Acoustic Parameters for Classification of Breathiness in Continuous Speech According to the GRBAS Scale

*Adam Stráník, *Roman Čmejla, and †Jan Vokřál, *†Prague, Czech Republic

Summary: Objectives. The purpose of this study was to classify objectively breathiness in continuous speech according to a subjective evaluation of voice based on the GRBAS scale.

Study Design. A retrospective, experimental study.

Methods. A total of 593 records with read text were twice evaluated by five experts according to the GRBAS scale within two sessions with a time delay of at least 2 weeks. The records were subsequently subjected to acoustic analysis using parameters which do not rely on the accurate estimation of fundamental frequency: Glottal-to-Noise Excitation ratio, Cepstral Peak Prominence Pearson *r* at autocorrelation peak, Breathiness Index, and the ratio of high- to mid/low-frequency energy. These parameters were subsequently analyzed and a total of 92 features were created for each record. After feature space reduction based on Correlation Feature Selection and Information Gain, the feature space was reduced to four parameters. These four parameters were used for classification of breathiness.

Results. In the final set of four, the acoustic parameters have significantly different mean ranks in every grade of breathiness according to the GRBAS scale (Kruskal-Wallis test [$P < 0.001$]). The accuracy of classifier for objective evaluation of level of breathiness based on the discrete scale of breathiness reached 77%. Assuming continuous grades of breathiness, the classifier reached $\rho = 0.92$ ($P < 0.001$).

Conclusions. The level of breathiness in continuous speech can be effectively described by automatic system–based analysis of acoustic measures. The proposed automatic system is able to determine the level of breathiness in continuous speech with sufficient precision.

Key Words: Dysphonic voice–Breathiness–GRBAS scale–GNE–HLR–BRI–RPK–CPP–Classification–Continuous speech.

INTRODUCTION

The human voice is the main instrument of communication and is used daily for a wide range of reasons. Dysphonic voices can cause negative attitudes in listeners, and it is, therefore, important to have reliable tools for their assessment and therapy.[1] The basic tools for voice assessment are subjective ones because every listener performs a subjective classification.

Subjective voice assessments are based on the perceptual quantization of basic voice markers, such as roughness and breathiness. There are several procedures that describe how to perform a subjective evaluation of hoarseness by professionals, for example, the recommendations by the Union of European Phoneticians,[2] GRBAS scale,[3] or Consensus Auditory Perceptual Evaluation of Voice.[4] There are also several patient-based methods such as the Voice Handicap Index[5] and Voice-Related Quality of Life.[6]

A breathy voice is a phonation in which the vocal folds vibrate, as they do in normal voicing, but are held further apart so that a larger volume of air escapes between them.[7] Acoustic measures have been developed to help quantify voice characteristics. The advantage of having such measures is that, if reliable and reproducible, they can be used as a means of following changes in voice over time and in comparing the efficacy of treatment regimens aimed at improving dysphonia.[8]

One of the most common tools for subjective voice assessment is the GRBAS scale, which is used by many researchers and in clinical practice.[9–13] The GRBAS scale uses a description of voice in five parameters: overall voice performance (G—grade); roughness (R), which is mainly related to vocal folds oscillation irregularity; breathiness (B), which summarizes the amount of additive noise in voiced parts of speech; and asthenicity (A) and strain (S), which are difficult to define. Every parameter can assume values from 0 to 3, where 0 means a normal voice or a voice without pathology and 3 means the most dysphonic voice or a voice where the pathology is most present. There are several studies, which investigate the reliability of the GRBAS scale, and the results are unconvincing: De Bodt et al[14] and Kranell et al[15] concluded that the parameter *G* is the most reliable, parameters *A* and *S* are the least reliable, and Wuyts et al[16] found parameter *A* to be the second most reliable.

There are several works focused on the objective classification of overall voice quality (G) based on a subjective evaluation according to the above-mentioned GRBAS scale.[9–13] Yet only one work takes breathiness B into account.[10]

There is no consensus regarding whether to use sustained phonation or continuous speech. The general opinion is the following: whereas sustained phonation is much easier to analyze, hoarseness is much more audible in continuous speech. This is mainly due to the fact that during continuous speech, the vocal folds have to start and stop vibrating repeatedly and the vocal tract is more burdened than solely during the sustained phonation. However, continuous speech is more common in daily communication and it is, therefore,

Accepted for publication July 22, 2013.

From the *Department of Circuit Theory, Faculty of Electrical Engineering, Czech Technical University in Prague, Prague, Czech Republic; and the †Department of Phoniatrics, First Faculty of Medicine, Charles University in Prague and General University Hospital in Prague, Prague, Czech Republic.

Address correspondence and reprint requests to Adam Stráník, Department of Circuit Theory, Faculty of Electrical Engineering, Czech Technical University in Prague, Technická 2, Prague 6, 16627 Czech Republic E-mail: stranik.adam@gmail.com

Journal of Voice, Vol. 28, No. 5, pp. 653.e9-653.e17
0892-1997/$36.00

http://dx.doi.org/10.1016/j.jvoice.2013.07.016

more of an effort to analyze it. This statement is in agreement with the latest trends in the automatic analysis of hoarseness.

In severely breathy voices, there is either no fundamental frequency (aphonic voice, $B = 3$) or the fundamental frequency is severely blurred by additive noise and accurate estimation is, therefore, almost impossible.[7] For this reason, only the acoustic parameters that do not rely on the accurate estimation of fundamental frequency will be taken into account for subsequent analyses.

Michaelis et al[17] introduced the Glottal-to-Noise Excitation ratio (GNE). GNE is based on determining the maximum cross-correlation between Hilbert envelopes of band-pass–filtered signals generated by the inverse filtration of a speech signal. A complete guide and workflow are described in Michaelis et al[17] and Godino-Llorente et al.[18] The idea of GNE can be summarized as follows: glottal closure in a normal speech signal will, after inverse filtration, excite the whole spectrum by a "Dirac impulse," and therefore, the Hilbert envelopes should have almost the same shape in all frequency bands, and thus Hilbert envelopes in different frequency bands will attain high cross-correlation coefficients. As pathologic, or to be more precise breathy, voices are less excited by vocal folds, the cross-correlation coefficients of the Hilbert envelopes in different frequency bands are lower. The significant advantage of GNE estimation lies in the fact that the fundamental frequency is not estimated.

Hillenbrand et al[19] proposed several parameters for breathy voice assessment: Cepstral Peak Prominence (CPP), Pearson *r* at autocorrelation peak (RPK), Breathiness Index (BRI), ratio of high- to mid/low-frequency energy (HLR), and the first harmonic amplitude.

CPP is commonly used for the examination of voice quality for both sustained vowels and continuous speech.[8,19–24] The idea of CPP is as follows: although the first several cepstral coefficients correspond with the transfer function of the vocal tract, the remaining cepstral coefficients correspond with the excitation of the vocal folds. This excitation appears as regularly spaced peaks in the quefrency domain. The position of the first peak corresponds to the fundamental frequency of the voice. CPP is the distance between the amplitude of this peak and the regression line of the underlying cepstrum. Whereas a healthy voice has a well-defined harmonic structure with a strong cepstral peak, a breathy pathologic voice does not have such a well-defined harmonic structure and the cepstral peak tends to be smaller. For the cepstral peak, the maximal value in the cepstrum above 1 millisecond (ie, below 1 kHz) is used.[19]

RPK is defined as a maximum in the autocorrelation function between 3.3 and 16.7 milliseconds (ie, between 60 and 300 Hz).[19] Voices with a well-defined harmonic structure tend to reach higher values than voices with additive noise.

BRI is defined as the ratio between the energy in the second derivative of a signal to the energy in a nonderived signal.[19] Larger values of BRI mean more energy at higher frequencies; hence, it is expected that breathy voices can reach higher values of BRI.

HLR measures the average spectral energy at or above 4 kHz and below 4 kHz. The idea of this parameter is clear and similar to the BRI: breathy voices have more energy at higher frequencies.[19]

Breathiness is audible in speech, and therefore, the severity of breathiness can be subjectively rated into several levels according to the GRBAS scale. It is hypothesized that acoustic parameters will differentiate the level of breathiness in continuous speech the same way as experts do during subjective rating. It is also hypothesized that detailed description of time behaviors of parameters will bring useful information for classification of breathiness into levels according to the severity. Having reliable acoustic parameters and their time description, an automated system for automatic classification of breathiness in voice will be proposed. Such an automated system will help in clinical practice.

The aim of this experiment was to select existing acoustic parameters and their time behaviors, which can differentiate an acoustic signal with continuous speech according to the breathiness expressed by parameter *B* from the GRBAS scale with the best precision. Therefore, the above-mentioned acoustic parameters associated with the measurement or quantification of additive noise in speech signals were analyzed in this study. The time behaviors of these parameters were subjected to subsequent analyses. After the selection of the best set of parameters, an attempt to design a classifier capable of classification between various levels of breathiness was performed.

METHODS

Voice samples

A database recorded by the Department of Phoniatrics, First Faculty of Medicine, Charles University in Prague and General Faculty Hospital in Prague was used for this work. This database was recorded from the 1970s to the 1990s as a common element of the voice examination of healthy and pathologic voices. All records were performed using professional recording equipment and were recorded in a soundproof booth with a level of ambient noise lower than 18 dB SPL. The database was originally recorded on tapes and then digitalized with a sampling frequency 44.1 kHz and 16 bit resolution.

A total of 593 records containing readings of a standard phonetically unbalanced text were selected from the database. The text is 34 words long and was evaluated by means of the GRBAS scale by five experts from the Department of Phoniatrics. Each expert made two assessments of each recording with a delay of at least 2 weeks between assessments. During the subjective assessment, the records were identified by a random ID, and the IDs were different for the second assessment.

The final grades according to the GRBAS scale were determined in two ways (a) using the modus from the final 10-element (five raters, two sessions) set of grades for discrete classification and (b) using the mean value from the final 10-element set of grades for continuous classification. Only the breathiness B from the whole GRBAS scale was taken into account for this experiment. The number of records for every grade of B are shown in Figure 1.