ACOUSTIC "BREATHINESS MEASURES" IN THE DESCRIPTION OF PATHOLOGIC VOICES

*Matthias Fröhlich *), Dirk Michaelis, Hans Werner Strube*

Drittes Physikalisches Institut, Universität Göttingen
Bürgerstr. 42-44, D-37073 Göttingen, Germany
*) matth@physik3.gwdg.de

ABSTRACT

One important perceptual attribute of voice quality is breathiness. Since breathiness is generally regarded to be caused by glottal air leakage, acoustic measures related to breathiness may be used to distinguish between different physiological phonation conditions for pathological voices. Seven "breathiness features" described in the literature plus one self-developed measure (the glottal to noise excitation ratio, GNE) are compared for their distinguishing properties between different well-defined pathological phonation mechanisms. It is found that only GNE allows a distinction between all the pathological groups and both the normal and aphonic reference group. Furthermore, GNE is among the measures showing the most significant distinctions between the different pathologic phonation mechanism groups. Therefore GNE should be given preference over the other features in the independent assessment of glottal air leakage or "breathiness" for moderately or highly disturbed voices.

1. INTRODUCTION

In the assessment of pathological voice quality in a clinical context acoustic measures are increasingly applied as diagnostic aid. Acoustic voice analysis offers several advantages to the phoniatrician and speech pathologist: it is cheap, easy to perform, and non-invasive. The major problem of using acoustic measures in the assessment of pathological voice quality is their interpretation in terms of laryngeal physiology or glottal function. Appropriate measures should allow a consistent interpretation of the resulting numerical values, both with regard to normal or aphonic voices marking the extremes of voice quality and with regard to other laryngeal pathologies.

One characteristic of pathologically disturbed voices is the degree of glottal closure during phonation. Information on this important functional or physiological issue can be obtained by laryngo-stroboscopy or high-speed imaging of the vocal folds. This information can be used to define groups showing the same phonatory conditions. By applying different acoustic measures to describe these groups their usefulness in the assessment of pathological vocal fold function can be tested.

This work was supported by the Deutsche Forschungsgemeinschaft under Kr 1469/2-1

In this paper we present statistical analysis results of eight acoustic features (described in the following) that were applied to voice groups showing different phonation mechanism after surgical treatment for laryngeal cancer. These groups will be described in detail in the methods section.

The choice of potentially useful acoustic features to specifically assess glottal closure is somewhat problematic. Only few studies have investigated the correlation between vocal fold physiology and acoustic measures. Holmberg et al. determined the spectral energy difference between the first and second harmonic (H1H2) and the spectral energy difference between the first and third formant (F1F3) as useful acoustic measures to assess different aspects of vocal fold function [11].

Most studies in the field of acoustic voice quality description, however, are concerned with the acoustic correlates of *perceptual* attributes. Since many researchers regard "breathiness" as being caused by glottal air leakage [1, 4], acoustic measures of breathiness may serve as a pool supplying useful candidates in the assessment of glottal closure. The "breathiness measure" most commonly encountered is the already mentioned H1H2 [3, 9–11, 13]. Additionally, de Krom determined the harmonics to noise ratio (HNR) and several measures of the spectral slope as – somewhat ambiguous – best correlates of breathiness in a thorough study on the relation between acoustic measures and perceptual voice attributes [3]. Finally, the glottal to noise excitation ratio (GNE) has been designed as an acoustic measure of additive noise [15]. It was found to show similarly high correlations with breathiness as the normalized noise energy (NNE, [12]) or the cepstral HNR (CHNR, [2]) [20].

2. METHODS

2.1. Data

From a data base of 454 pathological voices subjects were chosen with well-defined phonation mechanisms after surgical treatment for laryngeal cancer. The grouping of the voices described in the following was performed by an experienced phoniatrician on the grounds of laryngoscopic and laryngo-stroboscopic examinations.

Patients with small carcinoma may maintain a glottic phonation after surgery and wound healing. Two different phenomena occur: either the affected vocal fold still vibrates in the same way as for a healthy voice ("glottal phonation *with* vibration of the operated vocal fold", abbreviated as gp^+, 18 subjects) or the tissue properties have changed resulting in a stiff vocal fold ("glottal phonation *without* vibration of the operated vocal fold", abbreviated as gp^-, 8 subjects). If a major part of the vocal fold tissue had to be resected the ventricular folds may serve as voice source ("ventricular phonation", abbreviated as vp, 6 subjects). If even the ventricular folds are unsuitable for voice generation, ary-epiglottic phonation may take place. This term describes a phonation mechanism that uses the epiglottis and/or parts of the ary-cartilage as voice source ("ary-epiglottic phonation", abbreviated as aep, 5 subjects).

37 subjects with no history of voice problems served as reference group ("normal group"). 60 whispered vowels simulating complete aphonia served as reference group of the other voice quality extreme ("aphonic group"). For each subject the speech data consisted of 28 digital recordings (sampling frequency 48 kHz) of 6 different isolated vowels (for details see [7]).

2.2. Statistical tests

The significance of the group differences was calculated by the Wilcoxon two-sample test [18]. A significance level of $p \leq 0.01$ was chosen for the interpretation of the results.

2.3. Acoustic measures

The glottal cycle length was calculated using the waveform matching algorithm [17, 19]. The inverse of the mean cycle length of the analysis frame defined the frequency of the first harmonic. The power spectrum was calculated frame-wise by fast Fourier transform after applying a Hanning window.

The measure H1H2 was calculated as spectral energy difference (in dB) between the first and second harmonics. The measure F1F3 is defined as the energy difference (in dB) between the first and third formants [11]. The formants were determined by linear predictive coding using the covariance method. In order to yield more dependable results for highly disturbed voices with little harmonic structure, not the energy of the formant *peaks* was used (as in [11]) but the energy of the formant *regions*. A formant region in the spectrum was defined by the 27 samples interval centered around the sample closest to the corresponding formant.

According to de Krom [3], useful spectral regions for the definition of various spectral measures are 60 to 400 Hz ("region 0"), 400Hz to 2kHz ("region 1"), 2 to 5 kHz ("region 2"), and 5 to 8kHz ("region 3"). have been used in [8]). His results suggest as "breathiness measures" the energy difference between region 1 and region 0 (L1L0 stated in dB), between region 3 and region 2 (L3L2 stated in dB), and the energy difference between the first harmonic and region 1 (H1L1 measured in dB).

The normalized noise energy (NNE [12]) and cepstral harmonics to noise ratio (CHNR [2]) were calculated according to the definitions given in the original papers. For both measures a frequency range of 1-5kHz was used since for this frequency region the highest correlations to breathiness were found [20].

The glottal to noise excitation ratio (GNE) describes the correlation between Hilbert envelopes calculated for different frequency bands [14–16]. It is based on the assumption that glottal pulses resulting from the collision of the vocal folds lead to a synchronous excitation of different frequency bands. Turbulent noise generated at a constriction, on the other hand, leads to an uncorrelated excitation. GNE reaches its maximum value of 1.0 if the envelopes in two different frequency bands are exactly the same. It was calculated using a 3 kHz bandwidth for the Hilbert envelopes [14].

For the calculation of H1H2, F1F3, H1L1, L1L0, and L3L2 the speech signal was down-sampled to 16 kHz and analyzed on frames of 1024 samples using a shift of 512 samples. For GNE, NNE, and CHNR the signal was down-sampled to 10 kHz and analyzed on 500ms frames with a shift of 250ms. For each acoustic measure the median value of all analysis frames of the 28 vowels was used for the statistical analysis.

All described acoustic measures were calculated in a completely automatic and unsupervised way. While this approach will result in artifacts for measures based on the calculation of harmonics if no harmonics are present (e.g., for aphonic voices), it allows a quantitative description of *any* voice. The analysis results have to show *a posteriori* for which measures this extrapolation of the algorithms leads to interpretable results.

3. RESULTS

The group means and the significance of the differences between the groups are stated in Table 1. None of the features except GNE shows a significant difference for all cancer groups with regard to both the normal and the aphonic group. Looking at the five measures H1H2, F1F3, H1L1, L1L0, and L3L2 the only significant difference between individual cancer groups is found for H1L1 between gp^+ and aep. GNE shows significant differences between gp^+ and all the other phonation mechanism groups. Significant differences between gp^+ and the other groups are also found for CHNR and for NNE with the exception of the difference between gp^+ and gp^-. NNE and CHNR are the only two measures that show significant differences between gp^- and aep.

4. DISCUSSION

The findings that only GNE succeeded in showing significant differences between the cancer groups and both the normal and the aphonic voice group indicate that results obtained by the other measures should be interpreted with caution. From a physiological perspective all the described phonation mechanisms show considerable deviations from a healthy voice. On the other hand, all speech samples of the cancer groups were voiced. Therefore the failure to distinguish the cancer groups from the normal and aphonic references cautions against the exclusive use of any of the features except GNE (at least for moderately or highly disturbed voices).

For H1H2 the group means of gp^-, vp, and aep are significantly higher (5.2-8.5dB) than for the normal voices (0.6dB). This is in accordance with the assumption that a poor glottal closure leads to a stronger attenuation of the harmonics [3, 10]. On the other hand, the mean of the aphonic group (-8.4dB) can be interpreted as indication of the spectral shape of the noise excitation. This noise can be described as bandpass-filtered white noise [10], so that the negative value reflects the attenuation of low-frequent spectral energy. However, the transition from highly disturbed voices that still exhibit a harmonic structure (e.g., aep) to complete aphonia remains unclear. Therefore the ultimate extrapolation to aphonic voices seems questionable for H1H2.

Table 1: Differences (x-y) of the group means (stated in parentheses) for the different acoustic features (aphonic group means: -8.39dB (H1H2), 11.00dB (F1F3), 3.23dB (H1L1), 0.76dB (L1L0), -2.27dB (L3L2), -2.49dB (NNE), 3.53dB (CHNR), 0.10 (GNE)). Insignificant differences are indicated by an asterisk *.

($\downarrow x$, $y \rightarrow$)	gp^+	gp^-	vp	aep	aph
H1H2 [dB]					
norm (0.59)	0.73^*	-7.91	-4.56	-4.63	8.99
gp^+ (-0.14)		-8.64^*	-5.30^*	-5.36^*	8.26
gp^- (8.50)			3.35^*	3.28^*	16.90
vp (5.16)				-0.06^*	13.55
aep (5.22)					13.62
F1F3 [dB]					
norm (17.45)	-0.50^*	-0.34^*	-4.34^*	-0.58^*	6.45
gp^+ (17.94)		0.15^*	-3.85^*	-0.08^*	6.94
gp^- (17.79)			-4.00^*	-0.23^*	6.79
vp (21.79)				3.77^*	10.79
aep (18.02)					7.02^*
H1L1 [dB]					
norm (22.09)	-0.05^*	-1.53^*	7.49	9.34	18.86
gp^+ (22.14)		-1.48^*	7.55^*	9.39	18.91
gp^- (23.62)			9.02^*	10.87^*	20.39
vp (14.60)				1.84^*	11.36
aep (12.76)					9.52
L1L0 [dB]					
norm (-0.11)	0.32^*	0.59^*	0.52^*	0.23^*	-0.87
gp^+ (-0.43)		0.27^*	0.20^*	-0.09^*	-1.19
gp^- (-0.70)			-0.07^*	-0.36^*	-1.46
vp (-0.62)				-0.29^*	-1.39
aep (-0.33)					-1.10
L3L2 [dB]					
norm (-1.95)	-0.60	-0.79	-1.31	-0.36^*	0.33
gp^+ (-1.35)		-0.19^*	-0.71^*	0.24^*	0.93
gp^- (-1.16)			-0.52^*	0.43^*	1.12
vp (-0.64)				0.95^*	1.64
aep (-1.59)					0.69^*
NNE [dB]					
norm (-12.08)	-4.79	-7.52	-9.69	-10.49	-9.59
gp^+ (-7.29)		-2.73^*	-4.89	-5.69	-4.80
gp^- (-4.56)			-2.17^*	-2.97	-2.07
vp (-2.40)				-0.80^*	0.09^*
aep (-1.60)					0.89^*
CHNR [dB]					
norm (17.20)	5.28	9.98	12.10	13.56	13.67
gp^+ (11.92)		4.71	6.82	8.28	8.39
gp^- (7.21)			2.11^*	3.57	3.68
vp (5.10)				1.46^*	1.57^*
aep (3.64)					0.11^*
GNE					
norm (0.76)	0.15	0.49	0.43	0.51	0.65
gp^+ (0.61)		0.34	0.28	0.36	0.50
gp^- (0.27)			-0.06^*	0.02^*	0.16
vp (0.33)				0.08^*	0.22
aep (0.25)					0.14

The measures that allow a consistent ranking of the group means (normal voices and aphonia marking the extremes, cancer groups in between with gp^+ closer to the normal voices) are H1L1, GNE, NNE, and CHNR. Therefore only for these measures the extrapolation to aphonic voices seems permissible. For GNE, an extrapolation to highly disturbed and aphonic voices has been successfully used in the description of different pathological groups and in the monitoring of voice quality changes for individual subjects [6, 7, 14].

Of the acoustic "breathiness measures" tested, only GNE, NNE, and CHNR led to significant differences between the pathologic groups and the normal reference group. While such significances have to be regarded as prerequisite, the many significant differences among the pathologic phonation mechanisms indicate the usefulness of these measure in the quantitative description of different pathologic voice conditions. The significant difference between gp^+ and aep seen for H1L1 is also found for GNE, NNE, and CHNR. Therefore the overall distinction between the different phonation mechanisms is not increases if H1L1 supplements any of these three measures.

With regard to insignificance of the differences between the groups gp^-, vp, and aep for GNE several explanations are possible. While the relatively small number of subjects in these groups may be in part responsible for this behavior, an alternative explanation might be found in the multi-dimensional nature of voice quality itself. Assuming that GNE indicates glottal leakage, then pathologies with comparable glottal gaps cannot be expected to differ significantly if described by the GNE. To separate between such groups other signal characteristics have to be assessed, e.g. through measures of signal periodicity.

The laryngo-stroboscopic examinations reveal that the vibration patterns for the three groups gp^-, vp, and aep show considerable differences in the regularity of the vibration. Several studies have shown that measures like NNE or CHNR that are sensitive both to additive noise and aperiodicities of the speech signal [14, 15] should be considered as correlates of overall voice quality rather than of the subordinate voice quality "breathiness" [4, 5, 20]. Therefore it is not surprising that these measures have been determined as appropriate measures of breathiness as well [3]. The significant difference between gp^- and aep found for NNE and CHNR may therefore be attributed to this combined sensitivity to both additive noise and signal aperiodicity. This explanation is supported by the findings that a linear combination of three aperiodicity measures also leads to a significant difference between the two groups gp^- and aep [7].

5. CONCLUSION

The analysis of the different phonation mechanisms by acoustic measures that have been associated in the literature with the perceptual voice attribute "breathiness" shows that of all measures tested, only GNE meets the requirement to show significant differences between the disturbed voices and both normal and completely aphonic voices. GNE, NNE, and CHNR allow the best differentiation between the different phonation mechanisms found for cancer patients. Since the harmonics to noise ratio and related measures (e.g., CHNR and NNE) have shown higher correlations with the overall voice quality than with subordinate voice qualities such as breathiness, GNE should be given preference if an independent assessment of glottal function associated with “breathiness” for moderately and severely disturbed voices is desired.

References

[1] D. Childers and C. Lee. Vocal quality factors: Analysis, synthesis, and perception. *J. Acoust. Soc. Am.*, 90(5):2394–2410, 1991.

[2] G. de Krom. A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals. *J. Speech Hear. Res.*, 36:224–266, 1993.

[3] G. de Krom. Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments. *J. Speech Hear. Res.*, 38:794–811, 1995.

[4] P. Dejonckere. Principal components in voice pathology. *Voice*, 4:96–105, 1995.

[5] L. Eskenazi, D. Childers, and D. Hicks. Acoustic correlates of vocal quality. *J. Speech Hear. Res.*, 33:298–306, 1990.

[6] M. Fröhlich, D. Michaelis, H. W. Strube, and E. Kruse. Acoustic voice quality description: Case studies for different regions of the hoarseness diagram. In T. Wittenberg, P. Mergell, M. Tigges, and U. Eysholdt, editors, *Advances in Quantitative Laryngoscopy, 2nd ‘Round Table’*, pages 143–150, Erlangen, 1997.

[7] M. Fröhlich, D. Michaelis, H. W. Strube, and E. Kruse. Voice quality assessment using the hoarseness diagram. To be published, 1998.

[8] B. Hammarberg, B. Fritzell, J. Gauffin, J. Sundberg, and L. Wedin. Perceptual and acoustic correlates of abnormal voice qualities. *Acta Otolaryngol*, 90:441–451, 1980.

[9] J. Hillenbrand, R. A. Cleveland, and R. L. Erickson. Acoustic correlates of breathy vocal quality. *J. Speech Hear. Res.*, 37:769–778, 1994.

[10] J. Hillenbrand and R. A. Houde. Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. *J. Speech Hear. Res.*, 39:311–321, 1996.

[11] E. B. Holmberg, R. E. Hillman, J. S. Perkell, P. C. Guiod, and S. L. Goldman. Comparisons among aerodynamic, electroglottoghraphic, and acoustic spectral measures. *J. Speech Hear. Res.*, 38:1212–1223, 1995.

[12] H. Kasuya, S. Ogawa, K. Mashima, and S. Ebihara. Normalized noise energy as an acoustic measure to evaluate pathologic voice. *J. Acoust. Soc. Am.*, 80(5):1329–1334, 1986.

[13] J. Kreiman and B. R. Gerratt. The multidimensional nature of pathologic vocal quality. *J. Acoust. Soc. Am.*, 96(3):1291–1302, 1994.

[14] D. Michaelis, M. Fröhlich, and H. W. Strube. Selection and combination of acoustic features for the description of pathologic voices. *J. Acoust. Soc. Am.*, 1998 (in review).

[15] D. Michaelis, T. Gramss, and H. W. Strube. Glottal to noise excitation ratio - a new measure for describing pathological voices. *Acustica / acta acustica*, 83:700–706, 1997.

[16] D. Michaelis and H. W. Strube. Empirical study to test the independence of different acoustic voice parameters on a large voice database. In J. M. Pardo, E. Enriquez, J. Ortega, J.Ferreiros, J. Macias, and F. J. Valverde, editors, *Eurospeech ’95*, volume 3, pages 1891–1894, 1995.

[17] P. Milenkovic. Least mean square measures of voice perturbation. *J. Speech Hear. Res.*, 30(4):529–538, 1987.

[18] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling. *Numerical Recipes in C*. Cambridge University Press, 1988.

[19] I. R. Titze and H. Liang. Comparison of f0 extraction methods for high precision voice perturbation measurement. *NCVS Status and Progress Report*, 3:97–115, 1992.

[20] P. Zwirner, D. Michaelis, M. Fröhlich, H. W. Strube, and E. Kruse. Korrelationen zwischen perzeptueller Beurteilung von Stimmen nach dem RBH-System und akustischen Parametern. In M. Gross, editor, *Aktuelle phoniatrisch-pädaudiologische Aspekte 1997*, volume 5, Berlin, 1998. Renate Gross Verlag (in press).