Reproducibility of Automated Voice Range Profiles, a Systematic Literature Review

Printz, Trine; Rosenberg, Tine; Godballe, Christian; Dyrvig, Anne Kirstine; Grøntved, Ågot Møller

Published in:
The Journal of Voice

DOI:
10.1016/j.jvoice.2017.05.013

Publication date:
2018

Document version:
Final published version

Document license:
CC BY-NC-ND

Citation for pulished version (APA):
Printz, T., Rosenberg, T., Godballe, C., Dyrvig, A. K., & Grøntved, Å. M. (2018). Reproducibility of Automated Voice Range Profiles, a Systematic Literature Review. *The Journal of Voice*, *32*(3), 273-280. https://doi.org/10.1016/j.jvoice.2017.05.013

Reproducibility of Automated Voice Range Profiles, a Systematic Literature Review

***Trine Printz, *Tine Rosenberg, *Christian Godballe, †Anne-Kirstine Dyrvig, and ‡Ågot Møller Grøntved,** *†‡*Odense C, Denmark*

Summary: Objective. Reliable voice range profiles are of great importance when measuring effects and side effects from surgery affecting voice capacity. Automated recording systems are increasingly used, but the reproducibility of results is uncertain. Our objective was to identify and review the existing literature on test-retest accuracy of the automated voice range profile assessment.

Study design. Systematic review.

Data sources. PubMed, Scopus, Cochrane Library, ComDisDome, Embase, and CINAHL (EBSCO).

Methods. We conducted a systematic literature search of six databases from 1983 to 2016. The following keywords were used: phonetogram, voice range profile, and acoustic voice analysis. Inclusion criteria were automated recording procedure, healthy voices, and no intervention between test and retest. Test-retest values concerning fundamental frequency and voice intensity were reviewed.

Results. Of 483 abstracts, 231 full-text articles were read, resulting in six articles included in the final results. The studies found high reliability, but data are few and heterogeneous.

Conclusion. The reviewed articles generally reported high reliability of the voice range profile, and thus clinical usefulness, but uncertainty remains because of low sample sizes and different procedures for selecting, collecting, and analyzing data. More data are needed, and clinical conclusions must be drawn with caution.

Key Words: Phonetogram–Voice range profile–Voice assessment–Test-retest–Reliability.

INTRODUCTION

When treating voice disorders, measurement of outcome as well as side effects is important, and objective methods of measurement are of importance in ear-nose-throat departments and in speech-language therapy clinics.[1] Knowing the reliability of the different assessment methods and types must be considered a minimum requirement if treatment results are to be correctly interpreted. The European Laryngological Society[1,2] and the Union of European Phoniatricians[3] recommend the use of voice range profile (VRP) when assessing the voice. This measures the maximum voice capacity in terms of limits in vocal fundamental frequency (f_o) and intensity—parameters that can be changed by disease and by treatment, and are of great significance for the functionality of the voice. Knowledge of VRP assessment reliability is sparse. Most likely, many possible influencing sources cause variation in the assessment, for instance, natural variation in the voice from day to day, different times of the day, with and without vocal warm-up, clinician's motivation and elicitation strategies, preciseness of the protocol, and more.[1,4–18]

Previously, VRPs were recorded by manual procedures, where the patient had to match and hold a tone for up to 3 seconds, while the clinician evaluated the f_o and read the sound level from a sound level meter.[19] The reliability of these manual procedures has been investigated in test-retest studies of healthy voices, where studies find the test-retest variation varying from 1 to 10 dB in intensity range and 1–4 semitones in frequency range.[13,16,19,20]

At present, the measurement is automated by the use of computer programs and corresponding equipment, which facilitate the process for both patients and clinicians.[21] Although there is still a need for a consistent clinician and protocol, the demand for the patients to match their pitch to a musical note and hold it steady for up to 3 seconds is no longer required, as some of the new automated methods require only very short tone durations.[21,22] Nowadays, very short phonation times will be detected and the voice is recorded and analyzed precisely in real time, rendering direct comparability between the new and the old methods very difficult. In addition, former data of variability and reproducibility cannot be considered representative for the automated VRP.[5,6] It would be reasonable to assume larger SPL variation, and thus decreased reliability, of the automated method, when the vocal production needs only to last milliseconds. However, the programs typically do not register all these very short phonations. Instead, they accumulate them, and only include them in the voice analysis when a certain time threshold has been reached, for example 0.1 sec.[21,23]

It is important to note that only the reliability of the automated VRP is assessed, and not the validity. Whereas reliability concerns the difference between two equal measurements (the same clinician measuring VRP on the same subjects, under the same recording conditions, using the same protocol), the validity concerns the amount of measurement error, and the preciseness of the results reflects reality.[24]

Based on a systematic literature review, we aimed to identify differences between test and retest of VRP in normophonic,

Accepted for publication May 16, 2017.
Declaration of Interest: The authors report no declarations of interest.
From the *Department of ORL Head & Neck Surgery, Odense University Hospital and Institute of Clinical Research, University of Southern Denmark, Odense C, Denmark; †Department of Surgery, Odense University Hospital, Odense C, Denmark; and the ‡Department of ORL Head & Neck Surgery, Odense University Hospital, Odense C, Denmark.
Address correspondence and reprint requests to Trine Printz, Department of ORL Head & Neck Surgery, Odense University Hospital and Institute of Clinical Research, University of Southern Denmark, Sdr. Boulevard 29, Indg. 84, 1. sal, 5000 Odense C, Denmark. E-mail: trine.printz@rsyd.dk
Journal of Voice, Vol. 32, No. 3, pp. 273–280
0892-1997

https://doi.org/10.1016/j.jvoice.2017.05.013

healthy voices using automated measurement, thus achieving a clearer insight into the assessment variation. In the present study, we use the term *automated* to cover VRPs recorded with computer program with a clinician or experimenter providing guidance, coaching, and encouragement to the patient.

METHODS

Study design

A systematic literature review was conducted. We adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses checklist and statement recommendations.[25]

Review objective

The electronic search strategy was guided by the study question: identify f_o and/or intensity differences between test and retest in the automated VRP of healthy voices.[14,26,27] The variables of interest were highest and lowest frequency and intensity, frequency and intensity ranges, and area (number of cells; VRP size) (Table 1, applied abbreviations).

Literature search

Information sources and search

The systematic literature search was constructed as a block search and conducted electronically on June 7, 2016. It was supervised by The Medical Research Library, the medical special library for The University Library of Southern Denmark. Six databases, including PubMed, MedLine, and Embase (Table 2), were searched for relevant articles in the time period from 1983 to 2016.

Studies before 1983 were considered irrelevant, as this was the promotional year for the first automated VRP technology.[4,5] We applied a core set of key words and reviewed search terms pertaining to the VRP (phonetogram, voice range profile, acoustic voice analysis, voice capacity assessment, etc). The specific use of search terms and truncations is provided in Table 1. Reference lists were reviewed for relevant literature not included in the database search. All titles and abstracts were downloaded onto the reference management database EndNote X6, Thomson Reuters, New York, NY. Duplicates and references that clearly deviated from the subject were removed.

TABLE 1.
Voice Range Profile Parameters With Abbreviations and Measure Units

Parameter	Abbreviation	Measure
Highest intensity	Max SPL	dB SPL
Lowest intensity	Min SPL	dB SPL
Intensity range (lowest to highest SPL)	SPL range	dB
Lowest frequency	Min f_o	Hz/ST
Highest frequency	Max f_o	Hz/ST
Frequency range (lowest to highest tone)	ST range	ST
Area (semitones times decibels/number of cells)	Area	(ST × dB)/cells

Abbreviations: dB SPL, decibel sound pressure level; Hz, hertz; ST, semitones.

TABLE 2.
List of Databases, Search Terms, and Truncations

Databases	Search Terms	Truncations
PubMed	Phonetogram	f_o: fundamental frequency
Cochrane Library	Phonetography	SPL: sound pressure level
ComDisDome	Voice range profile	
Embase	f_o	
CINAHL (EBSCO)	SPL	
Scopus	Acoustic voice analysis	
	Voice evaluation	
	Voice capacity	
	Voice assessment	

Inclusion process

Two independent raters (TR and TP) assessed abstracts for further inclusion. In cases of disagreement, discussions between raters led to agreement. At full-text ratings, "reason for exclusion" codes were used. Discussions were conducted at all disagreements, including incongruence in the codes. The two investigators read the full text together and discussed whether the codes were correct. A third investigator, either author A-KD or ÅMG, was involved in case of doubt or disagreement of technical or statistical and other questions, respectively. Here, the issue in question was discussed informally, yet in accordance with the eligibility criteria until agreement was reached.

Eligibility criteria

For an article to be included, it was required to:

- measure VRP with the automated measurement
- present quantitative assessment of data, for instance means and standard deviations or intraclass correlation coefficient on *at least one* of the following parameters:
  - o maximum intensity measured in dB SPL
  - o minimum intensity measured in dB SPL
  - o SPL range measured in dB (lowest to highest dB)
  - o maximum f_o measured in Hz or ST (highest tone)
  - o minimum f_o measured in Hz or ST (lowest tone)
  - o semitone range measured in Hz or ST (lowest tone to highest tone)
  - o area, measured in cells (size of the VRP)
- measure healthy voices with no history of voice intervention or treatment
- report no intervention, treatment or other possible influencing factors between the two tests
- have uniform recording conditions—both under and between test and retest
- be written in Danish, Norwegian, Swedish, English, or German

Studies that met these criteria were assessed with the Critical Appraisal Skills Programme's "Case control checklist" from May 31, 2013,[28] which assesses level of evidence including risk of bias. During the selection process, all subsections of articles were of interest. If a small test-retest study was part of a clinical study, for instance a case-control study presenting a control group with no intervention between test and retest, this was included in our study to ensure that as much relevant test-retest data from the literature as possible were included. The checklists were completed, considering these minor reliability tests as the "primary focus" of the article in question.

Statistics

Kappa statistics were used to calculate the level of interrater agreement using *STATA/IC* 13.1 (StataCorp LP, College Station, TX).

RESULTS

Systematic literature search

Study selection and data extraction

The initial electronic search resulted in 5455 titles and abstracts (Figure 1). After removing duplicates, 483 abstracts remained. Of these, 231 were included for further full-text

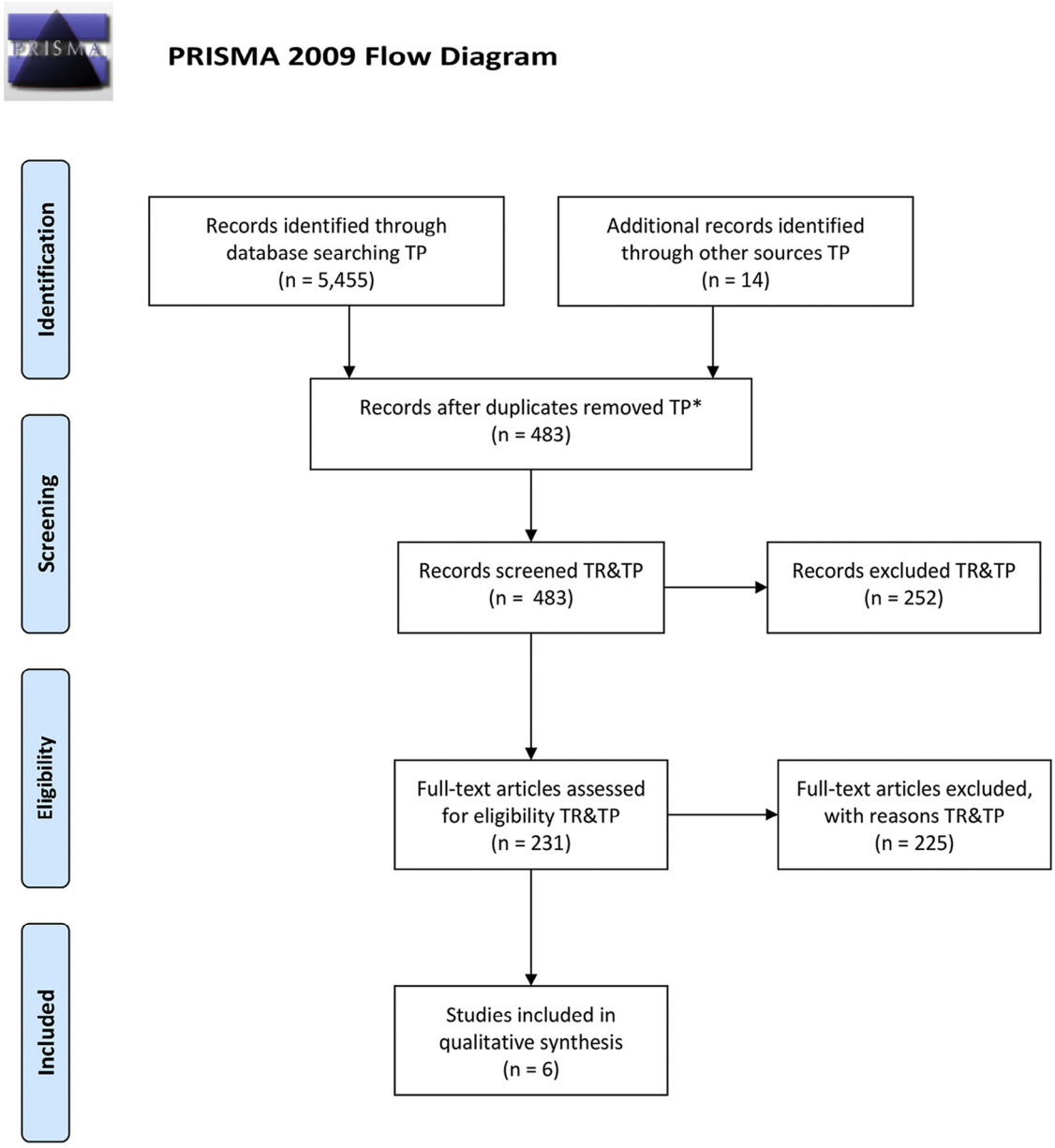

FIGURE 1. PRISMA flow diagram. The diagram shows the number of references retrieved through the entire search and selection process, starting with the database search (n = 5455) and other sources (n = 14) (*top*), through removal of duplications (n = 483), screening process, full-text inclusion (n = 231), and final inclusion in review (n = 6).

assessments. Kappa statistics for interrater agreement (abstract selection) was 0.82, equaling "substantial" interrater agreement ($P < 0.0001$). In full-text rating, two criteria were explored: (1) whether the study applied automated VRP equipment, and (2) whether the study held test *and* retest VRP data of healthy voices. These criteria were proven to be the two most frequent reasons for exclusion (180 full texts, comprising 77.9% of the studies). Six papers met the criteria for inclusion in the review (Tables 3 and 4). Kappa statistics for interrater agreement in full-text selection were 0.66 ($P < 0.0001$).

Six articles fulfilled all inclusion criteria (Tables 3 and 4). Together, they presented data for a total of 66 adult participants aged between 19 and 70 years with healthy voices. In one study, the voices were assessed in the morning and again in the afternoon; this was repeated 4 weeks later.[30] Between the morning and the afternoon recordings, all participants worked in a call center: more than 50% of them talked more than 8 hours a day, which has most likely put some strain on their voices. Therefore, the afternoon assessments were excluded from our review, whereas the set of morning data is included here. Automated VRPs and clinician guiding were used in all articles.

Four articles stated that test-retest was part of the aim and focus of the study[5,11,31,32]; the last two studies had different foci.[29,30] One tested voice use at work, and had two recordings well suited for our purpose[30]; the other was a therapy study, where only the control group receiving no intervention between the two tests was included in the review.[29] In only one study, both gender and specific age range of the test-retest participants was provided;

TABLE 3.
Studies Included in the Systematic Review (Year, Number, Gender and Age of Subjects, Aim and Focus of Study)

Author (year)	n	Gender and Age	Trained or Untrained Voices	Aim and Focus of Study
Sanchez et al (2013)[5]	6	3 males, 3 females(*)	Untrained healthy Australian voices	(1) To add to the body of knowledge about automatic phonetograms (2) Investigate the test-retest reliability of VRP data
D'Haeseleer et al (2013)[29]	7	Gender N/S (mean age 21.4, SD 1.8 y, range 19–25 y)	Students from a bachelor's program in music and performing musical art	Investigate the impact of manual circumlaryngeal therapy on the vocal characteristics of future elite vocal performers (the control group received no therapy, and is included in the present review)
Schneider-Stickler et al (2012)[30]	30	21 females, 16 males (mean age 29.6 y ± 8.5 y) (7 drop outs, leads to n = 30)	Untrained, but professional: employees at a tele-communication company	Examine the voice use at work and introduce biofeedback software into real-life workplace situation to improve vocal performance
Hallin et al (2012)[11]	3	Males(*)	Untrained Swedish speakers	Suggest protocols for recordings and analyses of speech range profiles and voice range profiles
Chen (2008)[31]	10	N/S(*)	Untrained Taiwanese Min or Mandarin speakers Hospital employees and university students	(1) Investigate the physiological frequency and intensity ranges of the tonal dialect of Min (2) Compare the physiological frequency and intensity ranges of Min with those of nontonal languages
Behrman et al (1996)[32]	10	8 females, 2 males (age 19–70)	Untrained hospital employees and speech-language pathology students	(1) Determine the important features of the contours of the VRPs of patients with organic pathology (2) Determine if the VRP is a clinically useful, within-subject measure of change in vocal function as a result of surgical intervention

* Specific age range for test-retest participants not stated.

TABLE 4.
Studies Included in the Systematic Review (Time Interval, Recording Protocol, and Applied Technology)

Study	Time Between Test and Retest	Recording Protocol	Technology
Sanchez et al (2013)[5]	2 wk	Protocol from Hallin et al, 2012[11]	Phog 2.0
D'Haeseleer et al (2013)[29]	20 min	Protocol from Heylen et al[33]	KayPENTAX Voice Range profile, Computerized Speech Lab (CSL)
Schneider-Stickler et al (2012)[30]	4 wk	Physiological VRP measurement ("singing VRP"): Glissandi up and down. Method thoroughly described, no protocol stated.	LingWAVES.
Hallin et al (2012)[11]	3–4 mo	Six-step protocol beginning with min SPL, lowering f_o followed by raising f_o. Then max SPL (loud voice lowering f_o followed by raising f_o. Then refining the max and min contours.	Phog
Chen (2008)[31]	1 wk	Sustained vowel /a/ for a minimum of 2 s for each semitone along the musical scale in the modal register and falsetto. At each target frequency, subjects produced softest and loudest voices.	Kay Elemetrics, Voice Range Profile, Model 4326 (CSL, Kay, model 4300)
Behrman et al (1996)[32]	2 wk	Phonation was elicited at each f_o of the semitone scale across the entire extent of the patients' frequency range. Upper contour was elicited as comfortably loud phonation and not as maximal physiological intensity	Kay Elemetrics Voice Range Profile Model 4326

mean age was, however, not stated.[32] The most frequent interval between test and retest was 1–4 weeks[5,30–32]; however, in one study, the retest was performed after 20 minutes with complete vocal rest in between[29] and in another after 3–4 months.[11] Recording protocols were somewhat similar, although variations in use of glissandi or tone-by-tone methods were applied. One study had a recording time limit.[30] Three types of equipment were used in the studies, Kay Elemetrics' Voice Profiler being the most frequent. With only six studies, it is not possible to analyze whether the reproducibility of these is alike.

Table 5 shows the VRP reproducibility reported by the six studies included in the review. Three studies provided reliability measures as correlations (r), and found high reliability, although one study found lower reliability in ST range and area. The other three studies reported results in means or medians. None of these provided data on all seven VRP variables. Difference from test to retest was 4 dB ($P = 0.107$) in max dB; and 1 dB in min SPL. Two studies reported max f_o differences of 78 Hz ($P = 0.500$) and 1 ST, respectively. Min f_o differences were 8 Hz ($P = 0.581$) and 2 ST. Differences in semitone range varied from almost no difference to 3 ST (+/− 5 ST). No studies reported specific differences in dB range or area.

Biases and limitations of the included studies

A risk of bias assessment was conducted by TR and TP (Table 6). Three types of bias were found: (1) lack of randomization, (2) lack of effectiveness of blinding, and (3) limitations in recording time. In one study, a limitation of the test was set to 20 minutes including questionnaires, voice recordings for acoustic analyses, and VRP measurements.[30] The limit of the retest was 10 minutes, including both voice recordings for acoustic analyses and VRP measurements. Another study provided only a 20-minute break for the voice to recover before retesting.[29] Neither of the studies used random selections of groups. Two studies included supra normal voice users, in the form of call center agents[30] and students from a bachelor's program in music and performing musical art,[29] which might lead to testing bias. In at least four of the studies, participants worked or studied in the same company or class; accordingly, a risk of ineffective blinding due to discussion of the study purpose was present.[29–32]

One study referred to the possible impact of giving the participants detailed information about study goals and vocal risk factors as well as the repetitions of VRP measurements as a potential bias of the study.[30] Although not mentioned, the latter might also apply to the other studies. Other potential biases emphasized in the studies are cooperation and motivation of the participants and lack of vocal warm-up before elicitation of the VRPs.

DISCUSSION

To the best of our knowledge, this is the first systematic review evaluating reproducibility in the automated VRP assessment. The precision of the assessment is essential when it should be trusted for clinical and research application.[1,5,7–12,14,15,18,34] Without this

TABLE 5.
Test-Retest Results of Included Studies

Study or Independent Variable	Max SPL (dB)	Min SPL (dB)	SPL range	Max f_o	Min f_o	ST range	Area
Sanchez et al (2013)[5]	High reliability(*)	High reliability(*)	High reliability(*)	High reliability(*)	High reliability(*)	High reliability(*)	High reliability(*)
D'Haeseleer et al (2013)[29]	Test: median: 108.0 dB Retest: median: 112.0 dB $P = 0.107$*	Test: median: 53.0 dB Retest: median: 54.0 dB $P = 0.071$*	–	Test: median: 1396.9 Hz Retest: median: 1318.5 Hz $P = 0.500$*	Test: median: 138.6 Hz Retest: median: 130.8 Hz $P = 0.581$*	–	–
Schneider-Stickler et al (2012)[30]	–	–	–	–	–	Test: 32 ST +/– 5 Retest: 35 ST +/– 5	–
Hallin et al (2012)[11]	–	–	r = 0.99	–	–	r = 0.69	r = 0.84
Chen (2008)[31]	r = 0.83 and 0.92	r = 0.83 and 0.92	r = 0.83 and 0.92	r = 0.83 and 0.92	r = 0.83 and 0.92	r = 0.83 and 0.92	–
Behrman et al (1996)[32]		Test: mean: 64.2 (mean SD 4.3) Retest: mean: 64.2 (mean SD 3.7)		Test: mean 40.0 ST Retest: mean 39.0 ST	Test: mean: 19.0 ST Retest: mean: 21.0 ST	Test: mean: 30.3 (mean SD 6.9) Retest: mean: 30.4 (mean SD 6.2)	–

Results are stated as they are reported in the respective papers according to the different statistic tests.
* No significant difference between test and retest, and thus indicates high reliability.

TABLE 6.
Risk of Bias Analysis in the Studies Included

Risk of Bias Assessment	Random Selection of Subjects	Blinding of Participants	Exposure Bias	Bias in Assessment Method	Selective Reporting
Sanchez et al (2013)[5]	No	?	Yes	No	No
D'Haeseleer et al (2013)[29]	No	?	Yes	Yes	No
Schneider-Stickler et al (2012)[30]	No	No	Yes	?	No
Hallin et al (2012)[11]	?	?	?	No	No
Chen (2008)[31]	No	?	Yes	?	No
Behrman et al (1996)[32]	No	?	Yes	?	No

knowledge, it remains unclear to what extent differences before and after treatment can be ascribed to changes in the voice as a result of the treatment or to general variability in the assessment. The majority of the literature analyzed addressed issues other than test-retest variance. Six articles were included in the final analyses.

We included only studies of healthy voices, as we were testing the reliability of the assessment, and not overall variation in dysphonic voices, which are very likely to vary more from test to retest.[13,27,35] This larger variation should be considered when applying the VRP (and all other voice assessments) clinically. Moreover, we included only physiological VRPs, as recommended by Pabon and the Voice Profiler Users Group.[36] In the physiological VRP, the aim is to detect the physiological boundaries, or extremes, and not only the most beautiful tones. For this reason, dB max and range results from one study were not included.[32] One could argue that the f_o ranges should also have been excluded, as f_o and intensity are related, and the participants in this article might not have reached their highest f_o, as only the maximal comfortable level, and not the extreme intensity level, was pursued. Including these figures would raise a question regarding the validity of the automated VRPs, which in this case might be reliable, but may be not valid in the sense of testing the extremes of the vocal range, which should be the aim, when assessing the VRP contour. Here, experimenter guidance and motivation, as well as a stable protocol, play a great role.[23]

The difference of 3 ST in semitone range when using automated VRP assessment is in accordance with the test-retest differences reported for the manual procedures.[19,20] However, in regard to measuring the dB level, the automated procedures, in general, seem to have a better reliability than the manual methods.[13,16,20,35] One possible explanation is the accumulation of time in the cells. The automated systems register the voice only when it hits a cell several times and then reach a predefined accumulated time.[23] Another explanation is the threshold for registering the min SPL value. The automated systems have different thresholds. For instance, Chen[31] and Behrman et al[32] used the Kay Elemetrics, Voice Range Profile, Model 4326 with a 50 dB SPL minimum threshold. Generally, healthy voices can phonate softer than the 50 dB SPL at a 30-cm measuring distance,[37] and therefore reach the 50 dB SPL threshold repeatedly during a recording. This might fictively improve reliability as the min SPL threshold is reached repeatedly, thus not showing the true variance of the voice. The differences in results support our assumption that reliability data for the manual VRP cannot be considered representative for the automated VRP.

Bias of included studies

Participants from two articles studied either speech-language pathology or music.[29,32] Accordingly, they might have special interest in, and insight into, the study goal, resulting in smaller test-retest differences than in a broader layperson population. In one article, participant selection relied only on the participants' subjective self-evaluation and reassurance of no previous history of dysphonia.[30] Moreover, they defined hypofunctionality as the inability to reach 90 dB SPL (at 30-cm microphone distance), but there was no clear statement as to whether participants with hypofunctional voices were excluded from the results, causing a potential bias in this article. Furthermore, they allowed only a very short and limited recording time for their VRPs.[30] In general, time limits in voice recordings can be problematic, owing to the variation in voice abilities, the participant's understanding of the task, need for breaks, etc.[2,5,6] This time limit might be part of the explanation for the larger test-retest differences found in this article. Another possible explanation might be that newer, and perhaps more sensitive, technology was deployed. Computer algorithms of the different automated VRP equipment are different and this could also lead to some bias.

The time intervals between test and retest varied from 20 minutes[29] to 3–4 months.[11] The VRP can be a strenuous test for the voice,[32] and a 20-minute break seems to be a relatively short time for the voice to recover. This could potentially induce a bias in that the voice is fatigued at the retest. Moreover, it is not clear whether there is a learning effect in the VRP, and how long this might last before wearing off, but it is probable. It is not possible to draw any conclusions about this from the present data, as there was no clear tendency for a learning effect, or the opposite, in the data, yet in the study that allowed the shortest break (20 minutes), "absence or minimal impact of learning effect" (page 4) was concluded.[29]

Generalization of results of this study

The six studies used different VRP equipment, and because this alone might induce variations owing to differences in computer algorithms, microphone stability, headset and microphone details, and sensitivity to noise,[23] generalization of the results is inadvisable, and no strong conclusions can be drawn. For the purpose of increasing the precision of the analyses, we tried

conducting a meta-analysis, but owing to limited amount of data, this was rejected.

Four of the studies outlined test-retest as a research focus. The largest study[30] included 30 participants, but their focus was to assess voice demands in call center employees and they had a time constraint on the VRP. This was suitable for their focus, but questionable regarding our study. Excluding their data, 26 participants are left in the review, instead of the previously stated 66. Most articles present only data on selected variables, and these vary between studies.[11,29–32] This might lead to a decrease in the power of the results, and thus the conclusions of the present study must be viewed with some caution. Estimation of disease-specific problems and the results of treatments are partly determined on the basis of VRP recording measurements, and even though clinical usefulness and reliable apparatus are indicated in this review, larger studies allowing for the clinical relevant differences in their estimation of number of participants are warranted.

CONCLUSION

This is the first literature review to specifically and systematically analyze the reliability of automated VRP assessment. The articles generally report high reliability of the VRP, and thus clinical usefulness, but uncertainty remains because of the low sample sizes and different procedures for selecting, collecting, and analyzing data. The current literature is not sufficient for clear results, and more studies with a higher level of evidence are warranted.

Acknowledgment

We acknowledge The Medical Research library at Odense University Hospital for their qualified assistance in the literature search.

REFERENCES

1. Dejonckere PH, Bradley P, Clemente P, et al. A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. Guideline elaborated by the Committee on Phoniatrics of the European Laryngological Society (ELS). *Eur Arch Otorhinolaryngol*. 2001;258:77–82.
2. Friedrich G, Dejonckere PH. [The voice evaluation protocol of the European Laryngological Society (ELS)—first results of a multicenter study]. *Laryngorhinootologie*. 2005;84:744–752.
3. Schutte HK, Seidner W. Recommendation by the Union of European Phoniatricians (UEP): standardizing voice area measurement/phonetography. *Folia Phoniatr (Basel)*. 1983;35:286–288.
4. Sulter AM, Wit HP, Schutte HK, et al. A structured approach to voice range profile (phonetogram) analysis. *J Speech Hear Res*. 1994;37:1076–1085.
5. Sanchez K, Oates J, Dacakis G, et al. Speech and voice range profiles of adults with untrained normal voices: methodological implications. *Logoped Phoniatr Vocol*. 2014;39:62–71.
6. Coleman RF. Sources of variation in phonetograms. *J Voice*. 1993;7:1–14.
7. Heylen L, Wuyts FL, Mertens F, et al. Normative voice range profiles of male and female professional voice users. *J Voice*. 2002;16:1–7.
8. Pabon P, Ternstrom S, Lamarche A. Fourier descriptor analysis and unification of voice range profile contours: method and applications. *J Speech Lang Hear Res*. 2011;54:755–776.
9. Speyer R, Wieneke GH, van Wijck-Warnaar I, et al. Effects of voice therapy on the voice range profiles of dysphonic patients. *J Voice*. 2003;17:544–556.
10. Holmberg EB, Ihre E, Sodersten M. Phonetograms as a tool in the voice clinic: changes across voice therapy for patients with vocal fatigue. *Logoped Phoniatr Vocol*. 2007;32:113–127.
11. Hallin AE, Fröst K, Holmberg EB, et al. Voice and speech range profiles and Voice Handicap Index for males—methodological issues and data. *Logoped Phoniatr Vocol*. 2012;37:47–61.
12. Ma E, Robertson J, Radford C, et al. Reliability of speaking and maximum voice range measures in screening for dysphonia. *J Voice*. 2007;21:397–406.
13. Sihvo M, Laippala P, Sala E. A study of repeated measures of softest and loudest phonations. *J Voice*. 2000;14:161–169.
14. van Mersbergen MR, Verdolini K, Titze IR. Time-of-day effects on voice range profile performance in young, vocally untrained adult females. *J Voice*. 1999;13:518–528.
15. Titze IR, Wong D, Milder MA, et al. Comparison between clinician-assisted and fully automated procedures for obtaining a voice range profile. *J Speech Hear Res*. 1995;38:526–535.
16. Gramming P, Sundberg J, Akerlund L. Variability of phonetograms. *Folia Phoniatr (Basel)*. 1991;43:79–92.
17. Titze IR. *Toward Standards for Voice Analysis and Recording*. Denver, CO: Denver Center for the Performing Arts; 1994.
18. Ferrone C, Galgano J, Ramig LO. The impact of extended voice use on the acoustic characteristics of phonation after training and performance of actors from the La MaMa Experimental Theater club. *J Voice*. 2011;25:e123–e137.
19. Awan SN. Phonetographic profiles and F0-SPL characteristics of untrained versus trained vocal groups. *J Voice*. 1991;5:41–50.
20. Siupsinskiene N, Lycke H. Effects of vocal training on singing and speaking voice characteristics in vocally healthy adults and children based on choral and nonchoral data. *J Voice*. 2011;25:e177–e189.
21. Pabon P. Voice profiler version 4.0. 2007.
22. Pabon J, Plomp R. Automatic phonetogram recording supplemented with acoustical voice-quality parameters. *J Speech Hear Res*. 1988;31:710–722.
23. Ternström P. The voice range profile: its function, applications, pitfalls and potential. *Acta Acust United Acust*. 2015;102:268–283.
24. Kottner J, Audige L, Brorson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *Int J Nurs Stud*. 2011;48:661–671.
25. Moher D, Shamseer L, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev*. 2015;4.
26. Rantala L, Vilkman E, Bloigu R. Voice changes during work: subjective complaints and objective measurements for female primary and secondary schoolteachers. *J Voice*. 2002;16:344–355.
27. Artkoski M, Tommila J, Laukkanen AM. Changes in voice during a day in normal voices without vocal loading. *Logoped Phoniatr Vocol*. 2002;27:118–123.
28. Critical Appraisal Skills Programme (CASP). *CASP Checklists*. Oxford: CASP; 2014.
29. D'Haeseleer E, Claeys S, Van Lierde K. The effectiveness of manual circumlaryngeal therapy in future elite vocal performers: a pilot study. *Laryngoscope*. 2013;123:1937–1941.
30. Schneider-Stickler B, Knell C, Aichstill B, et al. Biofeedback on voice use in call center agents in order to prevent occupational voice disorders. *J Voice*. 2012;26:51–62.
31. Chen SH. Voice range profiles for tonal dialect of Min. *Folia Phoniatr Logop*. 2008;60:4–10.
32. Behrman A, Agresti CJ, Blumstein E, et al. Meaningful features of voice range profiles from patients with organic vocal fold pathology: a preliminary study. *J Voice*. 1996;10:269–283.
33. Heylen L, Wuyts FL, Mertens F, et al. Evaluation of the vocal performance of children using a voice range profile index. *J Speech Lang Hear Res*. 1998;41:232–238.
34. Titze IR, The G. Paul Moore Lecture. Toward standards in acoustic analysis of voice. *J Voice*. 1994;8:1–7.
35. Sihvo M, Sala E. Sound level variation findings for pianissimo and fortissimo phonations in repeated measurements. *J Voice*. 1996;10:262–268.
36. Voice Profiler users Group CPP. Standard protocol phonetogram VRP recording. 2012.
37. Šrámková H, Granqvist S, Herbst CT, et al. The softest sound levels of the human voice in normal subjects. *J Acoust Soc Am*. 2015;137:407–418.