INTERSPEECH 2021
30 August – 3 September, 2021, Brno, Czechia

Modeling Dysphonia Severity as a Function of Roughness and Breathiness Ratings in the GRBAS Scale

Carlos A. Ferrer[1,2], *Efren Aragón*[1], *María E. Hdez-Díaz*[3], *Marc S. de Bodt*[3], *Roman Cmejla*[4], *Marina Englert*[5], *Mara Behlau*[5], *Elmar Nöth*[2]

[1]Informatics Research Centre, Central University "Marta Abreu" of Las Villas, Cuba
[2]Friedrich-Alexander-University Erlangen-Nuremberg, Germany
[3]Faculty of Medicine and Health Sciences, Antwerp University, Belgium
[4] Faculty of Electrical Engineering, Czech Technical University in Prague, Czech Republic
[5]Departamento de Fonoaudiologia, Universidade Federal de São Paulo, Brazil
cferrer@uclv.edu.cu, earagon@uclv.cu, mariae@gmail.com, mdbodt@uza.be, cmejla@fel.cvut.cz, marinaenglert@gmail.com, mbehlau@cevbr.com, elmar.noeth@fau.de

Abstract

Dysphonia comprises many perceptually deviating aspects of voice, and its overall severity perception is made by the listener according to methods of aggregating the single dimensions which are personally conceived and not well studied. Roughness and breathiness are constituent dimensions in most devised rating scales in clinical use. In this paper, we evaluate several ways to model the mapping of the overall severity as a function of the particular ratings of roughness and breathiness. The models include the simple linear averaging as well as several non-linear variants suggested elsewhere, and some minor adjustments. The models are evaluated on four datasets from different countries, allowing a more global evaluation of how the mapping is conceived.

Results show the limitations of the most widely assumed linear approach, while also hinting at a need for a more uniform coverage of the sample space in voice pathology datasets. The models explored in this paper can be expanded to higher-dimensional scales.

Index Terms: dysphonia, vocal quality, perceptual scales, roughness, breathiness.

1. Introduction

Dysphonia is a clinical term defined by the European Laryngological Society to comprise "any kind of perceived voice pathology: the deviation may concern pitch or loudness, as well as timbre or rhythmic and prosodic features" [1], [2]. The prevalence and impairing effects of dysphonia have prompted recommendations for its management and the procedures to search for the underlying diagnosis [3]–[5].

Efforts to quantify dysphonia severity can be broadly separated into perceptual or instrumental/objective.

1.1. The perceptual approach

Rating scales have been devised comprising a certain amount of dimensions, to be rated on a certain range of values.

1.1.1. Rating Scales

As a result from an effort to reduce the number of terms used in the perceptual ratings, the GRBAS (Grade, Roughness, Breathiness, Asthenicity, Strain) scale was devised in Japan in the 70's [6], with ratings from 0 (normal) to 3 (severe) on each dimension. The GRBAS scale is the core of a family of scales originating later. Reducing the number of dimensions, the German RBH [7] (where H can be roughly approximated to G in GRBAS), and the European Laryngological Society recommended the use of GRBAS or its reduced form GRB [1], [2]. As increments, the addition of I (instability) to form GRBASI [8], translated RASATI in Brazil [9], or H (Harshness) to form GRBASH [10]. The CAPE-V protocol developed in the U.S.A. established the rating of GRBSPL(P: Pitch, L: Loudness) as well as a possible couple of other clinician selected dimensions. Overall, G, R & B comprise the invariable core of most Dysphonia rating scales.

1.2. The Instrumental/Objective Alternative

Some of these dimensions are linked to physiological causes and specific acoustical cues, like R and B., prompting to the development of methods aiming to objectively appraise the purported acoustic cues [11]. Some objective methods measure a particular aspect of the voice process, and then relations are tried to be established with a perceptual dimension [12]. Other works perform manipulations on voice synthesizer's parameters, and then relate the perceived alterations to a now strictly defined change in the voice [13].

1.3. The measurement of Overall Severity

In spite of the efforts in the objective measurements of Dysphonia, the perceptual approach is still favored [14], [15]. The G in the perceptual scales stands for the overall severity, and as such can't be linked to a particular objective cue. By itself, G is a composite of its constituent dimension, and not individually definable.

1.3.1. Mapping as a Perceptual Composite

Since G is the "overall impression of voice deviance" [16], its value is somehow influenced by the perception of the individual dimensions of the particular scale. An expression relating G to its perceptual components is rarely given, e.g. in [17], it was suggested that G would be a Scalar Distance. However, it is most frequently assumed as a linear combination, either implicitly or explicitly [18].

http://dx.doi.org/10.21437/Interspeech.2021-1540

1.3.2. As an Acoustic Index / Composite

Even if G is not instrumentally definable, it must be acknowledged that it is the most successfully measured perception so far. The best single acoustic index, the Cepstral Peak Prominence, has been included as the only acoustic measure recommended in [19] for voice assessment, precisely as "a general measure of Dysphonia" [20]. Expressing G as a multiple linear regressions of acoustic indexes has also been the most successful, like AVQI [21] or CSID [22], recently followed by ABI, an index of B [11], [23].

1.4. Objectives of this paper

The actual way the individual perceptual dimensions map into the overall G is still unknown, and different mappings have never been compared. In this paper we explore the fit of several models of G as a function of only R and B to allow the use of a higher number of datasets. The effect of excluding some dimensions could hinder the detection of idiosyncratic rating aspects developed by raters when using particular derivations of the GRBAS scale, and also increase the unexplained variance of the models. Determining the actual perceptual mapping can have implications beyond the understanding of how judgments are made: the mappings obtained could be somehow translated to the case of using objective cues.

2. Materials and Methods

2.1. Models

In this paper we evaluate six model of the mapping of G in the (R,B) plane. Their grayscale representations are shown in Figure 1, from black for G=0 to white for G=3.

2.1.1. Linear

The simplest, straightforward method to express G as a composite of its constituent dimensions is a weighted sum of them. This linear model is implicitly assumed on the objective approach to measure G, where it is expressed a linear regression of acoustic correlates. The most general form of the linear model would be:

$$G_{Lin}\left(R,B\right)=k_R R+k_B B \quad (1)$$

where the k_R and k_B terms represent the possibly different weights perceptually given to each dimension. In this paper, we will assume R & B equally relevant, with both weights equal to ½, so that G cannot exceed 3, the maximum value of the scale. Whether k_R and k_B are actually different, and the estimation of their relative strengths might be the subject of future studies. The distribution of G_{Lin} values in the (R,B) plane is shown in the top-left graph of Figure 1, as intensities on a grayscale.

A useful concept we will use in discriminating the different models in this G(R,B) graph is the shape of the line joining points with same values of G (henceforth *iso-G* lines). For G_{Lin}, *iso-G* lines are straight lines with negative unity slope. The G_{Lin} model exhibits a huge contradiction which can be illustrated with the predicted values of G on the corners of the graph excluding the origin. We will refer to this as **the corner's problem**. It consists in the impossibility to rate as maximally distorted (G=3) a voice which is maximally distorted in only one of its particular dimensions (R or B), i.e., the (0,3) or (3,0) corners of the (R,B) plane. Conversely stated, a voice can only be perceived as maximally distorted if all its constituent dimensions are maximally distorted themselves: the (3,3) corner in the (R,B) plane. However, inspecting any available dataset of ratings would refute both statements too frequently to be sustained, i.e. maximum ratings of G are frequent on any of the three corners. In spite of G_{Lin} being straightforward and frequently assumed, the modeling of G(R,B) must be non-linear in order to solve (or mitigate) the corner's problem. Note than in G_{Lin} G(3,3) is twice as large as in the other corners.

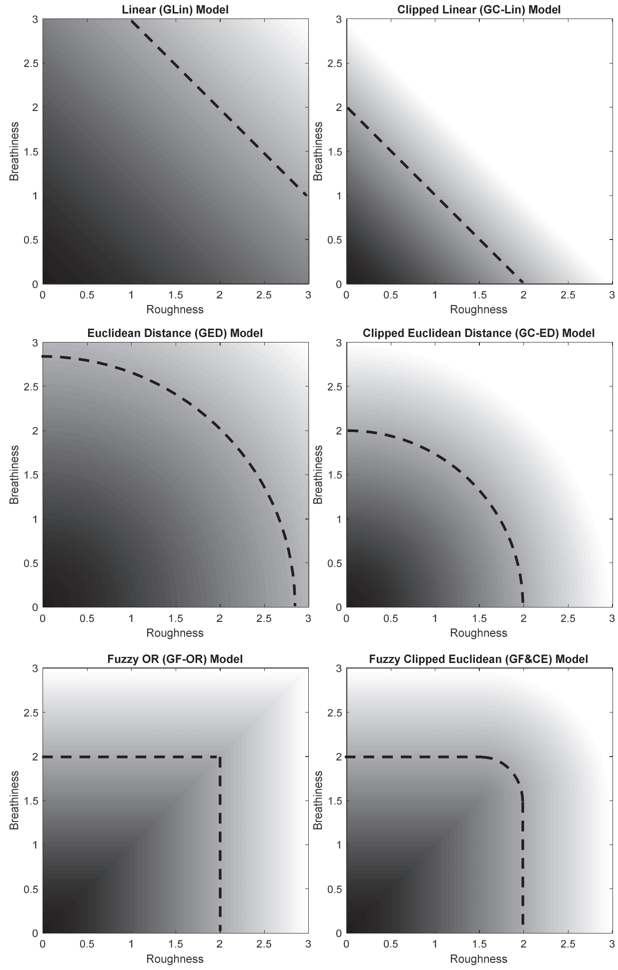

Figure 1: *Grayscale representation of the six models of G(R,B) evaluated. Top-Left: Linear (G_{Lin}), Top-Right: Clipped Linear ($G_{C\text{-}Lin}$), Middle-Left: Euclidean Distance (G_{ED}), Middle-Right: Clipped Euclidean Distance ($G_{C\text{-}ED}$), Bottom-Left: Fuzzy-OR ($G_{F\text{-}OR}$), Bottom-Right: Fuzzy + Clipped Euclidean ($G_{F\text{-}ED}$). Dashed curves: iso-G curve with G=2 per model.*

2.1.2. Clipping of Linear model

A simple modification to the G_{Lin} model which removes the corner's problem is to make both weights on equation (1) equal to unity, and limit (clip) all G values to 3. The values of G for this model (denoted $G_{C\text{-}Lin}$) are represented in the Top-Right graph in Figure 1.

A visual inspection of the *iso-G* lines for G=2 depicted in the graphs for both G_{Lin} and $G_{C\text{-}Lin}$ reveals unsatisfactory behaviors. For G_{Lin}, this value near the main diagonal (around R=2 & B=2) seems justified (there is room for more deviance in this diagonal direction). However, its *iso-G* don't even cross the R and B axis, leaving values below G=2 on both corners. For $G_{C\text{-}Lin}$, the opposite occurs: seemingly justified values in the crossings of its *iso-G* with the axis, but too much room left in the direction of the main diagonal. It is apparent that *iso-G*s should somehow go further in the main diagonal direction than the straight-line-shaped *iso-G*s of these linear based models.

2.1.3. *Euclidean*

In [17], it is proposed to conceive G as the scalar distance from the origin to the point on a 3-D metric space defined by the axis R, B, and A-S. Considering A and S as opposite directions (Hypo-Hyper) of a single dimension (Tonus) is not uncommon [17], [24], [25], however, under this assumption A and S could not be simultaneously different from zero. It is not hard to find examples countering this requirement on several datasets of ratings, including some samples provided by Hirano [6], although this is not a concern in this paper, dealing only with the R & B dimensions. The third model considered in this paper, G_{ED}, follows this definition of G as the Euclidean distance from the origin to the particular (R,B) point as described by equation (2), and is represented in the Mid-Left graph of Figure 1.

$$G_{ED}(R,B) = \sqrt{\left(R^2 + B^2\right)/2} \tag{2}$$

This G_{ED} model also suffers from the corner problem, although less pronounced than G_{Lin}: G(3,3)=3 is only 1.4 times larger than the other two corners. The *iso-G* curves in this model are arcs of circles with center in the origin, which seem as an improvement over G_{Lin} and $G_{C\text{-}Lin}$.

2.1.4. *Clipping of Euclidean Distance*

The 4th model evaluated tries to remove the corner problem from G_{ED} by making the values in the (0,3) and (3,0) corners equal to 3. It can be achieved by removing the division by 2 in equation (2), and clipping values larger to 3. The model is shown in the Mid-Right graph in Figure 1, denoted $G_{C\text{-}ED}$.

2.1.5. *Fuzzy OR*

In the German RBH scale, it is customary to enforce H to be higher or equal to the largest of R and B, to capture the fact that H is a composite (superclass) of both constituent dimensions [26]. From this space (G ≥ *max(R,B)*) we have chosen to evaluate the lowest value boundary, corresponding to the equality, as another model. Since *max(x,y)* is the most common way to define the OR operator in Fuzzy logic, we have named this model Fuzzy OR, and denoted it $G_{F\text{-}OR}$.

This model does not suffer from the corner problem, and the *iso-G* curves are straight lines parallel to the axis, intersecting at the main diagonal, as shown on the bottom-left graph in Figure 1.

2.1.6. *Fuzzy OR + Euclidean + Clipping*

The Goettingen Hoarseness Diagram [27] portrays the speaker's voice on a 2D representation broadly regarded as an 'objective' (R,B) plane. We are aware of a couple of commercially available software providing the users with the Goettingen Hoarseness Diagram (also known as Phonatory Deviation Diagram): LingWaves and VoxMetria. Both software provide an indication of the "Normal" voice region limited by an *iso-G* curve which is a mix of the straight lines from the $G_{F\text{-}OR}$ and the arcs from G_{ED} or $G_{C\text{-}ED}$ (see Figure 2).

We have included here a model mixing both *iso-G* curves, denoted G_{F+CE}, by evaluating equation (3):

$$G_{F+CE}(R,B) = \begin{cases} \max(R,B) & \text{if (R < th) or (B < th)} \\ \sqrt{(R-th)^2 + (B-th)^2} + th & \text{otherwise} \end{cases} \tag{3}$$

and then clipping values above 3. We have used *th*=1.5 to locate the change of behavior in the middle of the (R,B) plane.

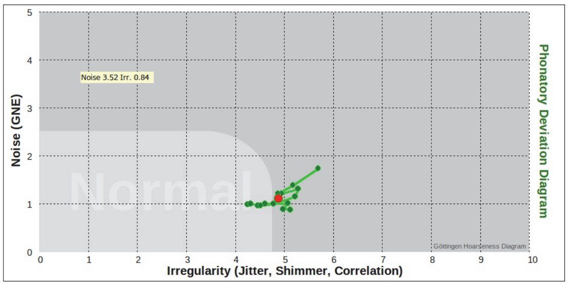

Figure 2: *Phonatory Deviation Diagram in VoxMetria.*

2.2. Datasets

The datasets considered were selected to be representative of specialists following slightly different variants of the GRBAS rating. They are described in the following subsections, in alphabetical order according to the 3-character country code. The 4x4 bin, 2D-Histograms of the four datasets are shown in Figure 3.

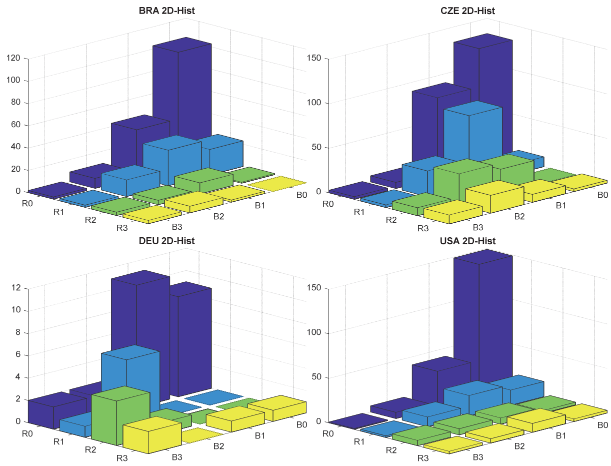

Figure 3: *4x4 histograms of the R & B ratings in the four datasets: Top-Left: Brazil (BRA), Top-Right: Czech Republic (CZE), Bottom-Left: Germany (DEU), Bottom-Right: USA (USA)*

2.2.1. *Brazil (BRA)*

This dataset consists of ratings of 258 voices (160 dysphonic, 98 vocally healthy) [28] by 9 judges in the GRBASI scale. Following the recommendations in [29], we show the 2-D Histogram of the average ratings on the (R,B) plane in the Top-Left pane of Figure 3. Bins have been denoted 0, 1, 2 or 3 for both R & B, although actual boundaries were set at [0, 0.75, 1.5, 2.25, 3].

A 4x4 bin histogram has been chosen since the dataset could warrant more than 10 samples on each bin. It can be seen, however, that most ratings fall in the ($R_{0\text{-}1}$,$B_{0\text{-}1}$) region of the histogram, and there is even a corner bin (R3,B0) which is unpopulated. The high number of healthy subjects (more than a third) certainly influences this distribution.

2.2.2. Czech Republic (CZE)

This dataset [25] comprises ratings by 5 judges, twice, of 469 voices comprising both healthy and pathologic subjects, in the GRBAS scale with A-S as a single Tonus dimension. The corresponding 4x4 histogram is shown in the Top-Right graph in Figure 3. Similar to the BRA dataset, here the samples cluster within the ($R_{0\text{-}1}$,$B_{0\text{-}1}$) bins, and there is an unpopulated bin (R2,B0). There are comparatively more samples towards the middle region.

2.2.3. Germany (DEU)

This dataset of ratings was extracted from the accompanying booklet to the two-CD RBH scale training samples from Nawka & Anders [30]. The sample recordings are limited to 40, with 10 for each value of H (i.e. G). This limited number of samples makes the 4x4 2D Histogram prone to the occurrence of zeros (5 from the 16 bins) as shown in Bottom-Left graph in Figure 3. We used the experts' (up to 7) average ratings. However being of smaller size, the DEU dataset, compiled to be illustrative of voice types, shows a smaller disproportion between the number of normal sounding samples and the ones in the corners.

2.2.4. United States of America (USA)

This dataset [31] contains ratings for 296 samples in both GRBAS and CAPE-V scales, performed by groups of 4-5 listeners. The 4x4 histogram is shown in the Bottom-Right graph of Figure 3. Here again there is large imbalance, with the (R_0,B_0) bin containing more than 50% of the samples.

2.3. Evaluation procedure

On every dataset, the Pearson's correlation coefficient r between actual G and the one predicted by each G(R,B) model according to the given ratings was calculated. The adequacy of the model is evaluated by the percentage of variance of the actual G explained (r^2).

3. Results and Discussion

The percent of variance explained by the different models on each dataset are shown in Figure 4. An additional dataset was made by adding all samples together, and denoted "All", to start analyzing the global trend of the models and then focus on specific behaviors on a given dataset.

A look at the total dataset shows that the adequacy of the four Euclidean and Fuzzy models is quite high (above 90% of variance explained) and almost indistinguishable, a fact confirmed by inspecting the results on individual datasets. The performances of both linear-based models is always worse than these four models, presumably due to the described inadequacy of their *iso-G* lines. The clipped version, intended to correct the corner problem, results in the worst performance in all cases.

Regarding the almost identical performance of G_{ED}, $G_{C\text{-}ED}$, $G_{F\text{-}OR}$ and G_{F+CE}, the origin of this result lies in the coverage of the (R,B) plane present in the datasets: to separate those models, it is necessary to have a larger population of the corners, which is where most models (not only these four) differ. Problems in the sample coverage have already been pointed out [29] as the possible reason behind non-conclusive results in Voice-Quality research.

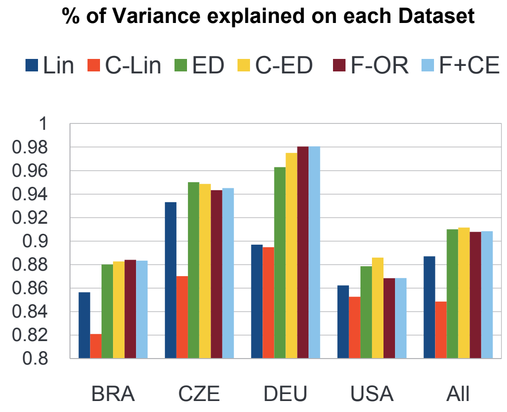

Figure 4: *Percent of variance explained by each model.*

With respect to the differences across datasets, DEU seems to contain the most suited ratings according to the extremely high percentage of variance explained (up to 98%) and the lower relative performance of G_{Lin}, expected due to its corner problem. The DEU dataset is the smallest, but its samples were chosen due to representativeness, unlike the other datasets more influenced by rates of appearances in clinical setting. As mentioned on its description, the DEU dataset shows a smaller disproportion between the number of samples in the (R_0,B_0) bin and the other 3 corners. Another influencing factor could be the presence of idiosyncratic patterns of rating for judges in the RBH scale, who might ignore the presence of other deviations (e.g. A or S) in the formation of the overall rating. For judges who actually consider more than R & B into their mapping of G, the fit of the evaluated models would certainly be worse ('noisier').

4. Conclusions

This paper has quantitatively evaluated several models for expressing G(R,B). Both the description of the models and their faults, as well as the results, points towards nonlinear combinations of R&B in the perceptual formation of the judgment of G. The non-linear models achieve more than 90% of variance explained overall, across all datasets.

In future works, the models considered could be optimized in terms of adjustable parameters. An obvious option would be to evaluate asymmetric models (i.e. different weights for R and B). However, it seems more important to first improve the coverage of the datasets available in terms of the number of samples with extreme values of R and B, in order to populate the corners of the (R,B) plane. Additionally, all models can (and should) be tested including dimensions other than the already considered B & R, in case the datasets contain them. Idiosyncratic judgment patterns could be explored in these studies.

5. Acknowledgements

This paper was funded in part by a Fellowship granted to the first author by the Alexander von Humboldt Foundation (Ref 3.2 - 1164728 - CUB - GF-E).

6. References

[1] P. H. Dejonckere *et al.*, “A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques: Guideline elaborated by the Committee on Phoniatrics of the European Laryngolo,” *Eur. Arch. Oto-Rhino-Laryngology*, vol. 258, no. 2, pp. 77–82, 2001.

[2] P. H. Dejonckere, “Assessment of Voice and Respiratory Function,” in *Surgery of Larynx and Trachea*, M. Remacle and E. H. Eckel, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 11–26.

[3] J. I. Chang, S. E. Bevans, and S. R. Schwartz, “Otolaryngology Clinic of North America: Evidence-Based Practice. Management of Hoarseness/Dysphonia,” *Otolaryngol. Clin. North Am.*, vol. 45, no. 5, pp. 1109–1126, 2012.

[4] T. Mau, “Diagnostic evaluation and management of hoarseness,” *Med. Clin. North Am.*, vol. 94, no. 5, pp. 945–960, 2010.

[5] S. R. Schwartz *et al.*, “Clinical practice guideline: Hoarseness (Dysphonia),” *Otolaryngol. - Head Neck Surg.*, vol. 141, no. 3 SUPPL, pp. S1–S31, 2009.

[6] M. Hirano, *Clinical examination of voice*, 1st ed. Vienna: Springer, 1981.

[7] T. Nawka, L. C. Anders, and J. Wendler, “Die auditive Beurteilung heiserer Stimmen nach dem RBH-System,” *Sprache, Stimme, Gehör*, vol. 18, pp. 130–133, 1994.

[8] P. H. Dejonckere, M. Remacle, E. Fresnel-Elbaz, V. Woisard, L. Crevier-Buchman, and B. Millet, “Differentiated perceptual evaluation of pathological voice quality: reliability and correlations with acoustic measurements,” *Rev. Laryngol.*, vol. 117, no. 3, pp. 219–224, 1996.

[9] E. J. Yamauchi, S. Imaizumi, H. Maruyama, and T. Haji, “Perceptual evaluation of pathological voice quality: a comparative analysis between the RASATI and GRBASI scales.,” *Logop. Phoniatr. Vocology*, vol. 35, no. 3, pp. 121–8, 2010.

[10] Ricardo Cardoso, R. F. Meneses, Lumini-Oliveira, J. P. Pestana, and B. Guimaraes, “Associations between Teachers’ Posture, Muscle Tension and Voice Complaints,” *J. Voice*, vol. In Press, 2020.

[11] B. Barsties v. Latoszek, Y. Maryn, E. Gerrits, and M. S. De Bodt, “The Acoustic Breathiness Index (ABI): A Multivariate Acoustic Model for Breathiness,” *J. Voice*, no. in press, 2017.

[12] E. H. Buder, “Acoustic analysis of voice quality: A tabulation of algorithms 1902–1990,” in *Voice quality measurement*, R. D. Kent and M. J. Ball, Eds. San Diego: Singular, 2000, pp. 119–244.

[13] B. R. Gerratt and J. Kreiman, “Measuring vocal quality with speech synthesis.,” *J. Acoust. Soc. Am.*, vol. 110, no. December 2000, pp. 2560–2566, 2001.

[14] A. Behrman, “Common practices of voice therapists in the evaluation of patients,” *J. Voice*, vol. 19, no. 3, pp. 454–469, 2005.

[15] E. S. Segundo, “International survey on voice quality: Forensic practitioners versus voice therapists,” *J. Exp. Phonetics*, vol. XXIX, no. February, pp. 8–34, 2021.

[16] B. Millet and P. H. Dejonckere, “What determines the differences in perceptual rating of dysphonia between experienced raters?,” *Folia Phoniatr. Logop.*, vol. 50, no. 6, pp. 305–310, 1998.

[17] J. F. Bonastre *et al.*, “Complementary approaches for voice disorder assessment,” *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*, vol. 2, pp. 1193–1196, 2007.

[18] T. Bhuta, L. Patrick, and J. D. Garnett, “Perceptual evaluation of voice quality and its correlation with acoustic measurement,” *J. Voice*, vol. 18, no. 3, pp. 299–304, 2004.

[19] R. R. Patel *et al.*, “Recommended Protocols for Instrumental Assessment of Voice: American Speech- Language-Hearing Association Expert Panel to Develop a Protocol for Instrumental Assessment of Vocal Function,” *Am. J. Speech-Language Pathol.*, vol. 27, no. 3, pp. 887–905, 2018.

[20] C. A. Ferrer Riesgo and E. Nöth, “What Makes the Cepstral Peak Prominence Different to Other Acoustic Correlates of Vocal Quality?,” *J. Voice*, Aug. 2019.

[21] Y. Maryn, M. S. De Bodt, and N. Roy, “The Acoustic Voice Quality Index: Toward improved treatment outcomes assessment in voice disorders,” *J. Commun. Disord.*, vol. 43, no. 3, pp. 161–174, 2010.

[22] S. N. Awan, N. Roy, and C. Dromey, “Estimating dysphonia severity in continuous speech: application of a multi-parameter spectral/cepstral model.,” *Clin. Linguist. Phon.*, vol. 23, no. 11, pp. 825–41, 2009.

[23] K. Hosokawa *et al.*, “Acoustic breathiness index for the Japanese-speaking population: Validation study and exploration of affecting factors,” *J. Speech, Lang. Hear. Res.*, vol. 62, no. 8, pp. 2617–2631, 2019.

[24] P. H. Dejonckere, C. Obbens, G. M. de Moor, and G. H. Wieneke, “Perceptual Evaluation of Dysphonia: Reliability and Relevance,” *Folia Phoniatr. Logop.*, vol. 45, no. 2, pp. 76–83, 1993.

[25] A. Stráník, R. Čmejla, and J. Vokřál, “Acoustic parameters for classification of breathiness in continuous speech according to the GRBAS scale,” *J. Voice*, vol. 28, no. 5, pp. 653.e9-653.e17, 2014.

[26] T. Haderlein, C. Schwemmle, M. Döllinger, V. Matoušek, M. Ptok, and E. Nöth, “Automatic Evaluation of Voice Quality Using Text-Based Laryngograph Measurements and Prosodic Analysis,” *Comput. Math. Methods Med.*, vol. 2015, p. 316325, 2015.

[27] M. Fröhlich, D. Michaelis, H. W. Strube, and E. Kruse, “Acoustic voice analysis by means of the hoarseness diagram,” *J. Speech, Lang. Hear. Res.*, vol. 43, no. June, pp. 706–720, 2000.

[28] M. Englert, L. Lopes, V. Vieira, and M. Behlau, “Accuracy of Acoustic Voice Quality Index and Its Isolated Acoustic Measures to Discriminate the Severity of Voice Disorders,” *J. Voice*, vol. In Press, 2020.

[29] C. A. Ferrer, T. Haderlein, Y. Maryn, M. S. De Bodt, and E. Nöth, “Collinearity and Sample Coverage Issues in the Objective Measurement of Vocal Quality: The Case of Roughness and Breathiness,” *J. Speech, Lang. Hear. Res.*, vol. 61, no. 1, pp. 1–24, 2018.

[30] T. Nawka and L. C. Anders, “Die auditive Bewertung heiserer Stimmen nach dem RBH-System. Doppel-Audio-CD mit Stimmbeispielen (Doppel-CD und Begleitheft).” Stuttgart, New York, 1996.

[31] P. R. Walden, “Perceptual Voice Qualities Database (PVQD): Database Characteristics,” *J. Voice*, vol. In press, 2020.