PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Comparison of acoustic and laryngeal parameters between healthy and disordered subjects

Patrick SCHLEGEL[1]; Stefan KNIESBURGES[1]; Stephan DÜRR[1]; Michael DÖLLINGER[1]; Anne SCHÜTZENBERGER[1]

[1] University Hospital Erlangen, Dept. of Otorhinolaryngology, Div. Phoniatrics & Pediatric Audiology

ABSTRACT

In voice research and clinics, various mathematical parameters are in use to describe different features of the recorded data. However, little is known about norm values of parameters and how they are affected by different pathologies such as functional dysphonia.

Keywords: Phonation, parameters, functional dysphonia, high-speed video endoscopy

1. INTRODUCTION

In voice research, various parameters are in use [1], all of which are intended to describe different features of the phonation process to allow differentiation between groups of subjects. To determine the usefulness of objective parameters for the detection of functional dysphonia related disorders, we investigated data recorded in a clinical environment applying many different objective parameters.

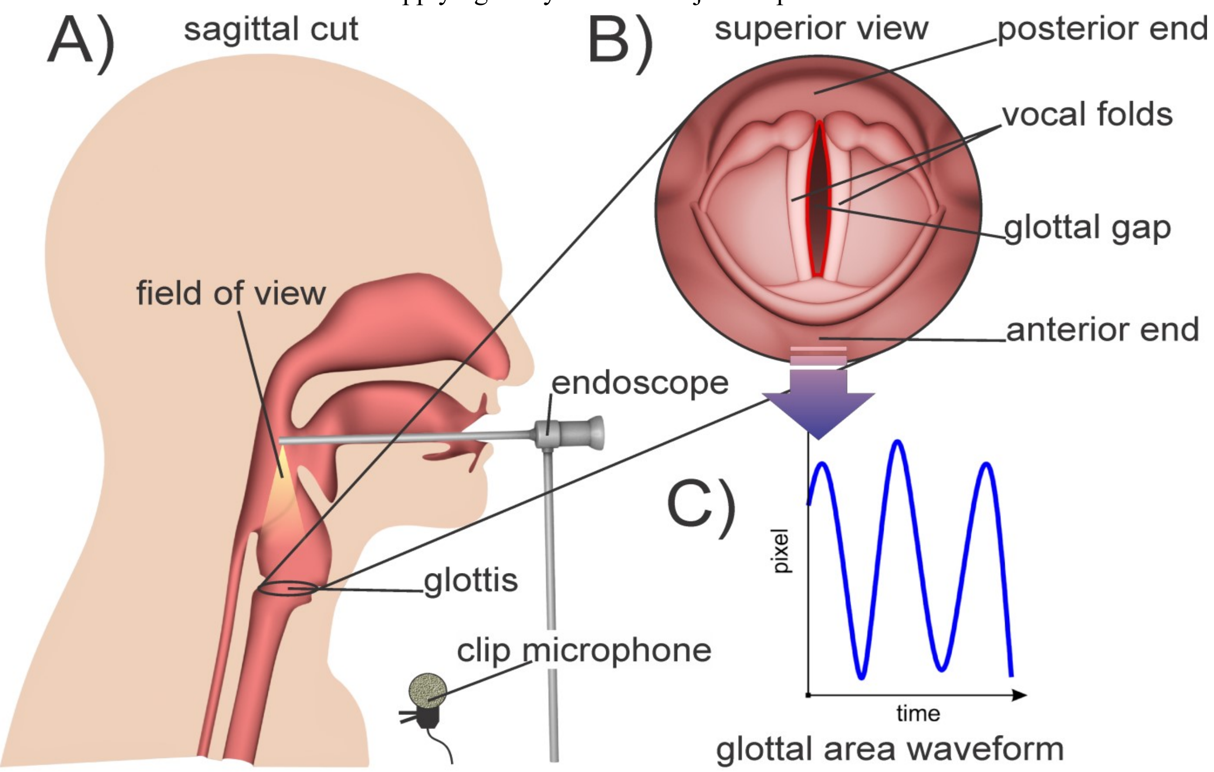

Figure 1 A) Recording of the vocal fold oscillations via a rigid endoscope being attached to a high-speed camera and parallel audio recording using a clip microphone. B) Superior view of the vocal folds as seen with the endoscope. C) Computed GAW: amount of pixel in the glottis over time.

[1] Patrick.schlegel@uk-erlangen.de

2. METHODS

In total, 352 endoscopically recorded HSV videos with synchronously recorded audio were investigated. Data collection and usage was approved by the ethic committee of the Medical School at Friedrich-Alexander-University Erlangen-Nürnberg (no. 290_13B). The videos were divided into four groups: females (162 healthy, 92 disordered) and males (67 healthy, 31 disordered) whereas all disorders were functional dysphonia related. All subjects phonated the vowel /i/ at a comfortable pitch and loudness level during examination. All videos were recorded by the clinically used Photron Fastcam MC2 with a spatial resolution of 512×256 pixels and a frame rate of 4000 fps. The synchronously recorded audio had a sampling rate of 40000 Hz (in 19 cases for disordered females and in 2 cases for disordered males the audio sampling rate was only 16000 Hz due to incorrectly set settings during recording). All videos included exactly 1000 frames of recorded phonation.

The videos were recorded over three years and, in case of the "healthy" groups, are part of different measurement series. Data from the disordered groups were not recorded as part of an organized series of measurements. Instead, the data was collected over the years from various ENT clinic patients. Apart from that, recording conditions stayed the same. For group-internal comparisons, subgroups were composed of the different measurement series for the healthy group and from the different years for the disordered group.

In this work, we investigate the expressiveness of in total 59 glottal area waveform (GAW) parameters (from which 33 were also calculated for the synchronously recorded audio signal).

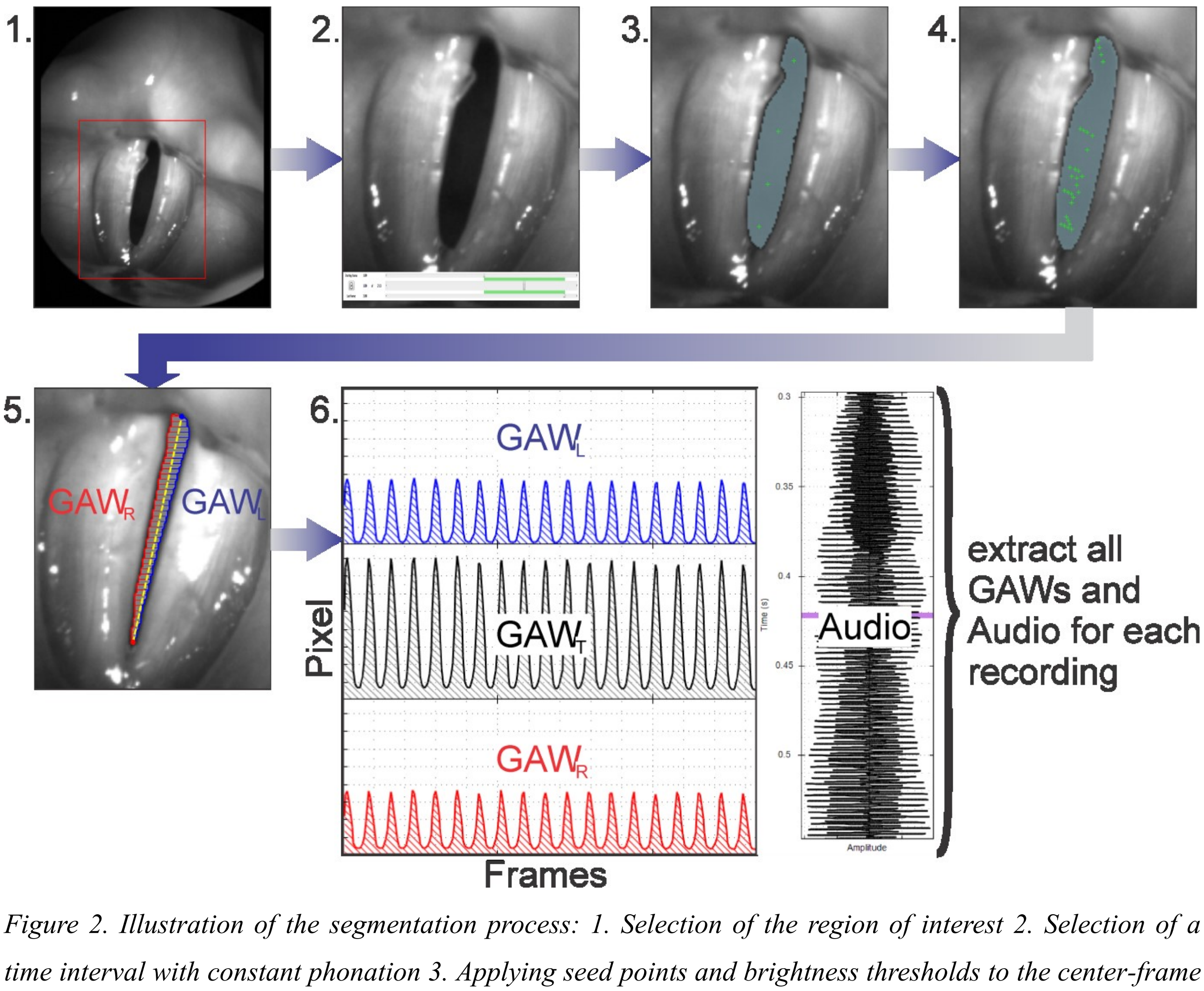

Figure 2. Illustration of the segmentation process: 1. Selection of the region of interest 2. Selection of a time interval with constant phonation 3. Applying seed points and brightness thresholds to the center-frame 4. Reviewing and adjusting 5. Computing partial GAWs 6. Extraction of total and partial GAWs and audio.

Table 1: Number of data sets for each subgroup. Each subgroup with at least 20 data sets is highlighted.

	females		*males*	
	healthy	*disordered*	*healthy*	*disordered*
Subgroup 1	*23*	*16*	*22*	*2*
Subgroup 2	*12*	*40*	*13*	*17*
Subgroup 3	*30*	*36*	*29*	*12*
Subgroup 4	*61*	*-*	*-*	*-*
Subgroup 5	*22*	*-*	*-*	*-*
Subgroup 6	*14*	*-*	*3*	*-*
total	*162*	*92*	*67*	*31*

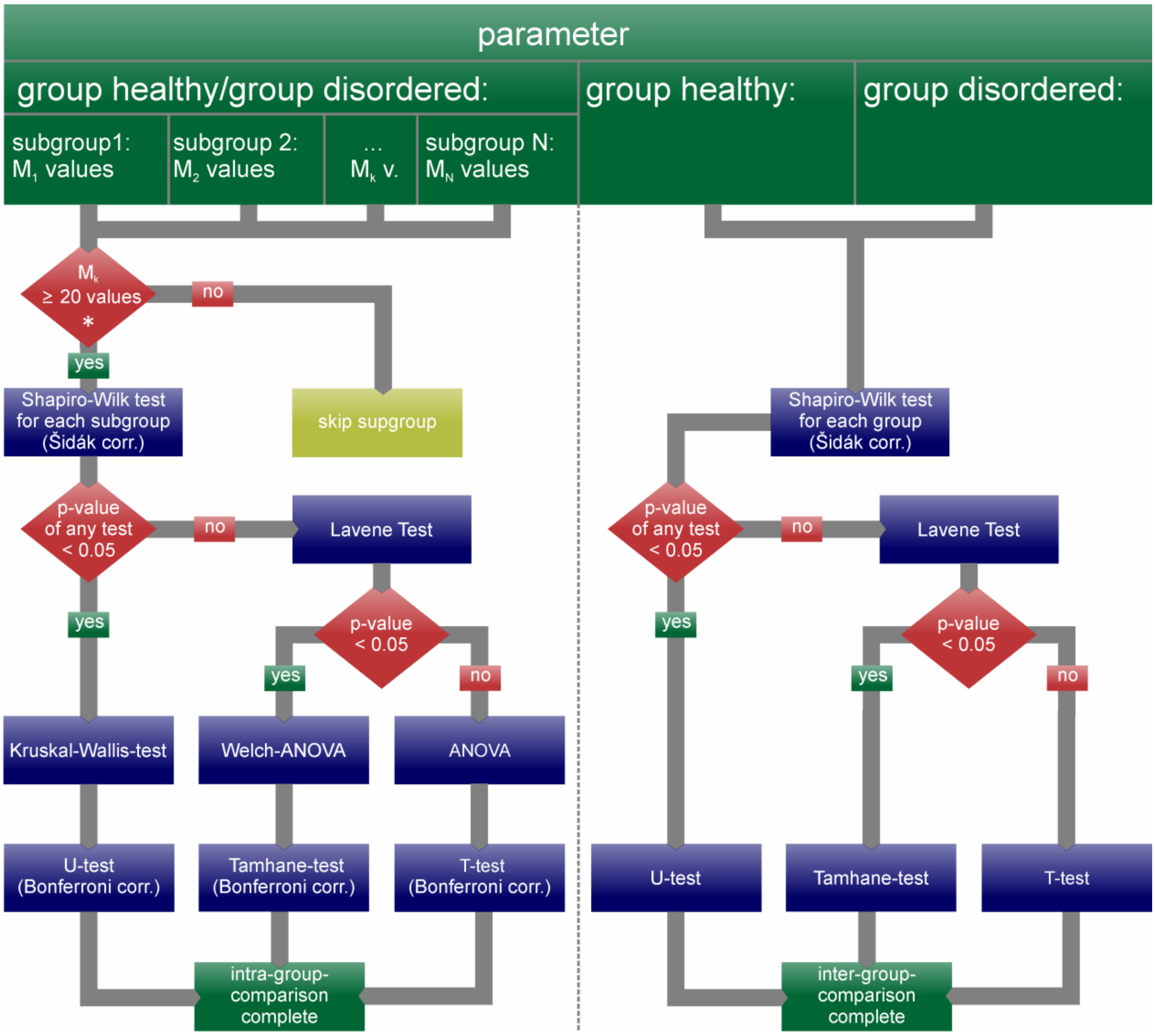

*Figure 3. The depicted statistical analysis workflow was performed for each parameter. *if only two subgroups contained 20 values or more, intragroup comparison was done analogously to intergroup comparison*

3. RESULTS AND DISCUSSION

Data analysis revealed that most parameters, for audio and GAW signals, did not show clear differences between the healthy and the disordered groups. In fact, some parameters differed considerably between different measurement series that were part of the healthy group. Out of 33 audio parameters only one parameter showed a statistically significant difference between healthy and disordered groups but not also between the different measurement series within the groups (intragroup comparisons). Also, this difference may only be attributable to the deviating sampling frequency for some of the audio signals in the disordered groups. Out of 59 GAW parameters, 30 parameters were statistically significantly different between healthy and disordered groups for males, females or both, but not also statistically significantly different for intragroup comparisons. The high proportion of statistically significantly different parameters for intragroup comparisons on the one hand may be attributable to minor changes in the recording conditions as it may take place in a clinical environment (e.g. one of the medical doctors is diseased and a colleague has to take his place). On the other hand, often parameters are developed and tested on rather small sample sizes of audio and video signals, recorded in a laboratory environment where sound-isolated rooms and optimal recording conditions are available [2-4]. In some cases, the parameters are even tested exclusively on artificially generated signals [5]. So the question arises if parameters tested by such means are still reliable if applied to clinical data.

4. CONCLUSIONS

As the results of this study imply, especially audio parameters seem to be not sufficiently robust to influences on the measuring process, as they may take place in a clinical setting. Also, in general parameters that changed statistically significantly between the healthy and disordered groups but not within the groups only showed limited deviations. However, it may be possible that a more complex, nonlinear multi-parametric based approach could allow for a differentiation between healthy and disordered subjects.

ACKNOWLEDGEMENTS

This project is funded by the Deutsche Forschungsgemeinschaft (grant no. 323308998).

REFERENCES

1. Pedersen M, Jønsson A, Mahmood S, Agersted A. Which mathematical and physiological formulas are describing voice pathology: An overview. Journal of General Practice. 2016;4(253). doi:10.4172/2329-9126.1000253
2. Hillenbrand J, Cleveland R, Erickson RL. Acoustic correlates of breathy vocal quality. Journal of Speech and Hearing Research. 1994;37:769-778
3. Yumoto 5, Gould WJ, Baer T. Harmonics-to-nose ratio as an index of the degree of hoarseness. Journal of the Acoustical Society of America. 1982;71:1544-1550
4. Qi Y, Hillman RE, Milstein C. The estimation of signal-to-noise ratio in continuous speech for disordered voices. Journal of the Acoustical Society of America. 1999;105:2532-2535
5. Noise estimation in voice signals using short-term cepstral analysis. The Journal of the Acoustical Society of America 2007;121(1679). doi: 10.1121/1.2427123

Table 2: p-values for intra- and inter-group-comparisons of parameters, which changed significantly between groups but not within a group.

Parameter name	Signal type	p-value intra-group females / males	p-value inter-group females / males
Plateau Quotient	GAW_T	0.690 / 0.817 *	0.000 / 0.038
Maximum-Area-Declination-Rate	GAW_T	- / 0.805	- / 0.012
Peak-Closing-Velocity	GAW_T	0.057 / -	0.046 / -
Amplitude-to-Length-Ratio	GAW_T	0.195 / -	0.000 / -
Harmonics Intensity	GAW_T	0.336 / 0.397	0.000 / 0.010
Harmonic to noise ratio	GAW_T	0.751 / -	0.002 / -
Normalized noise energy	GAW_T	0.105 / -	0.000 / -
Signal to noise ratio v1	GAW_T	0.180 / -	0.000 / -
Signal to noise ratio v2	GAW_T	0.575 / -	0.022 / -
Spectral flatness	GAW_T	0.481 / 0.123	0.005 / 0.042
Mean Waveform matching coeff.	GAW_T	0.198 / -	0.046 / -
Amplitude Variability Index	GAW_T	0.268 / -	0.018 / -
Energy Perturbation Quotient 5%	GAW_T	- / 0.142	- / 0.014
Energy Perturbation Factor	GAW_T	- / 0.067	- / 0.023
Time-Periodicity	GAW_T	- / 0.077	- / 0.028
Mean Jitter	GAW_T	0.117 / -	0.041 / -
Jitter Percent	GAW_T	- / 0.08	- / 0.026
Jitter Factor	GAW_T	- / 0.077	- / 0.023
Jitter Ratio	GAW_T	- / 0.08	- / 0.026
Period Perturbation Quotient 3%	GAW_T	- / 0.071	- / 0.033
Period Perturbation Quotient 5%	GAW_T	- / 0.118	- / 0.022
Period Perturbation Factor	GAW_T	- / 0.08	- / 0.025
Relative Average $Perturbation_B$	GAW_T	- / 0.098	- / 0.027
Relative Average $Perturbation_K$	GAW_T	- / 0.074	- / 0.032
Period Variability Index	GAW_T	0.609 / 0.128	0.005 / 0.005
Spatial Symmetry Index	GAW_L / GAW_R	- / 0.761	- / 0.022
Amplitude Symmetry Index	GAW_L / GAW_R	0.756 / -	0.007 / -
Dynamic range Symmetry Index	GAW_L / GAW_R	0.987 / -	0.012 / -
Waveform symmetry index	GAW_L / GAW_R	- / 0.392	- / 0.049
Phase Asymmetry	GAW_L / GAW_R	0.461 / -	0.001 / -
Harmonics Intensity	audio	0.436 / 0.790	0.000 / 0.018

*The p-values from Kruskal-Wallis-test, Welch-ANOVA or ANOVA in case of multiple intra-group comparisons. For pairwise comparisons, the p-values of U-test, Tamhane-test or t-test are given (Depending on whether the data was assumed to be normally distributed and if so with equal or unequal variances).