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, healthy voices using automated measurement, thus achieving a

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. ■■, No. ■■, pp. ■■-■■

0892-1997

http://dx.doi.org/10.1016/j.jvoice.2017.05.013

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