. (1996). Sound level variation findings for pianissimo and fortissimo phonations in repeated measurements. *Journal of Voice, 10*(3), 262-268.

Siupsinskiene, N., Adamonis, K., & Toohill, R. J. (2009). Usefulness of assessment of voice capabilities in female patients with reflux-related dysphonia. *Medicina (Kaunas), 45*(12), 978-987.

Speyer, R. (2008). Effects of voice therapy: A systematic review. *Journal of Voice, 22*(5), 565-580.

Speyer, R., Wieneke, G. H., van Wijck-Warnaar, I., & Dejonckere, P. H. (2003). Effects of voice therapy on the voice range profiles of dysphonic patients. *Journal of Voice, 17*(4), 544-556.

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

Storck, C., Brockmann, M., Schnellmann, E., Stoeckli, S. J., & Schmid, S. (2007). Functional outcome of vocal fold medialization thyroplasty with a hydroxyapatite implant. *Laryngoscope, 117*(6), 1118-1122.

Stout, B. (1938). The harmonic structure of vowels in singing in relation to pitch and intensity. *The Journal of the Acoustical Society of America, 10*(2), 137-146.

Sulter, A. M., Wit, H. P., Schutte, H. K., & Miller, D. G. (1994). A structured approach to voice range profile (phonetogram) analysis. *Journal of Speech, Language, and Hearing Research, 37*(5), 1076-1085.

Sundberg, J. (1987). *The Science of the Singing Voice*. Dekalb, Illinois: North Illinois University Press.

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

Tae, K., Kim, K. Y., Yun, B. R., Ji, Y. B., Park, C. W., Kim, D. S., & Kim, T. W. (2012). Functional voice and swallowing outcomes after robotic thyroidectomy by a gasless unilateral axillo-breast approach: comparison with open thyroidectomy. *Surgical Endoscopy, 26*(7), 1871-1877.

Teles-Magalhães, L. C., Pegoraro-Krook, M. I., & Pegoraro, R. (2000). Study of the elderly females' voice by phonetography. *Journal of Voice, 14*(3), 310-321.

Ternström, S., Pabon, P., & Södersten, M. (2016). The voice range profile: Its function, applications, pitfalls and potential. *Acta Acustica united with Acustica, 102*(2), 268-283.

Titze, I. R. (1992). Acoustic interpretation of the voice range profile (phonetogram). *Journal of Speech, Language, and Hearing Research, 35*(1), 21-34.

Titze, I. R., & Hunter, E. J. (2011). Feasibility of measurement of a voice range profile with a semi-occluded vocal tract. *Logopedics, Phoniatrics, Vocology, 36*(1), 32-39.

Titze, I. R., Wong, D., Milder Martin, A., Hensley Susan, R., & Ramig Lorraine, O. (1995). Comparison between clinician-assisted and fully automated procedures for obtaining a voice range profile. *Journal of Speech, Language, and Hearing Research, 38*(3), 526-535.

Tuomi, L., Johansson, M., Lindell, E., Folkestad, L., Malmerfors, M., & Finizia, C. (2017). Voice range profile and health-related quality of life measurements following voice rehabilitation after radiotherapy; A randomized controlled study. *Journal of Voice, 31*(1), 115 e119-115 e116.

Van Gogh, C. D. L., Leeuw, I. M. V., Boon-Kamma, B. A., Rinkel, R., de Bruin, M. D., Langendijk, J. A., . . . Mahieu, H. F. (2006). The efficacy of voice therapy in patients after treatment for early glottic carcinoma. *Cancer, 106*(1), 95-105.

Van Mersbergen, M. R., Verdolini, K., & Titze, I. R. (1999). Time-of-day effects on voice range profile performance in young, vocally untrained adult females. *Journal of Voice, 13*(4), 518-528.

Verdonck-de Leeuw, I. M., Hilgers, F. J. M., Keus, R. B., Koopmans-van Beinum, F. J., Greven, A. J., de Jong, J. M. A., . . . Bartelink, H. (1999). Multidimensional assessment of voice characteristics after radiotherapy for early glottic cancer. *Laryngoscope, 109*(2), 241-248.

Wingate, J. M., Brown, W. S., Shrivastav, R., Davenport, P., & Sapienza, C. M. (2007). Treatment outcomes for professional voice users. *Journal of Voice, 21*(4), 433-449.

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

Wolf, S., Stanley, D., & Sette, W. (1935). Quantitative studies on the singing voice. *The Journal of the Acoustical Society of America, 6*(4), 255-266.

World Health Organization: HIV/AIDS Definition of Key Terms. (2013). Retrieved from https://www.who.int/hiv/pub/guidelines/arv2013/intro/keyterms/en/

Wuyts, F., Heylen, L., Mertens, F., De Bodt, M., & Van de Heyning, P. (2002). Normative voice range profiles of untrained boys and girls. *Journal of Voice, 16*(4), 460-465.

Yiu, E., Xu, J. J., Murry, T., Wei, W. I., Yu, M., Ma, E., . . . Kwong, Y. L. (2006). A randomized treatment-placebo study of the effectiveness of acupuncture for benign vocal pathologies. *Journal of Voice, 20*(1), 144-156.

Yiu, E. M. L., & Chan, R. M. M. (2003). Effect of hydration and vocal rest on the vocal fatigue in amateur karaoke singers. *Journal of Voice, 17*(2), 216-227.