



Syddansk Universitet

Test-Retest Reliability of the Dual-Microphone Voice Range Profile

Printz, Trine; Sørensen, Jesper Roed; Godballe, Christian; Grøntved, Ågot Møller

Published in:
The Journal of Voice

DOI:
[10.1016/j.jvoice.2017.03.019](https://doi.org/10.1016/j.jvoice.2017.03.019)

Publication date:
2017

Document version
Peer reviewed version

Document license
CC BY-NC-ND

Citation for published version (APA):
Printz, T., Sørensen, J. R., Godballe, C., & Grøntved, Å. M. (2017). Test-Retest Reliability of the Dual-Microphone Voice Range Profile. *The Journal of Voice*. DOI: 10.1016/j.jvoice.2017.03.019

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Test-Retest Reliability of the Dual-Microphone Voice Range Profile

*†¹Trine Printz, *†¹Jesper Roed Sorensen, *Christian Godballe, and *Ågot Møller Grøntved, *Odense C, and †Denmark

Summary: Objectives. The voice range profile (VRP) measures vocal intensity and fundamental frequency. Phonosurgical and logopedic treatment outcome studies using the VRP report voice improvements of 3–6 semitones (ST) in ST range and 4–7 decibels (dB) in sound pressure level range after treatment. These small improvements stress the importance of reliable measurements. The aim was to evaluate the test-retest reliability of the dual-microphone computerized VRP on participants with healthy voices.

Study Design. This is a prospective test-retest reliability study.

Methods. Dual-microphone VRPs were repeated twice on healthy participants ($n = 37$) with an interval of 6–37 days. Voice frequency and intensity (minimum, maximum, and ranges) were assessed in combination with the area of the VRP.

Results. Correlations between VRP parameters were high ($r > 0.60$). However, in the retest, a statistically significant increase in voice frequency range (1.4 ST [95% confidence interval {CI}: 0.8–2.1 ST], $P < 0.001$), intensity ranges (2.2 dB [95% CI: 1.0–3.4 dB], $P < 0.001$), maximum frequency (1.0 ST [95% CI: 0.5–1.6 ST], $P < 0.001$), maximum intensity (1.4 dB [95% CI: 0.5–2.3 dB], $P = 0.002$), and area inside the VRP (148 cells [95% CI: 87–210 cells], $P < 0.001$) was observed.

Conclusion. The intra-examiner variation of the dual-microphone VRP is well below the differences seen after surgical or logopedic intervention, even when measuring in non-sound-treated rooms. There is a need for studies regarding inter-examiner reliability with a longer interval between test and retest before the assessment is fully reliable for clinical application.

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

INTRODUCTION

To complement the diagnosis of voice disorders and for documenting the outcomes after phonosurgery, both American and European associations of speech language pathologists and laryngologists recommend measuring vocal intensity and fundamental frequency.^{1–3} The measurements are presented in a two-dimensional diagram, the voice range profile (VRP). When using automated computerized methods for VRP recording, the fundamental frequency (f_0) and sound pressure level (SPL) can be measured in very short tone durations.^{4,5} This is designated the *computerized* VRP, as opposed to *manual* methods, requiring the patient to match vocal pitch to a musical note steadily for up to 3 seconds, allowing the examiner to judge pitch and measure SPL. In spite of the term “automated”, the assessment still requires a vigilant examiner providing guidance, coaching, and encouragement to the patient.

There are two types of computerized VRP methods: the single-microphone and the dual-microphone. The dual-microphone VRP

has improved stability for recordings at low SPLs due to a composition of a special headset with two microphones, one placed close to the mouth and the other 30 cm from the mouth. Before every new recording, an initial calibration detects the patient’s voice, which reaches the far microphone with a delay and thus a lower SPL. The system hereafter only accepts incoming sounds matching this pattern. Noise from the surroundings and the examiner’s voice is excluded and has no influence on the recording. Consequently, sound-treated rooms are not needed for the recording.⁵

Successful phonosurgical and logopedic treatment outcome studies report voice changes of 3–6 semitones (ST) in ST range^{6–8} and 4–7 dB in SPL range^{7,8} after treatment. Concerning the multiple causes for variation in the VRP, these small ST and dB differences stress the need for accurate assessments of measurement reliability. Previous VRP reliability studies using computerized setup with *single* microphones report high test-retest correlation (r) (defined as being $r > 0.60$ ⁹) in minimum f_0 (min f_0), maximum f_0 (max f_0), minimum SPL (min SPL), maximum SPL (max SPL),¹⁰ and VRP area.^{10,11} Behrman et al⁷ reported 1 ST difference from test to retest in min f_0 , and 2 ST in max min f_0 . D’Haeseleer et al¹² found 4 dB differences. Results from both studies are similar to or only just below the smallest treatment effect. However, due to the previously mentioned differences, as well as differences in microphone characteristics and algorithms for detecting and processing the incoming sound, test-retest results from single-microphone systems cannot be directly transferred to the dual-microphone system and new studies are warranted.¹³ We aimed at estimating the test-retest reliability in the VRP assessment of dual-microphone systems in participants with healthy voices.

Accepted for publication March 28, 2017.

Conflict of interest statement: None.

Source of financial support or funding: None.

The data were presented orally at the annual meeting for the Danish Society of Otolaryngology, Head & Neck Surgery, Nyborg, Denmark, May 12, 2016.

[†]The two authors contributed equally to the manuscript.

From the *Department of ORL Head & Neck Surgery, Odense University Hospital, Odense C, Denmark; and the †Clinical Institute, University of Southern Denmark, Denmark.

Address correspondence and reprint requests to Trine Printz, Audiopedagogics, Department of Clinical Research, University of Southern Denmark and Department of ORL Head & Neck Surgery, Odense University Hospital, J. B. Winsløvs Vej 4, 29 Indg. 84, 1. sal, Odense C 5000, Denmark. E-mail: trine.printz@rsyd.dk

Journal of Voice, Vol. ■■■, No. ■■■, pp. ■■■–■■■
0892-1997

© 2016 The Authors. Published by Elsevier Inc. on behalf of The Voice Foundation. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

<http://dx.doi.org/10.1016/j.jvoice.2017.03.019>

METHODS

The manuscript is in accordance with the Guidelines for Reporting Reliability and Agreement Studies.⁹ The guidelines recommend using the terms “interrater/intrarater reliability/agreement” in the title or abstract, but as there is no actual judgment in the VRP these terms have been replaced with test-retest reliability of the assessment.

Participants

For this prospective test-retest reliability study, we included adult (>18 years) normophonic participants. Exclusion criteria were prior voice disorders requiring treatment, ongoing upper respiratory tract infection, and trained singers, as the voice ranges of trained singers are not always representative of untrained individuals.^{14–16} JRS or TP made an informal perceptual voice assessment on all voices and excluded the participants if any abnormalities were found to be present. Also they completed the Voice Handicap Index (VHI) questionnaire. A VHI score of <18 points was accepted as no subjective voice complaints.¹⁷ All participants were recruited from hospital staff and their personal networks. Recruitment was conducted between June 2015 and February 2016.

Instrumentation, data collection, and analyses

For all voice recordings, the dual-microphone system Voice Profiler 5.0 (Alphatron Medical Systems, Rotterdam, The Netherlands) was employed. This device uses two cardioid-type microphones mounted on a headset: one positioned 2–3 cm from the mouth and the other 30 cm from the mouth. Although the close microphone produces signals with high signal-to-noise ratio, a small change in distance from the mouth will have a large effect on the SPL. Conversely, the far microphone stabilizes the SPL recording to prevent large SPL variations in the measurement if microphone distance to the subject changes.

Recordings were scheduled between 7:30 AM and 9:00 PM. Retests were scheduled within 6–37 days after the initial test. This period was chosen to limit the risk of voice changes between assessments. Two experienced examiners (a speech language pathologist and a medical doctor) handled the VRPs. Both examiners were experienced VRP users, having conducted >200 examinations independently. They were both trained in the VRP assessment protocol. Each patient had the same examiner throughout the study. The examiners were not blinded to the purpose of the study, but neither the participant nor the examiner had access to previous recordings. Recordings took place in the outpatient clinic. Room acoustics were not controlled or measured. All data were collected using REDCap electronic data capture tools hosted at Odense University Hospital, Odense, Denmark.¹⁸

Variables of interest were ST range, min f_0 and max f_0 , SPL range, min SPL, max SPL, and VRP area. Three independent variables were included: age, gender, and examiner.

Voice range profile recording procedure

The recording procedure was based on the principles of Hallin et al¹¹ and Sanchez et al,¹³ although extended to include all vowels. The microphone was situated just below the lower lip not touching facial hair. The mouth to microphone distance of the far

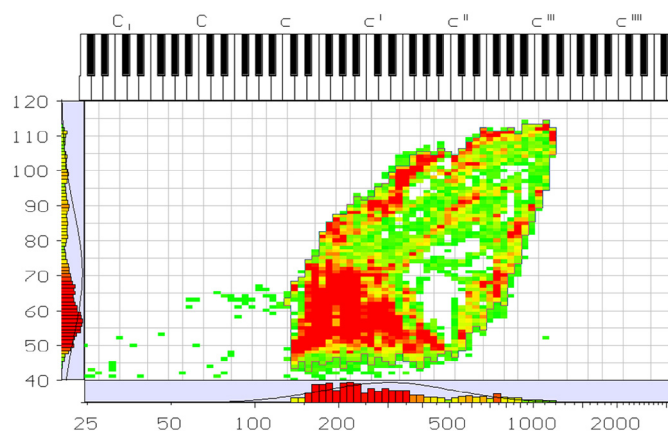


FIGURE 1. Example of voice range profile. Normal voice range profile with fundamental frequency in Hz on the x-axis, and vocal intensity in dB sound pressure level on the y-axis.

microphone was set to 30 cm. The directions of both microphones were checked to make sure they aimed directly at the mouth. The following calibration required the participant to say /he:i:./ until the Voice Profiler accepted the calibration.⁵

Both examiner and participant faced the computer screen, and guided the participant in how to reach the maximum boundaries of his or her voice. The participant went through the following steps: (1) soft tone using an easy pitch, (2) raise the pitch while staying soft, (3) recording the bottom octave: finding the lowest tone (yawn), (4) recording the bottom octave: singing loud, (5) highest and loudest tones in chest/modal voice, (6) head-/falsetto register: soft onsets, tone-by-tone upward, and (7) finalize with high and soft tones.

Elicitation strategies were the same in both test and retest. There was no time limit or upper boundary in how many times the participant could try to reach each ST and SPL. The computerized piano embedded in the VRP software, the examiner’s voice, and verbal plus visual cues guided the participants to the different VRP areas through the assessment. Primarily, tone-by-tone and gliding tones (high to low and low to high) were used, but also other forms of elicitation strategies, such as long and short both rising and falling tones in the upper contours, and shouting /haHA:./ on gliding tones from high to low. All strategies were applied in every recording, but for each participant the strategies that led to most cells in the VRP were preferred and exerted most. Visual assistance was provided on the two-dimensional graph. The maximal outside contour of the voice was the aim of the assessment, whereby the inner VRP contour (area inside) was not filled out (for an example of a VRP, see Figure 1). To adhere to the protocol, excessive glottal fry, strain, or “screaming” quality were excluded from the recording.¹³ Register changes, seen by a dip or disruption in the maximum contour between chest and falsetto register, were registered for most participants. An unlimited amount of water was provided before and during the recording; the amount was not measured. Before ending the VRP, the participant was encouraged to try all outer contours and see if they could be extended anywhere. The recording was ended when both tester and participants agreed that the maximum phonation area had been reached.

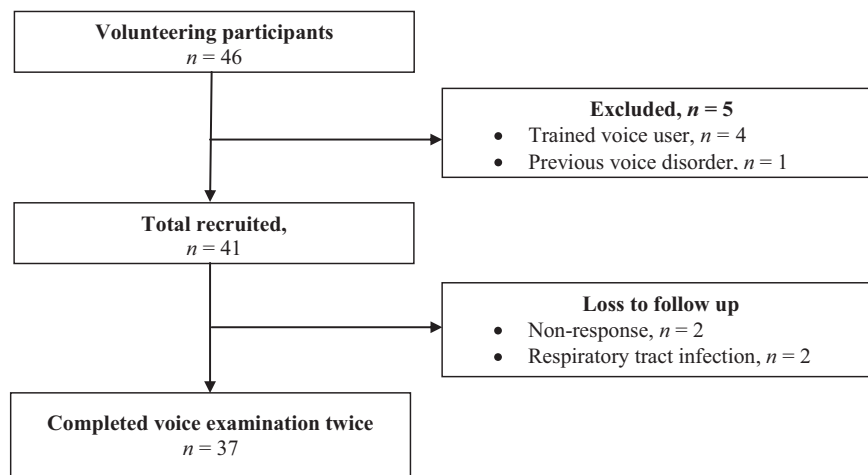


FIGURE 2. Flow diagram of participant inclusion.

Statistics

Pearson's correlation coefficient was used to evaluate the correlation between test and retest, as the data are from continuous scales. The differences between tests are presented by Bland-Altman plots. A paired *t* test was used to estimate the differences between analyses.⁹ The min SPL scale was reciprocally transformed to adjust for non-normality. To evaluate the effect of age, sex, and assessment number, we used a multiple linear regression analysis. The effect of time between assessments (above or below a median of 19 days between assessments) did not influence the results and was therefore not included in the final model. The sample size was calculated to 36 participants based on a minimum difference of 2.5 ± 5.2 ST in ST range.¹⁹

Ethics

The study was approved by the Danish Data Protection Agency. Quality control studies do not, by Danish law, require approval from the Regional Committees on Health Research Ethics in Southern Denmark, and approval can therefore not be obtained.

RESULTS

Participant characteristics

Of the 46 volunteering participants, five participants were excluded as they were either trained voice users or had a prior voice disorder, leaving 41 participants for the initial analyses (Figure 2). The retest was completed by 37 participants. Four participants were excluded due to either non-response or a respiratory tract infection at the time of the second assessment. Two minimum SPL recordings were excluded due to equipment malfunction in the minimum SPL recording, leaving 35 participants for the minimum SPL and SPL range measures. TP was responsible for 24 (65%) and JRS for 13 (35%) of the assessments, but all data were combined into a single group for statistical power.

The mean age of participants was 35 ± 11 years. Of the participants, 62% were female, and 8% consumed tobacco products on a regular basis (Table 1). No change in VHI scores was observed between the assessments and all scores were below 18 points, indicating no vocal disease among participants.¹⁴ No participants were excluded on the basis of the perceptual assessment.

Voice range profile

The VRP scores had a high correlation ($r > 0.60$) between test and retest for each participant (Table 2). In addition, Bland-Altman plots showed an equal distribution of differences throughout the VRP scores (Figure 3). The participants experienced a small increase in their voice capacity between test and retest. The initial ST range of 38.4 ± 3.7 ST increased by 1.4 ST (95% confidence interval [CI]: 0.8–2.1 ST) in the retest ($P < 0.001$). The increase was primarily in max f_0 by 1.0 ST (95% CI: 0.5–1.6 ST) ($P < 0.001$). Likewise, the SPL range of 69.8 ± 4.9 dB in the first test was amended by 2.2 dB (95% CI: 1.0–3.4 dB) in the retest ($P < 0.001$). The enlargement also took place in max SPL by 1.4 dB (95% CI: 0.5–2.3 dB) ($P = 0.002$). The participants enlarged the VRP area by 148 cells (95% CI: 87–210 cells) after they had performed the test once ($P < 0.001$). No change was observed in min f_0 or min SPL.

By multiple linear regressions, there was a significant difference in min SPL between the two examiners, as examiner 1 reached a 2.4 dB lower min SPL level than examiner 2. Further, males had, at baseline, a larger SPL range (coefficient 3.6 dB, $P = 0.045$) and ST range (coefficient 4.0 ST, $P = 0.01$) compared with females, but females increased more in both SPL range ($P = 0.003$) and ST range ($P = 0.002$) at the second assessment.

TABLE 1. Sociodemographic Characteristics of Participants Without Prior Voice Problems or Voice Training ($n = 37$)

	Mean \pm SD	
Age (years)		35 ± 11
Sex	Female, n (%)	23 (62%)
	Male, n (%)	14 (38%)
Consumer of tobacco	No, n (%)	30 (81%)
	Yes, n (%)	3 (8%)
	Not stated, (%)	4 (11%)
VHI score	Exam 1, mean \pm SD	6 ± 6
	Exam 2, mean \pm SD	6 ± 9
Days between exams	Mean (range)	16 (6–37)

Abbreviations: *n*, number; SD, standard deviation; VHI, Voice Handicap Index.

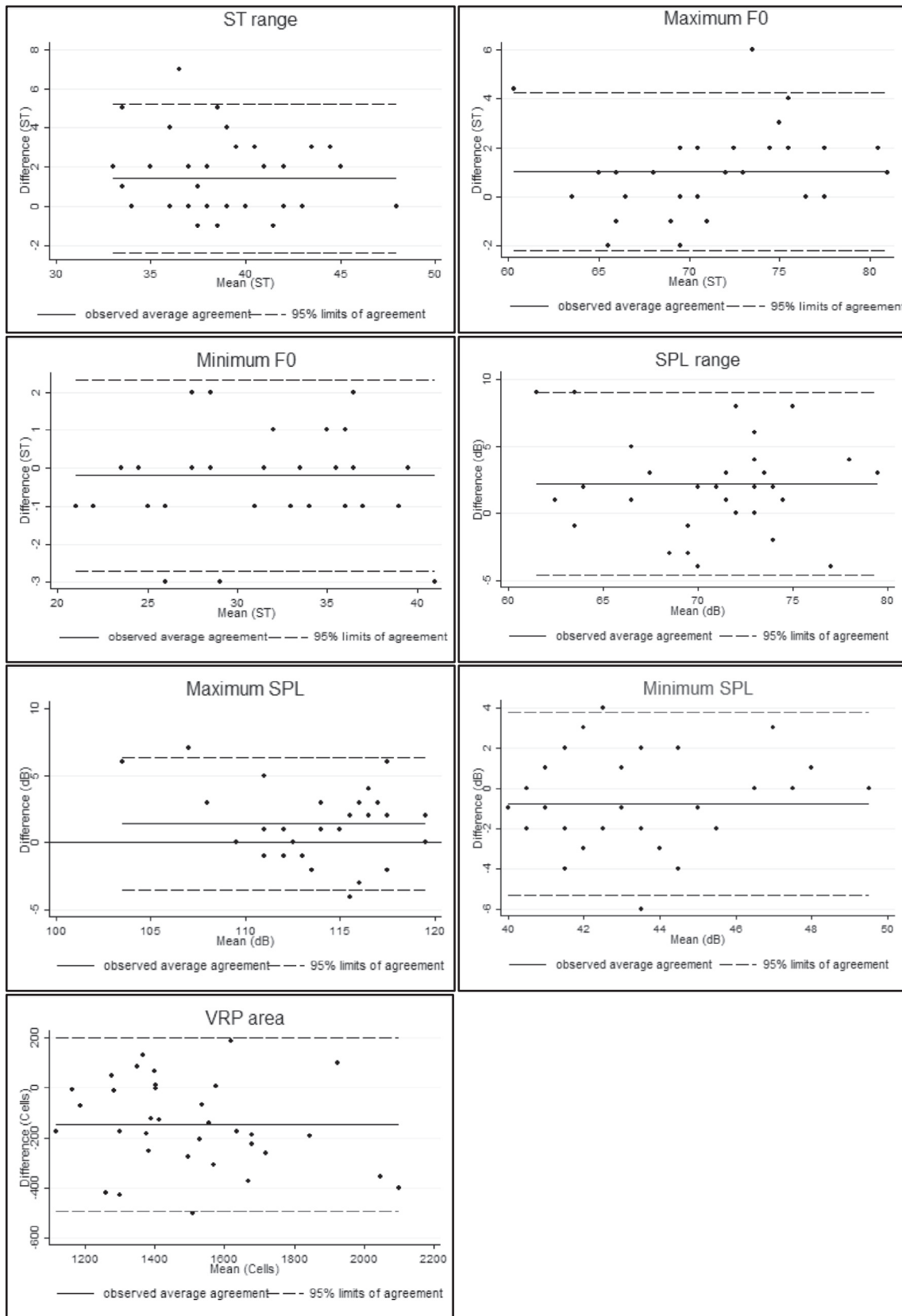


FIGURE 3. Bland-Altman plots of differences between first and second assessment showing observed average agreement and 95% limits of agreement. ST, semitone; f_0 , fundamental frequency; SPL, sound pressure level; dB, decibel; VRP, voice range profile.

TABLE 2.
Test-Retest Reliability in the VRP of Healthy Participants ($n = 37$)

Voice Range Profile	Test Mean \pm SD	Retest Mean \pm SD	Correlation r	Change Mean (95% CI)	P Value
Semitone range	38.4 \pm 3.7	39.8 \pm 3.4	0.85	1.4 (0.79–2.08)	<0.001
Minimum f_0 (ST)	31.6 \pm 5.2	31.4 \pm 5.3	0.97	–0.2 (–0.62 to 0.24)	0.38
Maximum f_0 (ST)	70.1 \pm 4.7	71.1 \pm 5.1	0.95	1.0 (0.49–1.59)	<0.001
SPL range (dB)	69.8 \pm 4.9	72.1 \pm 4.5	0.73	2.2 (0.96–3.39)	<0.001
Min SPL (dB)	43.7 \pm 2.4	42.8 \pm 2.8	0.61	–0.8 (–1.58 to 0.05)	0.05
Max SPL (dB)	113.5 \pm 4.3	114.8 \pm 3.5	0.80	1.4 (0.54–2.28)	0.002
VRP area (cells)	1442 \pm 239	1586 \pm 275	0.77	148 (87–210)	<0.001

Notes: Significant P values for comparison of mean scores between the two exams ($P < 0.05$).

Abbreviations: CI, confidence interval; dB, decibel; f_0 , fundamental frequency; SPL, sound pressure level; ST, semitone; VRP, voice range profile.

Younger people increased more in the SPL range (coefficient 0.13 dB, $P = 0.03$) and in area (coefficient 8.1 cells, $P = 0.005$) in the retest compared with older individuals, indicating that older participants were more consistent in their overall performance or that younger participants have more vocal flexibility.

DISCUSSION

Hitherto, this is the largest study to examine the test-retest reliability in the computerized dual-microphone VRP assessment.^{7,10–12,20} Further, more VRP parameters were included than seen in most previously published studies,^{7,10,11} and the loss to follow up was very limited (<10%).

The test-retest variation of the VRP assessment presented changes between the first and second assessment of 2 ST in ST range and 3 dB in SPL range. These changes are smaller than reported by single-microphone studies.^{7,12} Even more importantly, the variation is below the observed changes of 3–6 ST in ST range^{6–8} and 4–7 dB in SPL range^{7,8} after phonosurgical or logopedic intervention. In combination with other possible sources of variation, the total variation might be even higher. This underlines the need for a reliable VRP protocol if the assessment is intended to measure treatment effects.

The min f_0 and max f_0 correlations were higher than those reported by Chen.¹⁰ This might be a consequence of a more precise measurement device or protocol, or that we included more than three times the participants previously reported. Also, there were differences between our results and the previous SPL results. This may be caused by diversity in the VRP systems' microphone weighting and ability to detect, record, and store the voice signal.²¹ We used a dual-microphone system with a microphone close to the mouth capable of detecting the voice even at very low SPLs, even though it is the far microphone at 30 cm distance that actually records the voice and registers it on the screen.⁵ Hallin et al¹¹ measured the voices in their study at 15 cm microphone distance from the lips and added 6 dB after the test to adjust the test results for the difference between 15 and 30 cm in recommended measuring distance²². Furthermore, different min dB and max dB thresholds render comparisons difficult. The VRP device we used was able to detect min SPL as low as 40 dB SPL, whereas the min SPL threshold in Chen's study was 50 dB.¹⁰

This might also explain the higher variation in our study than reported by Chen.¹⁰ Other possible explanations for different study results might be patient-related factors and examiner experience.

The results showed higher scores on the majority of variables in the second assessment, with the lower limit of the 95% CI above 0. This indicates a systematic variation, which could be caused by a learning curve. In the second test, the participant knew the examiner and the assessment technique, allowing for a faster assessment putting less strain on the participant's voice. It was not possible to blind from knowing whether it was a test or a retest. Both examiner and participant perform the VRP in close cooperation. However, examiners were not allowed to look at the results from the VRP until having conducted the retest. The examiner could be biased by knowledge of the participant's abilities to reach specific f_0 s or SPLs. Thus, the observed increase in scores in the second examination might be a combination of the participant's familiarity with the VRP task and the examiner gaining knowledge of each participant's voice use. The VRP depends on the clinician's input and adjustment to the specific patient or research participant. The goal is to achieve the maximal VRP, and therefore different modeling strategies or ways of working with the patient within the limits of the protocol are necessary.

The recruitment strategy represents a weakness, as the use of family and friends of healthcare personnel promoted the study to more female than male participants and also possibly led to recruitment of younger individuals. This leads to some limitations in the generalizability of the results. The sample size of 37 individuals completing the VRP assessment twice is small, although more than three times the size of previously published studies.^{10,11,13}

Participants were examined with 6–37 days between assessments to estimate the test-retest reliability. With this short time frame between assessments, significant voice changes between tests were not expected. Nonetheless, this also leads to a potential risk of the participants having recollection bias of the previous assessment. This could explain the observed increase in VRP parameters at the second examination. In clinical assessments, there are often more than 37 days between each assessment, leaving the results with greater reliability than observed in our study.

With the present results, test-retest reliability of the VRP assessment using dual-microphone headset is thoroughly investigated. The uncertainties associated with the VRP assessment are with this study now known in greater detail. Although we observed significant increases in the majority of parameters in the second examination, these were too small to pose a clinical problem. Further, none of the parameters deteriorated in the second examination, making it easy to adjust for this uncertainty in future studies. However, it is important to point out that other variables such as time of day for recording, vocal warm-up, oral opening, inter-reliability variation, and differences in equipment might add further variation into the assessment. These variables must also be investigated further, in a structured and systematic manner, with large numbers of participants matched to the typical voice patient group in age and gender. Furthermore, the possible influences among examiners must be assessed, as education, experience, and training might affect VRP results. With this knowledge, adjustments to the VRP assessment protocol can be made and account for the variables in everyday clinical practice. Lastly, results from one dual-microphone setup are not directly transferable to other types of devices, due to differences in software, voice sound acquisition systems, and storage of the sound. Therefore, additional reliability studies conducted with other types of equipment are also warranted.

CONCLUSION

The intra-examiner variation of the dual-microphone VRP is well below the differences seen after surgical or logopedic intervention, even when measuring in non-sound-treated rooms. There is a need for studies regarding inter-examiner reliability with a longer interval between test and retest before the assessment is fully reliable for clinical application.

REFERENCES

- American Speech-Language-Hearing Association. Scope of Practice in Speech-Language Pathology [Scope of Practice]; 2016. Available at: www.asha.org/policy. Retrieved November 25, 2016.
- 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 Phoniatics of the European Laryngological Society (ELS). *Eur Arch Otorhinolaryngol*. 2001;258:77–82.
- 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.
- Pabon JPH, Plomp R. Automatic phonetogram recording supplemented with acoustical voice-quality parameters. *J Speech Hear Res*. 1988;31:710–722.
- Pabon P. Manual Voice Profiler Version 4.2. Rotterdam, The Netherlands: Alphon Medical Systems; 2007. Available at: <http://kc.koncon.nl/staff/pabon/OtherActivities/VoiceProfiler/ManualVPjan07.pdf>. Retrieved November 25, 2016.
- D'Alatri L, Galla S, Rigante M, et al. Role of early voice therapy in patients affected by unilateral vocal fold paralysis. *J Laryngol Otol*. 2008;122:936–941.
- 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.
- Schneider B, Denk DM, Bigenzahn W. Functional results after external vocal fold medialization thyroplasty with the titanium vocal fold medialization implant. *Laryngoscope*. 2003;113:628–634.
- 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.
- Chen SH. Voice range profiles for tonal dialect of Min. *Folia Phoniatr Logop*. 2008;60:4–10.
- 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.
- 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. doi:10.1002/lary.24050.
- 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.
- Hunter EJ, Svec JG, Titze IR. Comparison of the produced and perceived voice range profiles in untrained and trained classical singers. *J Voice*. 2006;20:513–526.
- Sulter AM, Schutte HK, Miller DG. Differences in phonetogram features between male and female subjects with and without vocal training. *J Voice*. 1995;9:363–377.
- Morris RJ, Brown WS, Hicks DM, et al. Phonational profiles of male trained singers and nonsingers. *J Voice*. 1995;9:142–148.
- Chandrasekhar SS, Randolph GW, Seidman MD, et al. Clinical practice guideline: improving voice outcomes after thyroid surgery. *Otolaryngol Head Neck Surg*. 2013;148(6 suppl):S1–S37.
- Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42:377–381.
- Leino T, Laukkanen AM, Ilomaki I, et al. Assessment of vocal capacity of Finnish university students. *Folia Phoniatr Logop*. 2008;60:199–209.
- D'Haeseleer E, Claeys S, Wuyts F, et al. Vocal quality in university teachers: a pilot study. *B-ENT*. 2009;5:101–109.
- Ternström S, Pabon P, Södersten M. The voice range profile: its function, applications, pitfalls and potential. *Acta Acust United Ac*. 2016;102:268–283.
- Schutte HK, Seidner W. Recommendation by the Union of European Phoniaticians (UEP): standardizing voice area measurement/phonetography. *Folia Phoniatr (Basel)*. 1983;35:286–288.