

The Influence of Voice Sample Length in the Auditory-Perceptual Judgment of Overall Voice Quality

*†Ben Barsties and *‡§Youri Maryn, *§ and ‡Antwerp, Belgium, and †Hamburg, Germany

Summary: Objectives. The aim of this study was to explore the influence of voice sample lengths (VSLs) on the perceived degree of severity of overall voice quality. To increase a valid judgment in voice quality, a consistent rating is essential to estimate the presence and degree of severity of a voice.

Methods. Three VSLs were defined by varying only the length of continuous speech followed by constant duration of 3 seconds of the mid-vowel portion sustained vowel [a:]. VSL-1 contained 17 syllables as is used for the Acoustic Voice Quality Index by Maryn et al (2010). VSL-2 consisted of a customized length (ie, with a mean number of 35.5 syllables of the original text, which corresponds to 3 seconds of only-voiced segments in continuous speech). VSL-3 had 93 syllables, which represent a complete phonetically balanced text. An expert panel of five judges rated 100 voice samples with various degrees of hoarseness. In total, three sessions were achieved judging one VSL variant in one session of the 100 voice samples.

Results. The results showed significant differences of auditory-perceptual judgment between VSL-1 and VSL-2 ($U = -4.985, P = 0.000$), and between VSL-1 and VSL-3 ($U = -5.642, P = 0.000$). However, judgments of VSL-2 and VSL-3 did not differ significantly ($U = -1.785, P = 0.074$).

Conclusion. Based on these results, it can be concluded that the judgment of VSL-2 is comparable with the judgment of VSL-3. Therefore, speech material for clinical judgment can be reduced to VSL-2 for rating overall voice quality.

Key Words: Voice quality–Auditory-perceptual judgment–Sample length–Continuous speech–Sustained vowel.

INTRODUCTION

Voice quality is a feature of the voice sound described as a perceptual phenomenon but is not clearly defined.¹ Commonly, voice quality is defined based on exclusion. For example, pitch, loudness, and phonetic categories do not take voice quality into consideration.² Voice quality may be highlighted as a multidimensional perceptual construct in contrast to pitch, loudness, and voiced phonemes. Furthermore, pitch, loudness, and phonetic categories are monodimensional using a single objective correlate like fundamental frequency, intensity, and format frequency, respectively. Therefore, voice quality cannot be objectively quantified by a single measure or rating.³ Although these definitions underline the complexity and difficulty of the voice quality concept, a measurement based on perceptual evaluation still remains as the logical candidate for a gold-standard assessment. Voice quality is a perceptual phenomenon by nature,⁴ namely a behavioral response to a stimulus or stimulus acoustic features in the voice sound, presumably related to periodicity prominence.⁵

Clinicians or researchers prevalently judge voice quality by auditory-perceptual methods. These methods include the grade, roughness, breathiness, asthenia, strain scale recommended by the Japan Society of Logopedics and Phoniatrics,⁶ the Consensus Auditory-Perceptual Evaluation of Voice proposed by the

American Speech-Language-Hearing Association,⁷ or other psychophysical methods such as various forms of rating scales, magnitude estimation, and matching. Explanations for wide use might be the standardization, the simplicity, and the efficiency of these methods. Furthermore, auditory-perceptual methods enable clinicians to document the presence, degree, and progression of abnormal voice quality.

Because type and severity of voice quality differ substantially between continuous speech and sustained vowel phonations,^{8,9} implementation of both speech types in clinical rating protocols has become undisputed.⁷⁻¹⁰ Continuous speech approximates everyday conversation more, but its voice quality rating shows greater variability because perceptual focus diverges from nonvocal phenomena (eg, prosodic fluctuation, phonetic context, and phonological use of dysphonia).^{8,10} Sustained vowels, on the other hand, are less susceptible to such phonetic variability but lack ecological validity (ie, are less representative for daily voice use patterns). Therefore, the concept of concatenating both speech types has been advocated and shown to be promising in recent studies.^{8,9} Single auditory-perceptual judgments of both speech types together have shown a strong proportional relationship and no significant difference to *post hoc* averaging of the ratings of continuous speech and sustained phonation separately, and to a bivariate model weighting of the separate speech type ratings.⁸

A central concern in the evaluation of perceptual measurement is validity and reliability. Validity refers to what the measurement actually measures and how useful the measurement is.^{11,12} The validity of perceptual measurement depends on characteristics of the raters and the stimuli. Reliability is the overall consistency of a measure producing similar results under consistent conditions.

Validity and reliability are dedicated to the degree of absence from measurement errors in the judgment scores. First, random

Accepted for publication July 13, 2016.

Disclosure: This article has no actual or potential conflict of interest and was not financially supported.

From the *Faculty of Medicine and Health Sciences, University of Antwerp, Belgium; †Medical School, Hochschule Fresenius University of Applied Sciences, Hamburg, Germany; ‡European Institute for ORL, Sint-Augustinus Hospital, Antwerp, Belgium; and the §Faculty of Education, Health & Social Work, University College Ghent, Belgium.

Address correspondence and reprint requests to Ben Barsties, Universiteitsplein 1, 2610 WILRIJK, Antwerp, Belgium. E-mail: ben.barsties@t-online.be

Journal of Voice, Vol. 31, No. 2, pp. 202–210

0892-1997

© 2017 The Voice Foundation. Published by Elsevier Inc. All rights reserved.

<http://dx.doi.org/10.1016/j.jvoice.2016.07.006>

TABLE 1.
Overview From Barsties and De Bodt¹ of Influencing Factors of Rater Reliability Divided into Three Category Groups

Category	Factors Affecting Reliability	Reference
Listener	Panel size	Bele ¹¹
	Internal factors (eg, lapses in memory, attention, fatigue, and mistakes)	Kreiman et al, ¹³ Gerratt et al ¹⁴
	Judgment experience of the raters	Kreiman et al, ^{15,16} Pfützner and Barry, ¹⁷ Sofranko and Prosek, ¹⁸ Misono et al ¹⁹
	Professional background of the raters, for example, otolaryngologists, speech-language pathologists, singing voice teacher, or phonetic teacher	Sofranko and Prosek, ¹⁸ Pfützner and Barry ¹⁷
	Listener bias in the knowledge of medical diagnosis background of a voice	Eadie et al ²⁰
	Musical background of the raters	Eadie et al ²¹
	Listener agreement/reliability improved after training of native listeners, undergraduate speech-language pathologists, or graduate speech-language pathologists	Chan and Yiu, ^{22,23} Chan et al, ²⁴ Iwarsson and Reinhold Petersen, ²⁵ Eadie and Baylor ²⁶
Stimulus	Restrict recognition memory in the auditory modality	Bigelow and Poremba ²⁷
	More disagreement among slightly and moderate voice disorders than in normal voices or extreme cases of voice quality	Kreiman et al, ¹³ Gerratt et al, ¹⁴ Kreiman and Gerratt ²⁸
	Uses or disuse of anchor voices (ie, reference pattern)	Gerratt et al, ¹⁴ Chan and Yiu, ²² Awan and Lawson, ²⁹ Eadie and Kapsner-Smith ³⁰
Scale	Drift in ratings caused by perceptual context (eg, after hearing a number of slightly severe pathological voices, the rating for a moderate severe pathological voice becomes more severe through a shift of the internal standard from the listener)	Kreiman et al ¹⁶
	Visual analog or ordinal scale	Wuyts et al ³¹

errors affect the score of a person because of purely chance happenings (ie, guessing distraction in the testing situation, administration errors, content sampling, scoring errors, and fluctuations in the individual examinee's state). Therefore, random errors may reduce both the reliability and the validity and thus the utility of the measurement.¹¹ Second, systematic errors measuring some particular characteristics of the person or the test that have nothing to do with the construct (eg, a rater who always uses the scale in the same manner) may also affect the score of a person. Commonly, systematic errors do not show inconsistent results of the judgment but may lead to low validity of the measurement, which reduces the utility.¹¹ Both types of errors have an impact on the clinical applicability in perceptual judgment. However, the most critical property of any measurement is its validity. Reliability is a condition for validity that places an upper limit on the validity of a measurement. Thus, it is necessary to estimate reliability, but reliability is not a sufficient prerequisite for the measurement to have validity.¹¹

There are many factors related to scale, listener, and stimulus that significantly influence the reliability and accuracy of the clinical judgment of voice quality (Table 1). However, to our knowledge, the influence of voice sample length (VSL), which is a fundamental characteristic of the voice sample, on voice

quality ratings has not been investigated before for vowels, continuous speech, or the combination of both speech types.

The present study aimed to examine the length of continuous speech in the judgment of voice quality for the clinical practice in a combined paradigm considering continuous speech and sustained phonation. Our main research question addresses the following: how much information of various VSLs, which are commonly used in clinical practice under consideration of the implementation of both speech types, does a rater need to steadily judge overall voice quality with one average score? Based on these results, a first suggestion can be drawn for the clinical professionals who examine and treat voice patients by finding significant differences in the degree of severity of the different VSLs. Furthermore, we wanted to estimate the reliability of the overall voice quality judgments from the raters of the concatenated voice samples with different lengths, controlling partially the validity of the perceptual measurement.

MATERIALS AND METHODS

Subjects

One hundred subjects with a voice disorder were recruited from the Ear-Nose-Throat caseload of the Sint-Jan General Hospital

Download English Version:

<https://daneshyari.com/en/article/5124395>

Download Persian Version:

<https://daneshyari.com/article/5124395>

[Daneshyari.com](https://daneshyari.com)