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

*,†,[1]Trine Printz, *,†,[1]Jesper 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_o) 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_o (min f_o), maximum f_o (max f_o), minimum SPL (min SPL), maximum SPL (max SPL),[10] and VRP area.[10,11] Behrman et al[7] reported 1 ST difference from test to retest in min f_o, and 2 ST in max min f_o. D'Haeseleer et al[12] 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.
[1]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, Audiologopedics, Department of Clinical Research, University of Southern Denmark and Department of ORL Head & Neck Surgery, Odense University Hospital, J. B. Winsløws 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

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_o and max f_o, 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 al[11] 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 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]

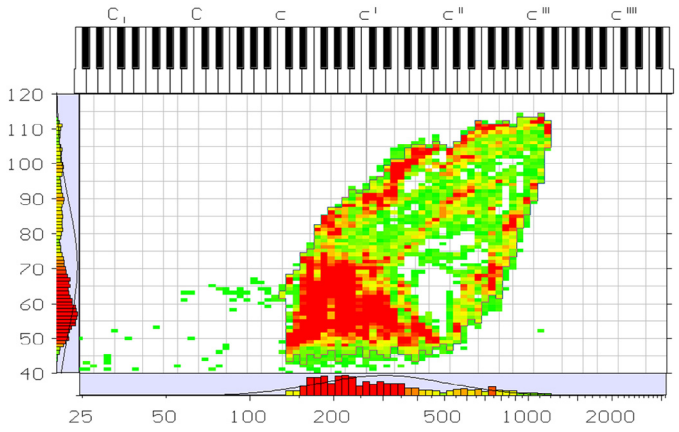

FIGURE 1. Example of voice range profile. Normal voice range profile with fundamental frequency in Hz on the x-axis, and vocal intensity in dB sound pressure level on the y-axis.

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.