Auris Nasus Larynx 42 (2015) 183–188

Contents lists available at ScienceDirect

Auris Nasus Larynx

journal homepage: www.elsevier.com/locate/anl

Assessment of voice quality: Current state-of-the-art

Ben Barsties [a,b,c,*], Marc De Bodt [b,d,e]

[a] Deutsche Stimmklinik, Hamburg, Germany
[b] Faculty of Medicine and Health Sciences, University of Antwerp, Belgium
[c] Medical School, Hochschule Fresenius University of Applied Sciences, Hamburg, Germany
[d] Department of Otorhinolaryngology and Head & Neck Surgery, University Hospital, Antwerp, Belgium
[e] Faculty of Medicine & Health Sciences, University of Ghent, Belgium

ARTICLE INFO

Article history:
Received 3 September 2014
Accepted 7 November 2014
Available online 28 November 2014

Keywords:
Voice quality
Voice evaluation
Voice assessment
Subjective approach
Objective approach

ABSTRACT

Voice quality is not clearly defined but it can be concluded that it is a multidimensional perceived construct. Therefore, there are broadly two approaches to measure voice quality: (1) subjective measurements to score a client's voice that reflects his or her judgment of the voice and (2) objective measurements by applying specific algorithm to quantify certain aspects of a correlate of vocal production. This paper proposes a collection and discusses a number of critical issues of the current state-of-the-art in voice quality assessments of auditory-perceptual judgment, objective-acoustic analysis and aerodynamic measurements in clinical practice and research that maybe helpful for clinicians and researchers.

1. What is voice quality?

Voice quality is a perceptual phenomenon in voice. In general, voice quality is not clearly defined in a lot of publications. Commonly, it is accepted that pitch, loudness and phonetic categories are not considered voice quality [1]. Thus, voice quality is a multidimensional construct and cannot be measured mono-dimensional such as pitch (i.e., measured in Hz) or loudness (i.e., measured in dB) [2]. This state makes it difficult to operationalize the concept [3].

Another dilemma in the definition of voice quality are the many different terms for describing voice quality, but only a few of these received wide acceptance like the major subtypes of dysphonic voice quality: 'breathiness', 'roughness', and 'strained' [4].

Further, there are physiological definitions of voice quality by considering the perceived result of the coordinated action from respiratory system, vocal folds and supra-glottal movement. These definitions are inadequate to specify the listeners' contributions quality that makes an essential part in defining what is after all a perceptual phenomenon [3]. The measured response of a voice is not necessarily constant across listeners and therefore physiologically based definitions cannot accommodate such effects [3].

The evaluation of voice quality may take one of two broad approaches. First, a subjective approach by listening to a client's voice and assigning a score that reflects his or her judgment of the voice. These are probably the oldest methods. Second, the use of a objective approach that applies specific algorithm to quantify certain aspects of a correlate of vocal production such as vocal acoustic signal, or the inverse-filtered oral airflow signal or its derivative [4].

2. Assessments of voice quality

In the past, different attempts are undertaken to create a consensus in voice assessment by consideration methods to evaluate voice production. These protocols contain five multidimensional aspects in voice assessments like visual analysis, perceptual evaluation, aerodynamic measures, acoustic analysis, and self-evaluation by the patient, e.g. description in the European Laryngological Society (i.e., ELS) Protocol provided by Dejonckere et al. [5] or Sri Ramachandra University Protocol from Boominathan et al. [6]. All these assessments are independent to each other and judge the impact of the voice disorder in a larger context especially at the evaluation of voice treatment. Therefore, no reduction of these five dimensions seems useful [7,8]. This status confirms the importance of using different types of assessments for the evaluation of voice quality. In the following paragraphs the different types of voice assessments are discussed which were recommended in the protocols of ELS and Sri Ramachandra. The

* Corresponding author at: Deutsche Stimmklinik, Hamburg, Germany. Tel.: +49 4051313007.
E-mail address: ben.barsties@t-online.de (B. Barsties).

http://dx.doi.org/10.1016/j.anl.2014.11.001
0385-8146/

aim of the collection is to create an overview of the current state-of-the-art and discuss a number of critical issues in these voice quality assessments.

3. Auditory-perceptual judgment

Auditory-perceptual judgment is a main part of routine clinical assessment of patients with voice disorders to document the voice quality because of its simplicity and efficiency. There are different possibilities to standardize the evaluation of auditory-perceptual judgment in the rating of voice quality. Internationally, some schemes or scales are specifically designed, such as the GRBAS scale or RBH scale, Consensus Auditory-Perceptual Evaluation of Voice (i.e., CAPE-V), Laver's Voice Profile Analysis, The Stockholm Voice Evaluation Consensus Model, Vocal Profile Analysis Scheme, Buffalo Voice Profile, and Hammarberg scheme [9,10].

Three rating schemes are the most frequently reported and most accepted rating schemes in the perceptual evaluation [5,11,12]. First, the GRBAS scale proposed by the Japan Society of Logopedics and Phoniatrics [13]. It is a five-dimensional scale by using four unidimensional parameters to focus the listener selectively on specific aspects in the voice which are linked to voice quality. These are labeled as *Roughness* (i.e., R), *Breathiness* (i.e., B), *Asthenia* (i.e., A) and *Strain* (i.e., S). The fifth parameter considers the overall severity in the impression of abnormality in the voice as labeled as *Grade* (i.e., G). All parameters are judged in a 4-point ordinal scale of 0 (normal) to 3 (severe).

Second, the RBH scale is used in German clinics [14] and is also recommended by the Committee on Phoniatrics of the European Laryngological Society [8]. In comparison with the GRBAS scale the judge can evaluate only three dimensions (i.e., roughness, breathiness, and hoarseness which is equal to G from the GRBAS scale) on a 4-point ordinal scale.

Third, the CAPE-V is proposed by the American Speech-Language and Hearing Association [15]. It enables in the standard analysis of the same parameters like mentioned in the GRBAS scale, except asthenia. Additionally, CAPE-V adopts a visual-analog scale and has predetermined vocal tasks and analysis criteria. Further, the evaluation of pitch, loudness and the addition of two more parameters (i.e., diplophonia, fry, falsetto, asthenia, aphonia, pitch instability, tremor, and wet/gurgly) are part of CAPE-V.

In the process of voice quality evaluation the choice of the stimuli plays an important role. Stimuli selected for these measurements include vowel(s) or running speech (i.e., syllables/words/sentences) [4]. The vocal behavior differs substantially between these two speech tasks potentially leading to perceived differences in type and severity of dysphonia [16]. Advantages of choosing running speech are providing a better approximation to everyday conversation than vowels alone, allowing evaluation of certain characteristics, the effect of co-articulation on voice quality may not be evident from vowel samples, and running speech shows greater variability in voice quality. Therefore running speech is preferred in the evaluation of voice quality [4]. The use of vowels in the evaluation of voice quality is meaningful to observe only few characteristics in the voice without focusing on non-vocal phenomena (e.g. prosodic fluctuation, phonetic context, and phonological use of dysphonia) or to establish a relationship between perceived quality and vocal fold physiology [4,16].

Otherwise, auditory-perceptual judgment has a large variation in both intra- and inter-rater reliability [17,18]. There are different reasons which influence the results of reliability and accuracy in perceptual evaluation. Thus, the results are disturbed in auditory-perceptual judgment without consideration of certain affecting factors. These factors can be selected in three categories (i.e., listener, stimulus, and scale). First, the listener category relates to the panel size [18], inherently unstable internal factors of the internal standard by listeners (e.g. lapses in memory, attention, fatigue, and mistakes) [17,19], the restrict recognition memory in the auditory modality [20], the judgment experience decrease [21,22] or increase [23–25] reliability, the professional background (e.g. otolaryngologists, speech-language-pathologists, singing voice teachers, and phonetic teachers) [23,25], listener bias in the knowledge of medical diagnosis background [26], musical background [27], and auditory-perceptual judgment training (e.g. native listeners [28–30], undergraduate speech-language-pathologists [31], or graduate speech-language pathologists [32]).

Second, the stimulus category covers aspects like more disagreement among slightly and moderate voice disorders than in normal voices or extreme cases of voice disorders [17,19,33], use of anchor stimuli (i.e., reference pattern) while rating [19,28,34,35], and the drift in ratings caused by perceptual context (e.g. 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) [22].

Third, the scale category includes the factor of differences in reliability using visual-analog or ordinal scale [36].

4. Objective-acoustic analysis

Objective-acoustic analysis of voice signals is the most used diagnostic instrument to identify voice disorders in research [37].

Voice signals can be acoustically analyzed in time, frequency, amplitude, and quefrency domain. An elaborated collection of diverse methods based on these domains is described and discussed by Buder [38].

Traditionally, acoustic methods are applied to yield objective data on only sustained vowels. Analysis based on one acoustic method (e.g. Jitter, Shimmer, etc.) on sustained vowels revealed in different studies poor reliability or poor documentation of improvement in voice quality [8,39]. Most of them also show low or poor correlation to auditory-perceptual judgment [40], that raises questions regarding the validity and usefulness of these acoustic determinants [41]. However, a combination of several parameters in a model shows higher reliability and validity with e.g. correlates to auditory-perceptual judgment [42–46].

The evaluation of voice quality on sustained vowels does not necessarily correspond with running speech and might be an exceeding limitation of most of the reported studies which use acoustic methods. Thus, it encompasses drawbacks in the objective-acoustic analysis [42,47]. However, objective-acoustic parameters can be scarcely used in direct analysis of running speech, i.e., spectral, and cepstral methods [48,49]. The commonly used application of 'classically' objective-acoustic parameters (e.g. Jitter, Shimmer, Harmonics-to-noise analysis, etc.) has first to be filtered in voice segments and non-voice segments before they are eligible for analysis in running speech.

To our knowledge, there are only two multiparametric models to evaluate voice quality successfully on sustained vowels and on running speech, called Acoustic Voice Quality Index (i.e., AVQI) proposed by Maryn et al. [47], and Cepstral Spectral Index of Dysphonia (i.e., CSID) published by Awan et al. [48] and Awan et al. [49]. In further investigations, both models could confirm accuracy as well as reliability in detecting voice abnormality: *AVQI* [50–54] and *CSID* [55–60].

The use of objective-acoustic analysis in research or clinical practice is dependent on different conditions to enable voice analysis on a high level of accuracy and reliability, as well. These conditions cover three categories of *Hardware*, *Software*, and *Examination Circumstances* and are described in the following paragraphs.