Feature Evaluation for Acoustic Pressure Signals from a Hybrid Simulation of Human Phonation

Florian Kraxberger[1], Andreas Wurzinger[1], Sebastian Falk[2], Stefan Schoder[1], Stefan Kniesburges[2], Michael Döllinger[2], Manfred Kaltenbacher[1]

[1] *Technische Universität Graz, Institut für Grundlagen und Theorie der Elektrotechnik (IGTE), 8010 Graz, Österreich, Email: kraxberger@tugraz.at*

[2] *Universitätsklinikum Erlangen, Abteilung für Phoniatrie und Pädaudiologie, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91054 Erlangen, Deutschland*

Introduction

Communication disorders, such as voice anomalies can have a severe effect on the affected persons' quality of life, especially for heavy occupational voice users, e.g. teachers, and can even lead to isolation and depression [1]. In Fig. 1, the anatomic structures involved in the human voice are depicted. Due to a pressure difference between the lungs and the mouth there is the tracheal airflow through the vocal folds (VF) and the ventricular folds (VeF). The vibrations of the VF interrupt the tracheal airflow, causing turbulences, vortices and flow instabilities in the supraglottal region, see Fig. 1.

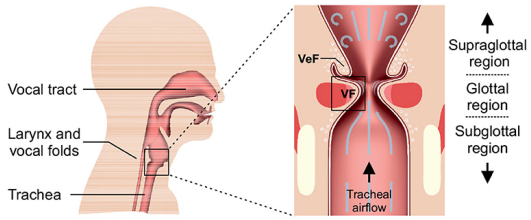

Figure 1: 2D view of a human head (left) with detail of the larynx (right). [2]

To investigate disorders and anomalies in human phonation, studies employing a hybrid aeroacoustic simulation model have been published recently [2, 3, 4]. The simulation model consists of a finite-volume incompressible computational fluid dynamics (CFD) simulation (the CFD geometry is depicted in Fig. 2), from which aeroacoustic source terms are computed. In the CFD, the VF motion is prescribed following the M5 model [5]. Successively, a computational acoustics (CA) simulation is performed using the geometry depicted in Fig. 3, based on the perturbed convective wave equation (PCWE), which is given in [6] as

$$\frac{1}{c^2}\frac{\mathrm{D}^2\psi^\mathrm{a}}{\mathrm{D}t^2} - \Delta\psi^\mathrm{a} = -\frac{1}{\bar{\rho}c^2}\frac{\mathrm{D}p^\mathrm{ic}}{\mathrm{D}t}, \tag{1}$$

where c is the speed of sound, ψ^a is the acoustic scalar potential, $\bar{\rho}$ is the mean air density, p^ic is the incompressible pressure (the result of the CFD simulation), and $\frac{\mathrm{D}}{\mathrm{D}t} = \frac{\partial}{\partial t} + \overline{\boldsymbol{v}}\cdot\nabla$ is the substantial derivative. The acoustic pressure p^a is evaluated from the acoustic scalar potential ψ^a with $p^\mathrm{a} = \bar{\rho}\frac{\mathrm{D}\psi^\mathrm{a}}{\mathrm{D}t}$. Due to the low Mach number of the flow, the convective part of the substantial derivative can be neglected in (1), such that the substantial derivative collapses to a pure time derivative [7]. An overview in form of a flow chart of the hybrid aeroacoustic workflow for human phonation employing the PCWE has been published in [8, Fig. 3], and a detailed description is available in [3].

The goal of this work is to find cause-effect relations, where the cause are different kinds of voice anomalies, and the effect is given by acoustic features evaluated from the simulated acoustic pressure signal.

Method

To find the cause-effect relation, the simulated voice anomalies (i.e. the 'cause') and the evaluated features (i.e. the 'effect') are described.

Simulation of Voice Anomalies

Three types of voice anomalies have been investigated in various degrees. These are (i) three different subglottal pressures, (ii) four different glottal closure types, and (iii) two cases of VF motion symmetry.

Subglottal Pressure: The inability to build up a pressure difference between the lungs and the mouth can be one reason for an dysphonic voice, e.g. caused by muscle tension disorders, as reported in [9, 10]. Subglottal pressures of $P_\mathrm{sub} \in \{385, 775, 1500\}$ Pa have been modeled by means of a constant pressure boundary condition at the inlet depicted in Fig. 2.

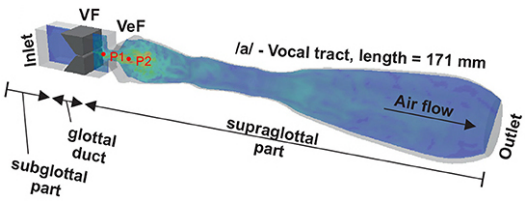

Figure 2: Geometry setup of the CFD simulations. The subglottal pressure is modeled as a constant inlet boundary condition. [2]

Glottal Closure (GC) Type: The vocal folds can have varying degrees of glottal insufficiency, which occur when the vocal folds do not close completely during each cycle. According to [11, 12], this irregularity occurs for many voice pathologies. As a consequence, concerned patients report, that a higher effort is necessary for phonation [13]. To model glottal insufficiencies, the initial glottal opening o_initial is varied in four levels (GC1 — GC4).

Thereby, the following values have been used for the initial opening: $o_{\text{initial}} \in \{0, 40, 70, 100\}\,\%$, corresponding to the GC types GC1, GC2, GC3, and GC4, respectively [2]. A detailed description as well as visualizations of the different symmetry cases are available in [14, Fig. 5.5].

VF Motion Symmetry: In regular phonation, the left and right vocal folds oscillate symmetrically. However, in the case of laryngeal hemiparesis or unbalanced muscle tension, the vocal folds oscillate asymmetrically [15, 16]. Therefore, the simulation model features two cases of vocal fold motion symmetry, as depicted in [2, Fig. 4]. In the symmetric case, both vocal folds have the same amplitude in their motion. In the asymmetric case, one vocal fold oscillates with 50 % of the amplitude of the other vocal fold.

All possible combinations of voice anomalies are simulated resulting in 24 simulation configurations. For all configurations, the vocal tract geometry is constant resembling the vowel */a/*, and the fundamental frequency of the VF vibration is 148 Hz.

Acoustic Features

To quantify the effect of the simulated voice anomalies, the acoustic pressure signal $p^{\text{a}}[n]$ is evaluated at the microphone position depicted in Fig. 3, where n is the discrete-time step number. Extracting features from audio recordings is a common practice in voice science and phonology [17, 18, 19, 20, 21, 22]. A detailed description of the acoustic features used in the present work is available in [23]. In the following, the features are described briefly.

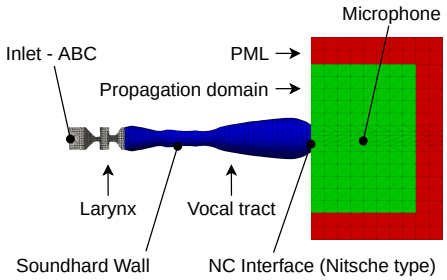

Figure 3: Cross-section of the CA simulation geometry [24]. The acoustic pressure is evaluated at the microphone point for computing acoustic features.

Sound Pressure Level (SPL): The subglottal pressure, as well as other simulation parameters, is expected to have an influence on the SPL of the acoustic signal. The SPL L_p is computed as follows

$$L_p = 20\,\log_{10}\frac{\tilde{p}}{20\,\mu\text{Pa}} \quad \text{with} \quad \tilde{p} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(p[n]\right)^2}, \quad (2)$$

where N is the total number of time steps, and $\tilde{p}$ denotes the root-mean-square of the acoustic pressure signal.

Harmonics to Noise Ratio (HNR): HNR is defined as the energy ratio between the harmonic signal component to the noise-like signal component in dB. The splitting of harmonic and noise-like components is performed via the auto-correlation function $r_p(\tau)$, as described in [18, pp. 77–78]. It is computed as follows

$$\text{HNR}_{\text{dB}} = 10\log_{10}\frac{\max_{\tau_{\text{marg}}}^{\tau_{\max}}\{r_p(\tau)\}}{r_p(0) - \max_{\tau_{\text{marg}}}^{\tau_{\max}}\{r_p(\tau)\}}, \quad (3)$$

where $\tau_{\max}$ is the maximum possible lag time (i.e. the time corresponding to the highest lag index) and $\tau_{\text{marg}} = 4.08\,\text{ms}$ is a lag margin.

Cepstral Peak Prominence (CPP): The CPP is computed as defined in [25, 26, 27]. Similar to HNR, the CPP is also a measure of harmonicity of the signal, and quantifies how prominent and numerous harmonic components are in a signal compared to non-harmonic components. It is the difference between the cepstral peak and a linear regression of the cepstrum [25, 26].

Hammarberg Index η and Alpha Ratio ρ_α: Hammarberg Index η (HBI) and Alpha Ratio ρ_α are measures of the energy distribution across the spectrum. The HBI η is the ratio of the most prominent energy peak in the range of 0 kHz–2 kHz to the most prominent energy peak in the region 2 kHz–5 kHz. It is calculated according to [18, p. 38], such that

$$\eta = \frac{\max_{k=1}^{k_{\text{pivot}}}\{\tilde{p}[k]\}}{\max_{k=k_{\text{pivot}}+1}^{k_{\max}}\{\tilde{p}[k]\}}, \quad (4)$$

where the pivot frequency bin k_{pivot} is the highest spectral bin for which $f \leq 2\,\text{kHz}$, and $k_{\max}$ is the highest spectral bin for which $f \leq 5\,\text{kHz}$. The alpha ratio is similar to the HBI described above, but instead of computing the ratio between energy peaks, the energy sum in the frequency bands is considered [28]. Using the frequency bands 50 Hz–1 kHz and 1 kHz–5 kHz, the alpha ratio ρ_α can be computed according to [18, pp. 38–39], such that

$$\rho_\alpha = \frac{\sum_{k=k_{\text{start}}}^{k_{\text{pivot}}}\tilde{p}[k]}{\sum_{k=k_{\text{pivot}}+1}^{k_{\max}}\tilde{p}[k]}, \quad (5)$$

where the k_{start} is the lowest spectral bin for which $f \geq 50\,\text{Hz}$, k_{pivot} is the highest spectral bin for which $f \leq 1\,\text{kHz}$, and $k_{\max}$ is the highest spectral bin for which $f \leq 5\,\text{kHz}$.

Spectral Slope: The spectral slope is a measure of energy distribution across the spectrum. In [18], spectral slope is defined as the result of a linear regression operation on the whole spectrum of the pressure signal.

Low-Pass Filtering: From the acoustic pressure signals obtained by the finite-element aeroacoustic simulation, the features defined above are evaluated. Prior to the feature evaluation, the signals are low-pass filtered with a cut-off frequency of 5 kHz. This is due to the upper frequency limit of the simulation setup [3]. Because we are only focused on the lowest two formants, which are below 2 kHz [29], the features which allow for a low-pass filtering, i.e. CPP, HNR and spectral slope, are evaluated *additionally* with a low-pass filter at 2 kHz. This results in a total of 9 evaluated features, which are analyzed in the following.

Results

In Fig. 4, the correlations between simulation parameters and evaluated features are visualized. Therefrom, it can be estimated, that subglottal pressure will be correlated mostly with SPL, ρ_α, and HNR. GC type is expected to correlate mostly with η, HNR and CPP. The symmetry property exhibits weak correlations as depicted in Fig. 4. Furthermore, it can be seen from Fig. 4, that features with a 2 kHz low-pass filter show stronger correlations than the 5 kHz low-pass filtered version of the same feature.

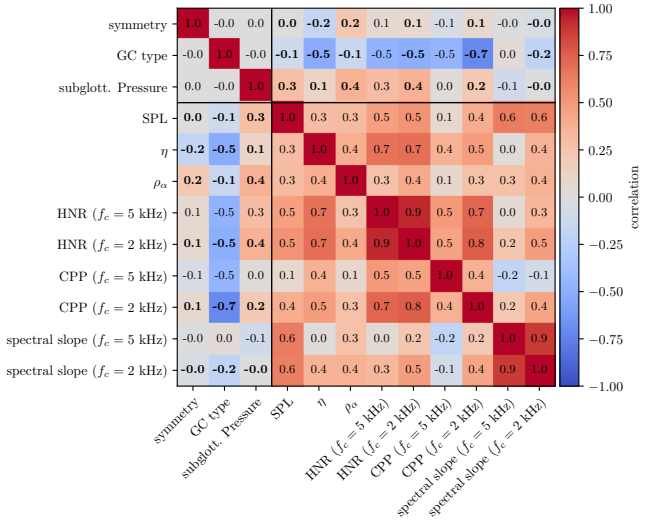

Figure 4: Visualization of the correlation matrix for simulation parameters and evaluated features.

To investigate the correlations in detail, boxplot diagrams are used. In Fig. 5, we see that the evaluated features allow to discriminate two groups of subglottal pressures: {358 Pa} on the one hand and {775, 1500}Pa on the other hand can be discriminated with SPL, ρ_α and to a certain extent also with η.

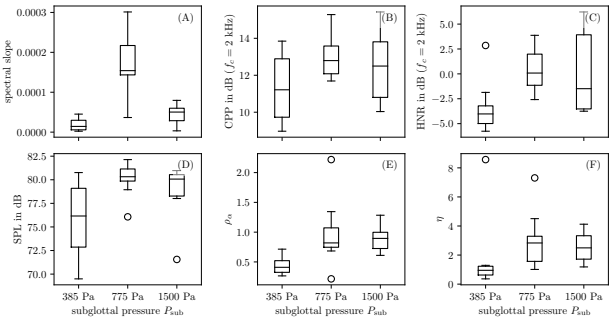

Figure 5: Boxplot diagrams of evaluated features and subglottal pressure.

In Fig. 6, we see that an increasing glottal insufficiency leads to a decrease in CPP. Furthermore, the case GC 1 can be discriminated from the other GC types using HNR and η.

The boxplot diagrams in Fig. 7 show, that the symmetry property cannot be discriminated sufficiently with the evaluated features.

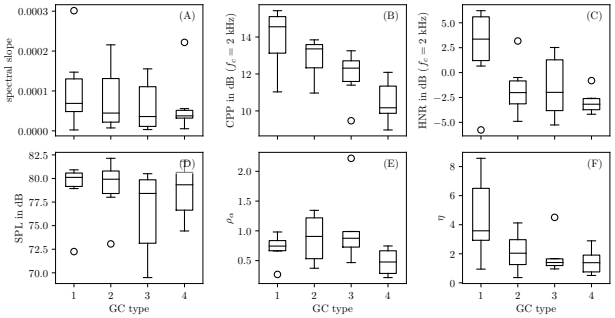

Figure 6: Boxplot diagrams of evaluated features and GC type.

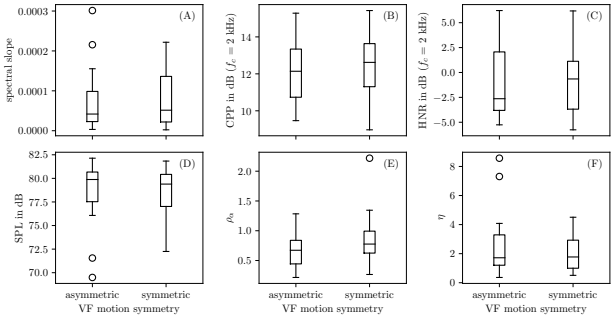

Figure 7: Boxplot diagrams of evaluated features and symmetry property.

Conclusion

The presented simulation-based investigation of regular or irregular oscillation characteristics aims to gain fundamental insights into the cause-effect chain of the human voice production mechanism. In summary, the following conclusions can be drawn: (I) Increasing glottal insufficiency causes a decrease in the CPP. (II) Increasing subglottal pressure correlates with increasing SPL and ρ_α. (III) The glottal symmetry cannot be discriminated based on the evaluated features. Future work could include a more detailed cluster analysis of the simulation configurations and the inclusion of more acoustic features aiming at the discrimination of the symmetry property.

References

[1] N. Roy, R. M. Merrill, S. Thibeault, R. A. Parsa, S. D. Gray, and E. M. Smith. Prevalence of voice disorders in teachers and the general population. *J. Speech Lang. Hear. Res.*, 47(2):281–293, 2004.

[2] S. Falk, S. Kniesburges, S. Schoder, B. Jakubaß, P. Maurerlehner, M. Echternach, M. Kaltenbacher, and M. Döllinger. 3D-FV-FE aeroacoustic larynx model for investigation of functional based voice disorders. *Front. Physiol.*, 12:226, 2021.

[3] S. Schoder, M. Weitz, P. Maurerlehner, A. Hauser, S. Falk, S. Kniesburges, M. Döllinger, and M. Kaltenbacher. Hybrid aeroacoustic approach for

the efficient numerical simulation of human phonation. *J. Acoust. Soc. Am.*, 147(2):1179–1194, 2020.

[4] P. Maurerlehner, S. Schoder, C. Freidhager, A. Wurzinger, A. Hauser, F. Kraxberger, S. Falk, S. Kniesburges, M. Echternach, M. Döllinger, and M. Kaltenbacher. Efficient numerical simulation of the human voice. *Elektrotech. Inftech.*, 138(3):219–228, 2021.

[5] R. C. Scherer, D. Shinwari, K. J. De Witt, C. Zhang, B. R. Kucinschi, and A. A. Afjeh. Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees. *J. Acoust. Soc. Am.*, 109(4):1616–1630, 2001.

[6] M. Kaltenbacher. *Numerical Simulation of Mechatronic Sensors and Actuators*. Springer, 2015.

[7] S. Schoder, P. Maurerlehner, A. Wurzinger, A. Hauser, S. Falk, S. Kniesburges, M. Döllinger, and M. Kaltenbacher. Aeroacoustic sound source characterization of the human voice production-perturbed convective wave equation. *Appl. Sci.*, 11(6), 2021.

[8] A. Wurzinger, S. Schoder, P. Maurerlehner, S. Falk, S. Kniesburges, M. Döllinger, and M. Kaltenbacher. Analysis of aeroacoustic sources in a hybrid simulation of human phonation using perturbed convective wave equation. In *DAGA*, Vienna, 2021.

[9] M. A. Belsky, S. D. Rothenberger, A. I. Gillespie, and J. L. Gartner-Schmidt. Do phonatory aerodynamic and acoustic measures in connected speech differ between vocally healthy adults and patients diagnosed with muscle tension dysphonia? *J. Voice*, 35(4):663.e1–663.e7, 2021.

[10] O. Garaycochea, J. M. A. Navarrete, B. del Río, and S. Fernández. Muscle tension dysphonia: Which laryngoscopic features can we rely on for diagnosis? *J. Voice*, 33(5):812.e15–812.e18, 2019.

[11] S. Kniesburges, A. Lodermeyer, M. Semmler, Y. K. Schulz, A. Schützenberger, and S. Becker. Analysis of the tonal sound generation during phonation with and without glottis closure. *J. Acoust. Soc. Am.*, 147(5):3285–3293, May 2020.

[12] R. Patel, D. Dubrovskiy, and M. Döllinger. Characterizing vibratory kinematics in children and adults with high-speed digital imaging. *J. Speech Lang. Hear. Res.*, 57(2):S674–S686, April 2014.

[13] Z. Zhang. Compensation strategies in voice production with glottal insufficiency. *J. Voice*, 33(1):96–102, 2019.

[14] P. Maurerlehner. Efficient FEM model of human phonation. Master's thesis, TU Wien, 2020.

[15] P. Woo, A. K. Parasher, T. Isseroff, A. Richards, and M. Sivak. Analysis of laryngoscopic features in patients with unilateral vocal fold paresis. *The Laryngoscope*, 126(8):1831–1836, 2016.

[16] S. S. Azar, P. Pillutla, L. K. Evans, Z. Zhang, J. Kreiman, and D. K. Chhetri. Perceptual evaluation of vocal fold vibratory asymmetry. *The Laryngoscope*, 131(12):2740–2746, 2021.

[17] F. Alias, J. C. Socoro, and X. Sevillano. A review of physical and perceptual feature extraction techniques for speech, music and environmental sounds. *Appl. Sci.*, 6(5), 2016.

[18] F. Eyben. *Real-time Speech and Music Classification by Large Audio Feature Space Extraction*. Springer, 2016.

[19] I. Mierswa and K. Morik. Automatic feature extraction for classifying audio data. *Mach Learn*, 58(2):127–149, 2005.

[20] C. Fang, H. Li, L. Ma, and M. Zhang. Intelligibility evaluation of pathological speech through multigranularity feature extraction and optimization. *Comput Math Method M*, 2017(2431573), 2017.

[21] S. Hegde, S. Shetty, S. Rai, and T. Dodderi. A survey on machine learning approaches for automatic detection of voice disorders. *J. Voice*, 33(6):947.e11–947.e33, 2019.

[22] L. Lopes, V. Vieira, and M. Behlau. Performance of different acoustic measures to discriminate individuals with and without voice disorders. *J. Voice*, 2020.

[23] F. Kraxberger. Understanding human voice disorders. Master's thesis, TU Graz, 2021.

[24] A. Wurzinger. Aeroacoustic simulation of flow parts in medical and automotive applications. Master's thesis, TU Wien, 2020.

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

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

[27] V. Birk, S. Kniesburges, M. Semmler, D. A. Berry, C. Bohr, M. Döllinger, and A. Schützenberger. Influence of glottal closure on the phonatory process in ex vivo porcine larynges. *J. Acoust. Soc. Am.*, 142(4):2197–2207, 2017.

[28] S. Patel, K. Scherer, J. Sundberg, and E. Björkner. Acoustic markers of emotions based on voice physiology. In *Proc. Speech Prosody*, volume 100865. ISCA, May 2010.

[29] J. Probst, A. Lodermeyer, S. Fattoum, S. Becker, M. Echternach, B. Richter, M. Döllinger, and S. Kniesburges. Acoustic and aerodynamic coupling during phonation in mri-based vocal tract replicas. *Appl. Sci.*, 9(17), 2019.