15th International Conference on Advances in
Quantitative Laryngology, Voice and Speech Research
March 30 – April 1, 2023
Mayo Clinic, Phoenix, Arizona

CHARACTERIZATION OF VOCAL MOTOR CONTROL USING LARYNGEAL KINEMATICS IN INDIVIDUALS WITH HYPERFUNCTIONAL VOICE DISORDERS

HASINI R. WEERATHUNGE, COURTNEY J. DUNSMUIR, SARAH J. COCROFT, MANUEL E. DÍAZ-CÁDIZ, CARA E. STEPP

Keywords: High-Speed Video Endoscopy; Vocal Motor Control; Laryngeal Kinematics; Vocal Hyperfunction

Introduction

Using high-speed video endoscopy techniques to estimate vocal fold kinematic measures in a population of individuals with and without hyperfunctional voice disorders (HVDs), two overarching research questions were investigated; 1) do vocal motor control strategies vary based on different target productions (i.e., different rate and effort conditions) and 2) does vocal motor control strategies for vocal production targets differ in individuals with and without HVDs.

Methods

Twenty-eight adults with HVDs and 28 age- and sex-matched controls produced repeated utterances of /ifi/, at varying speech rates (50, 65, and 80 beats-per-minute) and speaker-induced vocal effort (mild, moderate, and maximum effort). The glottic angle profiles of productions were extracted to calculate the kinematic stiffness ratio (i.e., ratio between maximum adductory velocity and maximum glottal angle), spatiotemporal index (i.e., cumulative sum of the standard deviations of the glottal angle trajectories), and asymmetry index (i.e., ratio between deceleration and acceleration durations of the glottal velocity profile) to objectively quantity vocal fold tension, production variability, and asymmetry, respectively.

Results

Results indicate that individuals with HVDs exhibit significantly higher kinematic stiffness ratios, spatiotemporal indices, and asymmetry indices across rates and effort conditions compared to controls, indicating higher vocal fold tension, production variability, and asymmetry. The magnitude and direction of change in measures across target productions also varied across the two populations, implying differential underlying vocal control strategies for different task requirements in individuals with and without HVDs.

Discussion and Conclusion

The study outcomes provide insight into the ability of laryngeal kinematics to differentiate underlying motor control strategies in individuals with HVD to guide future clinical intervention.

Acknowledgments

This work was supported by grant R01 DC016270 (Cara E. Stepp and Frank H. Guenther) from the National Institute of Deafness and Other Communication Disorders. It was also supported by a Graduate Fellow Award from the Rafik B. Hariri Institute for Computing and Computational Science and Engineering (Hasini Weerathunge) and an ASHFoundation New Century Doctoral Scholarship (Hasini Weerathunge).

15th International Conference on Advances in
Quantitative Laryngology, Voice and Speech Research
March 30 – April 1, 2023
Mayo Clinic, Phoenix, Arizona

The Effects of Voice Therapy Maneuvers on Vocal Fold Vibration in Older Adults

Samlan, Robin A, Monahan, Natalie, Ikuma, Takeshi, Smith, Dori, Zacharias, Stephanie, Kunduk, Melda

Keywords:

Note to Authors: Please use Times New Roman 11 font for content. Must be single page only.
The below section should be placed in two columns and include the following headers:

Introduction

Older adults report pitch, loudness, and voice quality changes with aging. While voice therapy is a common treatment for older adults with dysphonia, voice therapy techniques might work differently in older adults than younger adults, given typical age-related changes across the multiple systems. The aim of the current study was to assess the effects of four therapy maneuvers on vocal fold vibration after a short (approximately 3 minute) trial therapy period: 1) *Abdominal onset* 2) *Pulling* (forced laryngeal closure), 3) *Semi-occlusion* (/u/), and 4) *Assertive voice*.

Methods

Rigid high-speed videoendoscopy was completed at 4000 frames per second on sustained vowels /i/ or /u/. To date, 193 videos from 21 participants have been analyzed using a semi-automated segmentation system and harmonic-model based objective parameters. Fifty-four (54) of the productions were baseline, 40 abdominal onset, 30 semi-occlusion, and 34 assertive. Linear mixed model ANOVAs were completed for harmonic richness factor (*HRF*), relative glottal gap (*RGG*), open quotient (*OQ*), and fundamental frequency standard deviation (f_oSD).

Results

Significant main effects were present for HRF ($F(4,170)=4.539$, $p=0.002$), RGG ($F(4,171)=4.532$, $p=0.002$), and OQ ($F(4,169)=4.653$, $p<0.001$). There were no significant changes in f_oSD with the interventions. Results from the larger data set will be available for the AQL meeting.

Discussion and Conclusion

Findings will be interpreted relative to the vibratory changes that occur with specific techniques.

Acknowledgments

NIDCD R21DC016356

References

15th International Conference on Advances in
Quantitative Laryngology, Voice and Speech Research
March 30 – April 1, 2023
Mayo Clinic, Phoenix, Arizona

VOCAL FOLD DETECTIVE EDGE ANALYSIS IN HIGH-SPEED VIDEOENDOSCOPY DURING RUNNING SPEECH IN ADDUCTOR SPASMODIC DYSPHONIA

AHMED M. YOUSEF, DIMITAR D. DELIYSKI, MOHSEN ZAYERNOURI, STEPHANIE R.C. ZACHARIAS, MARYAM NAGHIBOLHOSSEINI

Keywords: Deep Learning; Image Analysis; Machine Learning; High-Speed Imaging; Voice Disorders

Introduction

Adductor spasmodic dysphonia (AdSD) is a neurological voice disorder that disrupts laryngeal muscle control in connected speech [1]. Coupling high-speed videoendoscopy (HSV) with a flexible endoscope, video data from the vocal fold (VF) vibration can be obtained in connected speech which can be beneficial to understand the disrupted vocal function in AdSD [2]. Previously, we developed several methods to study VF dynamics by segmenting the glottal area in HSV data during running speech to study normophonic and AdSD voices [3-6]. This work builds upon our prior contributions and provides an automated method to detect the left/right VF edges in HSV for examining VF vibrations in AdSD during connected speech.

Methods

A monochrome high-speed camera, coupled with a flexible nasolaryngoscope, was used for the data acquisition (recording rate of 4,000 frames per second and spatial resolution of 256x224 pixels). The video data were collected from four AdSD patients and three normophonic speakers during the reading of six CAPE-V sentences and part of the Rainbow Passage at Mayo Clinic, Scottsdale, AZ. The deep learning approach we previously developed [6] was utilized in this work to segment the glottal area in the HSV recordings. The first image moment of inertia was then computed for each row of the detected glottal area pixels (glottis center) where the midline was predicted as a fitted second-order curve. Accordingly, the spatial location of the left/right VF edges was automatically determined in the HSV images. The segmentation accuracy of VF edges was evaluated against visual/manual analysis.

Results

The developed approach accurately detected the right and left VF edges in the HSV dataset. Fig. 1 illustrates a sample of four automatically segmented frames showing the original images before and after segmentation, in the left and middle panels, respectively. The segmented area is highlighted in red in the middle panels. The detected left and right VF edges based on the image moments are indicated by the cyan and magenta lines in the right panels.

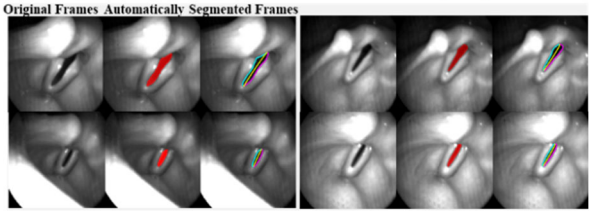

Fig. 1. Results of segmenting four HSV frames and detecting the right and left edges of the VFs.

Discussion and Conclusion

The proposed edge detection method was accurate even with significant image noises and excessive laryngeal movements in AdSD in running speech. This method enables developing HSV-based measures for assessing VF function in AdSD.

Acknowledgments

We would like to acknowledge NIH NIDCD K01DC017751, R21DC020003, and the MSU Discretionary Funding Initiative.

References

[1] Roy, N. et al., 2005. Task specificity in adductor spasmodic dysphonia versus muscle tension dysphonia. *Laryngoscope,* 115(2), pp. 311-316.

[2] Yousef, A. M. et al., 2022. Detection of Vocal Fold Image Obstructions in High-Speed Videoendoscopy During Connected Speech in Adductor Spasmodic Dysphonia: A Convolutional Neural Networks Approach. *J Voice*; S0892-1997(22)00027-3.

[3] Yousef, A. M. et al., 2022. A Deep Learning Approach for Quantifying Vocal Fold Dynamics during Connected Speech using Laryngeal High-Speed Videoendoscopy. *JSLHR,* 65(6), pp. 2098-2113.

[4] Yousef, A. M. et al., 2021. A hybrid machine-learning-based method for analytic representation of the vocal fold edges during connected speech. *Appl. Sci.,* 11(3), pp. 1179.e1-1179.e15.

[5] Yousef, A. M. et al., 2023. Spatial segmentation for laryngeal high-speed videoendoscopy in connected speech. *J Voice,* 37(1), pp. 26-36.

[6] Yousef, A. M. et al., 2022. Deep-learning-based representation of vocal fold dynamics in adductor spasmodic dysphonia during connected speech in high-speed videoendoscopy. *J Voice;*S0892-1997(22)00263-6

15th International Conference on Advances in
Quantitative Laryngology, Voice and Speech Research
March 30 – April 1, 2023
Mayo Clinic, Phoenix, Arizona

Experimental investigation of source-filter interaction in a synthetic larynx model

Christoph Näger, Stefan Kniesburges, Stefan Becker

Keywords: Source-filter interaction, synthetic larynx model, laser-Doppler-vibrometry

Introduction

The human voice is generated in a complex interaction of fluid flow, structural oscillation and acoustic waves. This process has long been described with a linear source filter theory that contains no back coupling of the vocal tract's (VT) acoustic properties onto the vocal fold (VF) oscillation and glottal flow. There are however conditions where this is not valid, e.g. when a maximum in the frequency dependent VT input impedance is located closely to the oscillation frequency of the VFs [1]. In this work, we use a synthetic larynx model to investigate source filter interaction in the human voice.

Methods

We employ a VT of variable length to tune its acoustic properties to the oscillation frequency of the synthetic VFs. We perform laser-Doppler-vibrometry measurements to measure the surface oscillation velocity of the VFs. To investigate the relationship between the supraglottal flow field and VT acoustic properties, the time resolved, static pressure in the supraglottal channel is measured via pressure transducer. The acoustic properties of the VT are determined via transmission line model [2].

Results

The vibrometry measurements show a clear connection between the oscillation frequency f_o and the impedance maximum frequency f_{Z1}. If in vicinity to each other, the oscillation frequency tuned itself to the impedance maximum. No interaction between impedance maxima and higher harmonics $2f_o, 3f_o, \ldots$ could be found however. In contrast to this, the pressure data suggests a complex interaction between impedance maxima and the supraglottal flow field. The oscillation frequency shift could also be seen in the pressure data. However, there is also a clear interaction between higher harmonics and VT input impedance visible. We hypothesize, that this interaction does not change the main characteristics of the pulsating glottal jet flow. Instead, it results in a changed jet stability behavior resulting from changes in supraglottal vortices.

Discussion and Conclusion

We see a clear interaction between the acoustical properties of the VT and the oscillation of the VFs. Further investigations with the help of laser scanning vibrometry and particle image velocimetry have to be made to confirm these findings. In addition, a theoretical explanation of the observed physical behavior is necessary to enhance understanding of the observed source-filter interaction phenomena.

Acknowledgments

This research is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under project no. 446965891

References

[1] I. R. Titze: Nonlinear source-filter coupling in phonation: Theory, J. Acoust. Soc. Am., 2008

[2] M. Sondhi, J. Schroeter: A hybrid time-frequency domain articulatory speech synthesizer, IEEE Transactions on Acoustics, Speech, and Signal Processing (1987)

15th International Conference on Advances in
Quantitative Laryngology, Voice and Speech Research
March 30 – April 1, 2023
Mayo Clinic, Phoenix, Arizona

Evaluation of the One-Dimensional Glottal Flow Model in Predicting Voice Outcomes of Left-Right Asymmetric Vocal Fold Vibrations

Tsukasa Yoshinaga[1], Zhaoyan Zhang[2], Akiyoshi Iida[1]
[1]Toyohashi University of Technology, [2]University of California Los Angeles
Keywords: Vocal folds, Aeroacoustics, Flow model, Two-mass model.

Introduction

In voice production simulations, one-dimensional (1D) flow models are often used to reduce computational costs [e.g., 1-2]. However, the airflow around the oscillating vocal folds has been shown to exhibit three-dimensional (3D) phenomena [3], and it is unclear how accurate 1D flow models are in modeling fluid-structure interaction under left-right asymmetric vocal fold conditions. In this study, we compared voice production outcomes predicted by 1D and 3D flow models and evaluated the accuracy of 1D flow models in left-right asymmetric vocal fold conditions.

Methods

To focus on the flow models and avoid the high computational costs associated with modeling 3D vocal folds, both the 1D and 3D flow models were coupled to a two-mass model [1]. The 3D flow was simulated by solving the compressible Navier-Stokes equations with a high-order accuracy finite difference method, whereas the 1D flow was modeled by Bernoulli's equation and equivalent circuits. Simulations were performed for different values of the ratio of left-right tension imbalance Q. Details are reported in [4].

Results and Discussion

The maximum flow declination rate (MFDR) was calculated for both 1D and 3D flow models and plotted for different tension imbalance conditions in Fig. 1. The MFDRs predicted by the 1D model agreed well with those of the 3D model, indicating that the 1D flow simplification is sufficient in modeling phonation in the range of left-right asymmetric conditions. This agreement was obtained despite mismatches in intraglottal pressure distributions, indicating that vocal fold properties play a larger role than the glottal flow.

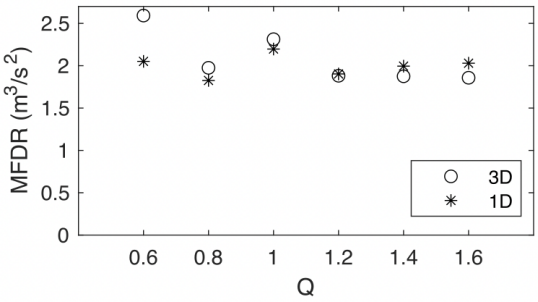

Fig. 1. Maximum flow declination rates with different left-right tension imbalance parameter Q.

Acknowledgments

This work was supported by MEXT as JSPS KAKENHI (Grant No. JP20K14648) and the National Institute on Deafness and Other Communication Disorders, the National Institutes of Health (Grant No. R01DC001797).

References

[1] Ishizaka, K., Flanagan, J. L. (1972). Bell Syst. Tech. J. 51(6), 1233–1268.
[2] Farahani, M. H., Zhang, Z. (2014). J. Acoust. Soc. Am. 136(5), EL369–EL375.
[3] Neubauer, J., Zhang, Z., Miraghaie, R., Berry, D. A. (2007). J. Acoust. Soc. Am. 121(2), 1102–1118.
[4] Yoshinaga, T., Zhang, Z., Iida, A. J. Acoust. Soc. Am. 152(5), 2557–2569.

15th International Conference on Advances in
Quantitative Laryngology, Voice and Speech Research
March 30 – April 1, 2023
Mayo Clinic, Phoenix, Arizona

DESCRIBING NON-PHONOTRAUMATIC VOCAL HYPERFUNCTION WITH AN ASYMMETRIC TRIANGULAR BODY-COVER MODEL WITH UNBALANCED INTRINSIC MUSCLE ACTIVATION

Calvache Carlos[1,2,3], Parra Jesús A.[4], Soláque Leonardo[2], Matías Zañartu[4]

[1] Department Communication Sciences and Disorder, Corporación Universitaria Iberoamericana, Bogotá, Colombia

[2] Department of Mechatronics Engineering, Universidad Militar Nueva Granada, Bogotá, Colombia

[3] Vocology Research, Vocology Center, Bogotá, Colombia

[4] Department of Electronic Engineering, Federico Santa María Technical University, Valparaíso, Chile

Keywords: Vocal fold modeling, muscle activation, MTD, Vocal Hyperfunction

Introduction

The effective clinical management of non-phonotraumatic vocal hyperfunction (NPVH) is hampered by the limited understanding of the role that numerous potential factors play in its etiology and pathophysiology[1]. This study aims to represent key biomechanical features of NPVH, through an asymmetric triangular body-cover model (TBCM)[2] with unbalanced intrinsic muscle activation.

Methods

Four pathophysiological and biomechanical features were selected to explore NPVH: (1) Increased Subglottal Pressure (SP), (2) Amplitude Asymmetry (AA), (3) Phase Asymmetry (PA) and (4) Arytenoid Adduction Asymmetry (AAA). Each feature was represented in the model, considering the contribution of the intrinsic laryngeal muscles, with a muscle imbalance factor between left and right vocal folds.

Results

Model behavior is presented in relation to each of the four selected features, according to the asymmetry factor in the muscle activations. Elevated SP and AAA are obtained through an increased activation of the Posterior Cricoarytenoid (PCA) muscle, which results in an incomplete glottal closure. Elevated AA and PA are obtained in the model through an unbalanced activation of Thyroarytenoid (TA) muscle.

Discussion and Conclusion

The aerodynamic and mechanical behavior reported in the literature agrees with the initial pilot results of this study. Thus, it is possible to represent features of NPVH through the asymmetric TBCM with muscle unbalance.

Acknowledgments

Research funded by the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health under grant P50DC015446, ANID under grants BASAL FB0008, FONDECYT 1191369, and Corporación Universitaria Iberoamericana under search grant 202210D038.

References

[1] R. E. Hillman, et al., “An updated theoretical framework for vocal hyperfunction,” *JSLP.*, vol. 29, no. 4, pp. 2254–2260, 2020.

[2] G. A. Alzamendi, et al., “Triangular body-cover model of the vocal folds with coordinated activation of the five intrinsic laryngeal muscles,” *JASA,* vol. 151, no. 1, pp. 17–30, 2022.

15th International Conference on Advances in
Quantitative Laryngology, Voice and Speech Research
March 30 – April 1, 2023
Mayo Clinic, Phoenix, Arizona

AN ASYMMETRIC TRIANGULAR BODY-COVER MODEL WITH UNBALANCED INTRINSIC MUSCLE ACTIVATION

Parra Jesús A.[1], Calvache Carlos[2,3,4], Matías Zañartu[1]

[1] Department of Electronic Engineering, Federico Santa María Technical University, Valparaíso, Chile

[2] Department Communication Sciences and Disorder, Corporación Universitaria Iberoamericana, Bogotá, Colombia

[3] Department of Mechatronics Engineering, Universidad Militar Nueva Granada, Bogotá, Colombia

[4] Vocology Research, Vocology Center, Bogotá, Colombia

Keywords: Vocal fold modeling, muscle activation, asymmetry oscillation.

Introduction

Previous efforts representing asymmetric vocal fold oscillation in lumped-element models have described the physics by introducing left-right mass and spring differences[1]. The triangular body cover (TBCM) provides posture and viscoelasticity control for the vocal folds, through the independent activation of all intrinsic laryngeal muscles. The present study provides an adaptation of TBCM for asymmetric vocal fold vibration through an unbalanced left-right muscle activation (a-TBCM).

Methods

The a-TBCM features independent left-right parameters according to activation rules of the TBCM, a new asymmetric vocal fold collision description, and a revised formulation of the glottal area waveform. We explore the proposed asymmetric description for various muscles and compare these results against those from previous spring-mass asymmetric descriptions.

Results

The behavior of the masses, springs and glottal angle is studied as the ratio between left and right cricoarytenoid (CT), thyroarytenoid (TA) and lateral cricoarytenoid (LCA) activation. For CT/TA ratio variation, the vocal process remains constant, but the viscoelasticity changes. On the other hand, for LCA ratio variation, the viscoelastic properties are maintained, but the vocal process changes.

Discussion and Conclusion

The proposed a-TBCM scheme allows for introducing differences in viscoelasticity and posture using the ratio of muscle activations. Future work will focus on constructing rules for the ratio between the activations of the folds, which would allow for representing various vocal fold pathologies.

Acknowledgments

Research funded by the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health under grant P50DC015446 and ANID under grants BASAL FB0008 and FONDECYT 1191369.

References

[1] D. D. Mehta, et al. “Investigating acoustic correlates of human vocal fold vibratory phase asymmetry through modeling and laryngeal high-speed videoendoscopy,” *JASA*, vol. 130, no. 6, pp. 3999–4009, 2011.

15th International Conference on Advances in
Quantitative Laryngology, Voice and Speech Research
March 30 – April 1, 2023
Mayo Clinic, Phoenix, Arizona

MACHINE LEARNING BASED ESTIMATION OF HOARSENESS SEVERITY USING SUSTAINED VOWELS RECORDED DURING HIGH-SPEED VIDEOENDOSCOPY

Tobias Schraut[1], Anne Schützenberger[1], Tomás Arias-Vergara[1], Melda Kunduk[2], Matthias Echternach[3], Michael Döllinger[1]

[1] Division of Phoniatrics and Pediatric Audiology, Department of Otorhinolaryngology, Head and Neck Surgery, University Hospital Erlangen, Germany

[2] Department of Communication Sciences and Disorders, Louisiana State University, USA

[3] Division of Phoniatrics and Pediatric Audiology, Department of Otorhinolaryngology, Head and Neck Surgery, University Hospital Munich (LMU), Germany

Keywords: high-speed videoendoscopy, voice disorder, acoustic signal, machine learning

Introduction

High-speed videoendoscopy (HSV) is a promising method for quantitative assessment of voice disorders. The addition of a synchronous acoustic recording to HSV allows simultaneous analysis of vocal fold movements and the acoustic voice signal. Acoustic assessment of voice impairment is commonly performed by subjective perceptual evaluation of continuous speech based on a grading system such as the roughness, breathiness, hoarseness (RBH) scale. Here, we present an automatic machine learning based approach to objectively quantify hoarseness severity based on sustained vowels recorded during HSV.

Methods

Acoustic recordings of the sustained vowel /i/ were collected during high-speed videoendoscopy. Temporal, spectral and cepstral features were extracted from 250ms of each recording. All recordings have an assigned RBH score (R, B, H ϵ [0,1,2,3], H = max(R, B)), which was determined subjectively by an expert based on continuous speech of the respective subject. In order to account for the label noise introduced by different valuation bases, subjects were divided into two levels of hoarseness $H < 2$ (normal, mild) and $H \geq 2$ (moderate, severe). Logistic Regression was employed as classification model, using the resulting output probabilities as continuous severity rating. Relevant features were selected using a sequence of filter methods and backward elimination.

Results

Results of severity estimation were evaluated in terms of classification performance as well as correlation between output probabilities and the subjective H score.

Discussion and Conclusion

The presented method describes a promising approach for objective acoustic grading of hoarseness which, together with HSV analysis, allows multimodal quantification of treatment progress in patients with voice disorders [1].

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) through project numbers DO 1247/8-2 and SCHU 3441/3-2.

References

[1] Schlegel P. (2020). Assessment of clinical voice parameters and parameter reduction using supervised learning approaches. (Dissertation, Universität Erlangen-Nürnberg). Shaker Verlag, Düren. DOI: 10.2370/9783844074352.

15th International Conference on Advances in
Quantitative Laryngology, Voice and Speech Research
March 30 – April 1, 2023
Mayo Clinic, Phoenix, Arizona

GIVING STATISTICAL TITLE BACK TO MACHINE LEARNING: STATISTICAL POWER ANALYSIS AND STATISTICAL MODEL CONFIDENCE OF MACHINE LEARNING

Hamzeh Ghasemzadeh, Robert E. Hillman, Daryush D. Mehta

Keywords: Machine learning, cross-validation, statistical significance, power analysis, sample size

Introduction

The application of machine learning in speech, language, and hearing sciences has increased significantly over the last decade. However, many prior studies have adopted the single holdout cross-validation method. The main aim of this study was to quantify the effect of different cross-validation methods on the statistical characteristics of the model found by machine learning.

Methods

A generative paradigm was used to systematically create datasets with different sample sizes, different dimensionalities, and different discriminative power of features. The generated datasets were then passed through a machine learning processing pipeline that included a wrapper-forward feature selection. Four major cross-validation methods of single holdout, 10-fold, training-validation-test, and nested 10-fold were implemented. Distributions of the null and alternative hypotheses were estimated from each scenario using the Monte Carlo method. The required sample size for obtaining a statistically significant outcome (α=0.05, β=0.2) was estimated separately for each cross-validation method. Additionally, the statistical confidence of the model was defined and calculated as the probability of correct features getting selected and hence being included in the final model. The effect of different cross-validation methods on the statistical confidence of the model was quantified.

Results

Our analysis showed that the accuracy of the model generated based on the single holdout had a significant bias, very low statistical power, and very low statistical confidence. Conversely, not only the estimated accuracy of nested 10-fold cross-validation was unbiased, but it also offered much higher statistical confidence and statistical power. As an example, the single holdout may need as high as 50% more samples to achieve the same statistical power as the nested cross-validation. Also, the confidence in the model generated using a single holdout could be as low as 1/4 of the model generated based on nested cross-validation.

Discussion and Conclusion

Cross-validation is an essential part of supervised machine learning and is the component that can prevent overfitting and governs the generalization of the results if used appropriately. This study quantified the statistical power and statistical confidence of the model generated based on different cross-validations and showed significant differences between the performance of different cross-validation methods. Based on our findings future studies are advised to adopt the unbiased and more robust method of nested cross-validation.

Acknowledgments

This research was funded by the National Institutes of Health - National Institute on Deafness and Other Communication Disorders (grants: T32 DC013017, P50 DC015446).

15th International Conference on Advances in
Quantitative Laryngology, Voice and Speech Research
March 30 – April 1, 2023
Mayo Clinic, Phoenix, Arizona

Prediction of Breathy Voice Quality by Frequency-Segmented Spectral Power Levels

Takeshi Ikuma, PhD, Andrew J. McWhorter, MD, Melda Kunduk, PhD

Keywords: acoustic analysis, breathiness correlates, multiple linear regression

Introduction

Majority of spectral voice parameters either employ the full spectral information (e.g., harmonics to noise ratio and cepstral peak prominence, CPP) or compare two harmonics (e.g., H1-H2 and H1-A1)[1]. This presentation demonstrates the effectiveness of measuring the harmonic and noise power levels in smaller frequency segments and recombining them to quantify breathy perceptual vocal quality. The segmentation allows different regions of the frequency spectrum to be related independently to perceptual breathiness and to each other.

Methods

Sustained /a/ recordings of speakers with voice disorders were sampled from Saarbrueken Voice Database. Four speech language pathologists perceptually rated the recordings for breathiness on a 100-point scale, and their averages were used as the ratings of the recordings. Only the speakers without any roughness in voice were included in the study ($n = 88$). The acoustic spectra ($f_{\max} = 4$ kHz) were analyzed with four different sets of frequency segmentation rules: (1) a single band, (2) equispaced bands in linear frequency, (3) equispaced bands in logarithmic frequency, and (4) formant-based spacing [2]. Harmonic and noise power of each band was evaluated using time-varying harmonic model [3], and multiple linear regression analyses were conducted to predict the perceptual rating by the spectral power levels and f_o in different frequency configurations.

Results

The variation in breathiness rating was poorly predicted by the single-band configuration ($R^2_{\text{adj}} = 54\%$) compared to the 5-segment logarithmic configuration (84%) and the formant-based rule (85%). The 5-segment linear configuration resulted in $R^2_{\text{adj}} = 73\%$.

Discussion and Conclusion

The log and formant-tuned spacing were the best suited approaches to capture the breathiness by segmented spectral power measures. The results align well with the known effectiveness of the CPP and psychoacoustic-based correlates of breathiness. These proposed measures could be attractive input to more complex machine learning systems for predicting voice qualities.

References

[1] B. Barsties v. Latoszek, Y. Maryn, E. Gerrits, and D. B. Marc, "A meta-analysis: acoustic measurement of roughness and breathiness," *J. Speech Lang. Hear. Res.*, vol. 61, no. 2, pp. 298–323, Feb. 2018, doi: 10.1044/2017_JSLHR-S-16-0188.

[2] G. de Krom, "Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments," *J. Speech Hear. Res.*, vol. 38, no. 4, pp. 794–811, Aug. 1995.

[3] T. Ikuma, B. Story, A. J. McWhorter, L. Adkins, and M. Kunduk, "Harmonics-to-noise ratio estimation with deterministically time-varying harmonic model for pathological voice signals," *J. Acoust. Soc. Am.*, vol. 152, no. 3, pp. 1783–1794, Sep. 2022, doi: 10.1121/10.0014177.

15th International Conference on Advances in
Quantitative Laryngology, Voice and Speech Research
March 30 – April 1, 2023
Mayo Clinic, Phoenix, Arizona

Simultaneous Segmentation and Sub-Pixel Localization in Structured Light High-Speed Video Endoscopy

Jann-Ole Henningson, Marion Semmler, Michael Döllinger, Marc Stamminger

Keywords: Keypoint Detection, Semantic Segmentation, Structured Light, High-Speed Video

Note to Authors: Please use Times New Roman 11 font for content. Must be single page only.
The below section should be placed in two columns and include the following headers:

Introduction - Malign changes of human vocal folds are conventionally observed by the use of (high-speed) video endoscopy that measures their 2D deformation in image space. However, it was shown that their dynamics contain a significant vertical deformation. This led to the development of varying methods for the 3D reconstruction of human vocal folds during phonation. In these works, especially active reconstruction methods that project a pattern onto the surface of vocal folds have been researched [1-3]. These pipelines initially predict a segmentation of the vocal folds and detect 2D pixel positions of the projected pattern. However, currently used segmentation methods either need manual input or were not developed with an additional light source in mind. This renders them either infeasible for clinical use or susceptible towards artifacts introduced by the structured light projector.

Methods - To address this, we present a method for the simultaneous segmentation, detection, and tracking of human vocal folds and speckle projection patterns in structured light high-speed video endoscopy. The proposed method is based on two stages: First, a convolutional neural network predicts a segmentation of the vocal folds, the glottal area and projected speckles in a temporal-coherent manner. Secondly, we compute sub-pixel accurate 2D point locations based on the pixel-level speckle class probabilities via gaussian fitting.

Results - Our approach drastically improves the robustness, accuracy and applicability of current human vocal fold 3D reconstruction pipelines.

Discussion and Conclusion – Our proposed method can easily be integrated into existing vocal fold reconstruction pipelines, supporting physicians by providing immediate feedback about the reconstruction process.

Acknowledgments - This work was supported by Deutsche Forschungsgemeinschaft (DFG) under grant STA662/6-1 (DFG project number: 448240908).

References :

[1] – Real-Time 3D Reconstruction of Human Vocal Folds via High-Speed Laser-Endoscopy, Henningson et al.

[2] – Three-dimensional optical reconstruction of vocal fold kinematics using high-speed videomicroscopy with a laser projection system, Luegmair et al.

[3] - 3d reconstruction of human laryngeal dynamics based on endoscopic high-speed recordings, Semmler et al.

15th International Conference on Advances in
Quantitative Laryngology, Voice and Speech Research
March 30 – April 1, 2023
Mayo Clinic, Phoenix, Arizona

SHORT-TERM PREDICTION OF DISEASE SYMPTOMS SEVERIT OF PARKINSON'S DISEASE

DARIA HEMMERLING[1], MAGDALENA WOJCIK-PEDZIWIATR[2,3], MAREK WODZINSKI[1], MATEUSZ DANIOL[1], PAWEL JEMIOLO[1]

[1] AGH University of Science and Technology, Faculty of Electrical Engineering, Automatics, Computer Science, and Biomedical Engineering, Krakow, Poland
[2] Department of Neurology, The John Paul II Hospital, Krakow, Poland
[3] Department of Neurology, Andrzej Frycz Modrzewski Krakow University, Poland

Keywords: signal processing, vocal biomarkers, neurodegenerative diseases, disease monitoring

Introduction

Prognosis in medicine is an extremely important aspect in planning a patient's therapy and in assessing prognosis. The main purpose of this study was to conduct research aimed at determining to what extent changes in the speech signal, detectable and quantifiable on the basis of phonetic-acoustic analyses of this signal, can be used as a source of prognostic information in Parkinson's disease. For this purpose, 150 recordings were collected from 30 patients at JP2 Hospital in Krakow, Poland. Voice recordings were made at well-defined moments of time in the "off" state, 30-, 60-, 120- and 180-minutes after taking the drug levodopa. Each patient uttered 5 specific vowels: /a/, /e/, /i/, /o/, /u/. At each measurement point, patients were examined by a physician and the UPDRS scale, Part III corresponding to motor changes (which has a range of 0 – 108 points), was determined.

Methods

Each speech signal was described by more than 600 acoustic features describing tone, energy, spectral features, frequency, amplitude, mel-cepstral features, harmonic to noise ratio, among others. Based on the created vector of acoustic features regression was performed to determine UPDRS-III scores. The regression was performed using the RANdom SAmple Consensus (RANSAC), Bayesian Ridge Regression, Random Forest Regression and XGBoost Regressor.

Results

The Mean Squared Error between predicted and neurologist-determined scores is about 4 points with XGBoost regressor. The results obtained should be considered promising. Based on the presented research methodology, it is possible to automate the assessment of a patient's disease progression at current moment on the UPDRS-III scale and reduce the examination time. Such tool could be also used in home-monitoring, bringing high value to more often measurements.

Discussion and Conclusion

The results of the study indicate the great success of using acoustic analysis in monitoring the course of Parkinson's disease and obtaining an objective measures to assess its severity at a given time point. For future works, we plan to combine voice signals with other sensors to measure motor changes and eye gaze with augmented reality technology.

Acknowledgments

The project was funded by The National Centre for Research and Development, Poland under Lider Grant no: LIDER/6/0049/L-12/20/NCBIR/ 2021.

15th International Conference on Advances in
Quantitative Laryngology, Voice and Speech Research
March 30 – April 1, 2023
Mayo Clinic, Phoenix, Arizona

Quantitative Analysis of Swallowing Using Time-Dependent Bolus Tracking

Luisa Neubig, Melda Kunduk, Deirdre Larsen, Rebecca Leonard, Andreas M. Kist

Keywords: Deep neural networks, Semantic Segmentation, Swallowing, Kinematics

Introduction

Videofluoroscopic swallow studies are a common diagnostic tool, where a patient is given to swallow radiopaque material to determine the changes in swallowing physiology. Segmentation and tracking of this so-called bolus are key to the objective evaluation of swallowing [1].

Methods

Deep neural networks (DNNs) were trained on a dataset of patients not suffering from dysphagia to discriminate bolus material from the background. We used a modification of the U-Net architecture [2] featuring Instance Normalization. The bolus was defined as the main material moving from the oral cavity to the esophagus. Segmentation quality was assessed using the Dice coefficient (DC), ranging from 0 (poor) to 1 (excellent). The bolus was further examined for its shape and main path (trajectory) during swallowing events.

Results

We performed a systematic analysis to determine the hyperparameters, identifying the best parameter setting with an overall DC of 0.86 on the validation and 0.84 on the test dataset. The segmented area was used for quantifications, particularly for bolus shape and kinematics analysis. The eccentricity of the bolus area over time showed a correlation between bolus shape and the swallowing phase. High eccentricity values were measured in the oral and pharyngeal phases, indicating a more elongated bolus. After fitting an exponential function to the moving bolus centroids, we observed a slightly upward-bent curve for normal swallowing. Our trajectory analysis is consistent across individual swallows and subjects.

Discussion and Conclusion

DNNs allow high-quality bolus segmentation in previously unseen data. Downstream analysis of bolus shape and kinematics resembles swallowing phases and clearance behavior and is important for further studies in patient collectives.

Acknowledgments

None.

References

[1] Leonard, R. (2017). Two methods for quantifying pharyngeal residue on fluoroscopic swallow studies: Reliability assessment. Annals of Otolaryngology and Rhinology, 4(3), 1168.

[2] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18 (pp. 234-241). Springer International Publishing.

15th International Conference on Advances in
Quantitative Laryngology, Voice and Speech Research
March 30 – April 1, 2023
Mayo Clinic, Phoenix, Arizona

Applying bioinformatic methods to temporality quantification of swallowing kinematics in ultrasonographic examinations

Lam, W.Y.S., Kwong, E., Chan, H.W.T.

Keywords: swallowing kinematics, ultrasonography, information theory, ordinality, variability, concurrency

Introduction

Swallowing is a process involving complex movement coordination. In particular, kinematic events across and within the oral and pharyngeal phases are overlapping, and previous studies had identified individual variability in their temporal sequence. The current study aimed to explore the feasibility of applying existing bioinformatic analyses to the study of ordinality, variability, and concurrency (i.e. overlaps) of swallowing kinematic events.

Methods

Forty-three participants, including 17 healthy adults, 15 healthy elderly, and 11 dysphagic individuals, received ultrasonographic examinations (B-mode) of swallowing on the mid-sagittal plane of the submental region, while performing swallowing trials of dry swallows and various water bolus types with reference to the International Dysphagia Diet Standardization Initiative (IDDSI). Three swallowing movements, namely, tongue base retraction (TB), geniohyoid muscle contraction (GH), and hyoid bone excursion (HB), were of interest. Key frames and the order of the event onset, maximum, and offset were extracted. Longest Common Subsequence, Shannon's Entropy, and Overlap Coefficient were used to analyze the ordinality, variability, and concurrency of swallowing events respectively.

Results

For ordinality, only four non-trivial obligatory event pairs, but not multiple event sequences, were identified. For variability and concurrency, statistical analyses revealed significant differences in the entropy of swallowing sequences ($F(2, 343)=56.195$, $p < 0.001$), and the overlapping onset-to-maximum and maximum-to-offset movements duration. Post hoc analyses confirmed significant differences among healthy adults, elderly, and the dysphagics.

Discussion and Conclusion

The findings will be discussed with reference to previous studies of swallowing sequence using videofluoroscopy (e.g [1]) and ultrasonography (e.g. [2]). The current study also suggested the feasibility and plausibility of using ultrasonography and the modality-independence of applying existing bioinformatic methods to study/examine swallowing kinematic sequence.

References

1. Molfenter SM, Leigh C, Steele CM. Event Sequence Variability in Healthy Swallowing: Building on Previous Findings. Dysphagia. 2014;29(2):234–42.

2. Kwong E, Shek PTC, Leung MT, Zheng, YP, Lam, WYS. Temporal measures of oropharyngeal swallowing events identified using ultrasound imaging in healthy young adults. Plos one. 2022; 17(6):e0270704

15th International Conference on Advances in
Quantitative Laryngology, Voice and Speech Research
March 30 – April 1, 2023
Mayo Clinic, Phoenix, Arizona

QUANTIFYING AMBULATORY VOICING-RESTING BEHAVIORS REVEALS SUBTYPES IN PATIENTS WITH PHONOTRAUMATIC VOCAL HYPERFUNCTION

Hamzeh Ghasemzadeh, Robert Hillman, Andrew Ortiz, Jarrad Van Stan, Daryush Mehta

Keywords: Voice disorder subtyping, phonotraumatic vocal hyperfunction, ambulatory voice monitoring, machine learning, voicing-resting behavior.

Introduction

The classic profile of patients with phonotraumatic vocal hyperfunction (PVH) is moderate-to-high voice use, high social potency (extraversion), and low impulse control. This study combines the power of a novel measure of ambulatory voicing-resting behavior and unsupervised machine learning and presents preliminary results showing that not all patients with PVH fit this classic profile.

Methods

An ambulatory voice monitoring device was used to record the neck-surface acceleration signals over approximately seven days from 116 female patients diagnosed with PVH and 116 age-, sex, and occupation-matched vocally healthy controls. The recordings were classified as voicing and resting (non-voiced) segments using a voice activity detector, and then voiced segments that were separated by less than 0.5 seconds (to account for pauses) were merged into a single voice phrase segment. A singing detector was applied on each voice phrase segments and only non-singing (i.e., speech) segments were retained for the rest of the analysis. Voicing-resting behavior of each participant was quantified using the voicing-resting ratio (VRR), defined as the duration of a speech phrase segment divided by the sum of the durations of the preceding and succeeding non-voiced segments. The distribution of the 15$^{th.}$ percentile of patients' ambulatory VRR (VRR15) exhibited a bimodal distribution and divided patients into two subtypes using unsupervised Gaussian mixture modeling. The two subtypes were then characterized using the Daily Phonotrauma Index (DPI) and a personality questionnaire and compared against the matched control group.

Results

Patients were clustered into two subtypes consisting of 83 (high VRR15) and 33 (low VRR15) individuals. The DPI classified the subtype with high VRR15 as patients; however, the subtype with low VRR15 was more likely to be classified as controls. Personality analysis showed that patients with high VRR15 exhibited the classic profile of higher social potency and lower impulse control than controls; in contrast, patients with low VRR15 exhibited no significant difference in these personality metrics relative to controls.

Discussion and Conclusion

Subtyping is an important step toward precision medicine, as it allows an inhomogeneous group of patients with a certain diagnosis to be divided into several more homogenous subtypes, with possibly different etiologies. This study provided preliminary results indicating the presence of at least two PVH subtypes. The first subtype followed the classical profile of PVH and was significantly different from their matched controls.

Acknowledgments

This research was funded by the National Institutes of Health - National Institute on Deafness and Other Communication Disorders (grants: T32 DC013017, P50 DC015446).

15th International Conference on Advances in
Quantitative Laryngology, Voice and Speech Research
March 30 – April 1, 2023
Mayo Clinic, Phoenix, Arizona

Ambulatory Monitoring of Subglottal Pressure Estimated from Neck-Surface Vibration in Individuals with and without Voice Disorders

Juan P. Cortés, Jon Z. Lin, Katherine L. Marks, Víctor M. Espinoza, Emiro J. Ibarra, Matías Zañartu, Robert E. Hillman, Daryush D. Mehta.

Keywords: Subglottal pressure; clinical voice assessment; accelerometer; ambulatory voice monitoring.

Introduction

This study, recently published as [1], reports on the development and validation of a method for subglottal pressure (Ps) estimation that can be applied to monitoring of vocal function and behavior in ambulatory settings. Ps plays a major role in voice production and has been shown to differentiate individuals with typical voices from those with voice disorders and to act as a clinical outcome measure. Therefore, the estimation of Ps during daily monitoring could provide better insights into the etiology and pathophysiology of voice disorders.

Methods

Subject-specific multiple regression models were derived for each subject to estimate Ps from running speech. The input to the model consisted of a sparse combination of raw neck-surface accelerometer features, as well as from glottal airflow estimates using an impedance-based inverse filter. The model was then applied to ambulatory recordings during activities of daily living in patients diagnosed with phonotraumatic vocal fold lesions, primary muscle tension dysphonia, and unilateral vocal fold paralysis.

Results

The proposed method exhibited significantly lower error than alternative Ps estimation methods in the literature, with average errors ranging from 1.13 to 2.08 cm H_2O for the participant groups. Application of the model to day-long recordings provided for the first ambulatory estimates of subglottal pressure in these populations.

Discussion and Conclusion

Results point to the feasibility and potential of real-time monitoring of subglottal pressure during an individual's daily life for improved prevention, assessment, and treatment of voice disorders. Further studies will investigate changes in the estimation of Ps after patients undergo voice therapy and/or surgery.

Acknowledgments

We thank Drs. James A. Burns and Tiffiny A. Hron whose patients were enrolled in the study.

References

[1] Cortés, J.P.; Lin, J.Z.; Marks, K.L.; Espinoza, V.M.; Ibarra, E.J.; Zañartu, M.; Hillman, R.E.; Mehta, D.D. Ambulatory Monitoring of Subglottal Pressure Estimated from Neck-Surface Vibration in Individuals with and without Voice Disorders. *Appl. Sci.* **2022**, *12*, 10692.

15th International Conference on Advances in
Quantitative Laryngology, Voice and Speech Research
March 30 – April 1, 2023
Mayo Clinic, Phoenix, Arizona

PHONATION ONSET IN AN IN-VIVO CANINE HEMI LARYNX MODEL: 3D MOVEMENT PATTERNS OF THE MEDIAL SURFACE

Patrick Schlegel, Hye Rhyn Chung, Michael Döllinger, Dinesh Chhetri

Keywords: hemilarynx, 3D reconstruction, in-vivo, dog, onset

Introduction

Phonation begins with the onset of vocal fold oscillation that is dependent on various factors such as airflow and different laryngeal muscle activations. Hence, onset is of great interest in voice research. However, it is often only assessed from a superior view, neglecting the vertical vibration component and the vibration onset along the medial surface.

Methods

We investigated vocal fold oscillation onset in an in-vivo canine hemi larynx model. Recurrent and superior laryngeal nerves were stimulated at average levels at nine different activation combinations approximately midway between minimum and maximum muscle response. Oscillation onset was achieved by providing a linearly rising airflow. The medial surface of the vocal fold was recorded through a glass prism at 3000 frames per second. Positions of 30 Landmarks tattooed on the medial surface were 3D reconstructed and interpolated as described in our previous work [1]. Empirical Eigenfunctions were used to de-noise and emphasize certain sections of vibrations.

Results

Vocal fold vibration started in general with a subtle near-periodic movement of the inferior part of the vocal fold. During this part of pre-onset the inferior part of the vocal fold was pushed further away from the glass prism with increasing airflow. Then the superior edge on the anterior side started to oscillate at a higher frequency. With further increase in airflow, the entire vocal fold started oscillating. This change happened gradually with first single complete oscillations moving from inferior to superior, breaking the pattern of fast anterior vibration. Then a stable state was reached as slower full vibrations replaced fast anterior vibrations.

Discussion and Conclusion

Looking at phonation onset along the medial surface reveals subtle patterns of movement preceding the from above visible vibration onset. This illustrates the importance of looking at vocal fold oscillations from different perspectives, also capturing the medial surface.

Acknowledgments

This project was funded by Grant No R01 DC11300 from the NIH/NIDCD. Contribution by M.D. are supported by Deutsche Forschungsgemeinschaft (DFG) under grant no. DO1247/16-1.

References

[1] Reddy NK, Schlegel P, Lee PY, Chhetri DK: 3D Reconstruction of Phonatory Glottal Shape and Volume: Effects of Neuromuscular Activation. Laryngoscope, 133(2):357-365, 2022

15th International Conference on Advances in
Quantitative Laryngology, Voice and Speech Research
March 30 – April 1, 2023
Mayo Clinic, Phoenix, Arizona

VIDEOSTROBOSCOPY VS HIGH-SPEED VIDEOENDOSCOPY; FACTORS INFLUENCING RATINGS OF LARYNGEAL OSCILLATION

ROBERT BRTINON FUJIKI, CAITLIN K. CROEGAERT-KOCH, & SUSAN L. THIBEAULT
University of Wisconsin-Madison, Department of Surgery

Keywords: Videostroboscopy, laryngeal imaging, high-speed videoendoscopy

Introduction: Characterizing vocal fold oscillation using videostroboscopy and/or high-speed videoendoscopy (HSV) is of critical importance for evaluating voice disorders.[1] It is unclear, however, what factors may influence the relationship between laryngeal oscillation ratings made from these exam types. This study examined the impact of patient diagnosis, severity of dysphonia, and rater experience on the relationship between videostroboscopic and HSV ratings of laryngeal oscillation and closure.

Methods: Ten licensed speech-language pathologists rated laryngeal oscillation using videostroboscopic and HSV exams from 15 patients with Adductor Spasmodic Dysphonia (ADSD) and 15 with benign vocal fold lesions. SLPs were categorized into lower (<5 years) and high (>5 years) experience groups. The Voice Vibratory Assessment of Laryngeal Imaging (VALI)[2] was adapted and used to rate vocal fold amplitude, mucosal wave, periodicity, phase symmetry, non-vibrating portion of the vocal fold, and glottal closure. Mean differences between ratings from both exam types were analyzed.

Results: Stroboscopy and HSV ratings were more strongly positively correlated for patients with benign vocal fold lesions (r between .43 and .75) than for those with ADSD (r between .40 and .68). Differences in ratings from videostroboscopy and HSV exams were significantly larger for ratings of amplitude, mucosal wave, glottic closure and periodicity in patients with ADSD than for patients with benign vocal fold lesions. For vocal fold amplitude and non-vibrating portion, raters with <5 years' experience displayed significantly greater differences between videostroboscopy and HSV ratings for patients with ADSD. More severe dysphonia was associated with significantly greater differences between ratings of periodicity and phase symmetry.

Discussion and Conclusion: Patient diagnosis, severity of dysphonia, and rater experience may impact the relationship between HSV and videostroboscopic exam ratings. Future study should expand this work to broader patient populations and rater professions.

Acknowledgments: We thank Nathan Welham, Carol Tolejano, Lisa Vinney, Alexandra Schenck, Felicia Francois, Renee King, Alex Foote, Maia Braden, Emerald Doll, and Anumitha Venkatraman for their roles in data collection.

References

1. Patel RR, Awan Shaheen N., Barkmeier-Kraemer Julie, 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. *American Journal of Speech-Language Pathology*. 2018;27(3):887-905. doi:10.1044/2018_AJSLP-17-0009

2. Poburka BJ, Patel RR, Bless DM. Voice-Vibratory Assessment With Laryngeal Imaging (VALI) Form: Reliability of Rating Stroboscopy and High-speed Videoendoscopy. *Journal of Voice*. 2017;31(4):513.e1-513.e14. doi:10.1016/j.jvoice.2016.12.003

15th International Conference on Advances in
Quantitative Laryngology, Voice and Speech Research
March 30 – April 1, 2023
Mayo Clinic, Phoenix, Arizona

A Novel Approach to High-Speed Transoral Fiberoptic Phonoscopy of the Glottis

Michal H. Tyc, Marcin Just, Anna Racino, Krzysztof Izdebski, Monika Morawska-Kochman

Keywords: High-Speed Digital Phonoscopy, fiberscope

Introduction

This study tested application of transoral flexible fiberscope to capture phonatory activity with a high-speed camera via oral route using a patent-pending tube. Transoral method was promising when compared to traditional methods, a rigid laryngoscope and the transnasal flexible fiberscopy

Methods

The study was conducted using High-Speed Digital Phonoscopy (HSDP) ALIS system consisting of an ALI Cam-HS1 camera module and ALI Lum-MF1 laser light source. Two rigid and one flexible endoscopes were used. All videos were recorded in HD mode and high-speed mode at 2400 fps (image 512x480 pixels) and 3200 fps (ca. 416x368 pixels). A single recording was about 1 second long. Dozens of recordings were obtained producing of different types of phonations.

Results

The quality of recordings using different types of endoscopes and different methods of insertion, transnasal and transoral differed. Compared with rigid optics, the overall visibility, especially the anterior commissure and glottis were significantly better for all vowels when viewed with a transoral flexible scope. The gag reflex was reduced and an almost-natural phonation was possible to record.

Discussion and Conclusion

The transoral examination with flexible fiberoptic endoscopy is comfortable for the patient and clinician. The visibility of the glottis was similar compared to a transnasal examination and the phonatory conditions for vowels were similar.

With the help of the laser light source, the image quality from the fiberscope was comparable with the image from older HSDP systems with rigid optics, and the transoral insertion option can be considered promising.

15th International Conference on Advances in
Quantitative Laryngology, Voice and Speech Research
March 30 – April 1, 2023
Mayo Clinic, Phoenix, Arizona

Comparing intraglottal geometry and velocity flow fields between an excised canine larynx and a synthetic vocal fold model

Jacob Michaud-Dorko[1], Charles Farbos de Luzan[2], Ephraim Gutmark[3], and Liran Oren[2]

Keywords: Synthetic vocal fold, excised canine larynx, flow separation vortices, and glottal flow waveform.

Introduction. Complementary tools for voice production research are emerging in the form of synthetic vocal folds. Made of silicone, these models replicate the geometry and material properties of human vocal folds and offer potential advantages over excised canine larynges, such as greater accessibility and longer lifespans with vibration frequencies similar to human phonation[1]. However, the synthetic models are structurally idealized, and their main limitation is that they do not represent human vocal folds as well as canine larynges. As the preferred model for studying the aerodynamics of phonation, canine larynges have been found to exhibit vortical structures in the flow fields during closing. Flow separation vortices (FSVs) near the superior edge are associated with changes in flow rate, leading to greater flow skewing and an increase in the maximum flow declination rate (MFDR), which enhance voice loudness and quality[2-5]. The current gap in knowledge is the comparison of the intraglottal geometry and velocity flow fields between an excised canine larynx and a synthetic vocal fold model.

Methods. Two-dimensional particle imaging velocimetry (PIV) was utilized to measure the intraglottal geometry and velocity at the mid-coronal plane in a synthetic vocal fold model and an excised canine larynx. The PIV measurements were used to determine the divergence angle, glottal width between the vocal folds, and the flow rate at the glottal exit.

Results. The findings indicate that synthetic models display only a slight divergence angle and no flow separation when the subglottal pressure is low. As the pressure rises, the increased divergence angle causes flow separation along the walls of the synthetic vocal folds and entrainment into the glottis. In contrast, the canine model exhibited the presence of FSVs along the superior edge at high subglottal pressure during the maximum divergence angle. As the subglottal pressure increased, the formation of FSVs resulted in greater glottal flow skewing towards closing and an increase in the MFDR.

Discussion and Conclusion. These findings highlight the limitations of synthetic vocal fold models in accurately representing the human larynx, particularly regarding the flow dynamics in the glottis. Further research is needed to improve the design and characterization of synthetic vocal fold models to better match the glottal flow fields to excised canine and human larynges.

Acknowledgments. None

References.

1. Murray, 2011. JoVE.
2. Oren, 2014. The Journal of the Acoustical Society of America.
3. Sundström, 2022. Journal of Voice.
4. Oren, 2014. Journal of Voice.
5. de Luzan, 2021. Journal of Voice.

15th International Conference on Advances in
Quantitative Laryngology, Voice and Speech Research
March 30 – April 1, 2023
Mayo Clinic, Phoenix, Arizona

INFLUENCE OF THE VOCAL FOLD KINEMATIC PARAMETERS ON THE KYMOGRAPHIC WAVEFORM

Svec JG[1], Pravin Kumar S[2], Vencálek O[3], Sandhanakrishnan R[2], Lehoux H[1]

[1]Voice Research Lab, Department of Experimental Physics, Faculty of Science, Palacky University, Olomouc, Czechia.
[2]Department of Biomedical Engineering, Sri Sivasubramaniya Nadar College of Engineering, Chennai, India.
[3]Department of Mathematical Analysis and Applications of Mathematics, Faculty of Science, Palacky University, Olomouc, Czechia.
Contact: jan.svec@upol.cz

Keywords: high-speed videolaryngoscopy, kymography, kinematic model, vocal fold vibration

Introduction: High-speed laryngoscopic and kymographic imaging has shown a great potential for "functional voice diagnosis". However, wider clinical implementation of this technology requires deeper understanding of how the observed vibration parameters are shaped by the underlying vocal fold dynamics.

Methods: To advance the understanding, we used a previously developed mucosal-wave-based kinematic model of the vocal fold oscillations [1] and simulated kymograms resembling those found in clinical practice. We varied five model input parameters within a clinically relevant range: amplitude of the vocal fold's upper margin (0.1 to 1.1 mm), amplitude of the vocal fold's lower margin (0.1 to 1.1 mm), vertical phase differences between the upper and lower vocal fold margins (0° to 125°), glottal halfwidth (-0.05 to 1.2 mm) and glottal convergence angle (-15° to 35°). We studied how these parameters influenced the following kymographic parameters: vocal fold opening speed, vocal fold closing speed, open quotient and peak-to-peak amplitude.

Results: We found that: 1) vocal fold opening speed was primarily influenced by the upper margin amplitude; 2) vocal fold closing speed was primarily influenced by the lower margin amplitude; 3) open quotient was influenced primarily by glottal half-width and secondarily by vertical phase differences; and 4) peak-to-peak kymographic amplitude was substantially influenced by all five input parameters.

Discussion and Conclusion

These insights can improve interpretation of the findings from high-speed and kymographic recordings in healthy and disordered voices.

Acknowledgments

Research was supported by the projects TH04010422 "VKG. 3.0", and OPIE project "Improvement of Doctoral Student Grant Competition Schemes and their Pilot Implementation", reg. no. CZ.02.2.69/0.0/0.0/19_073/0016713.

References

[1] Kumar, S. P., & Svec, J. G. (2019). Kinematic model for simulating mucosal wave phenomena on vocal folds. Biomedical Signal Processing and Control, 49(March), 328-337.

15th International Conference on Advances in
Quantitative Laryngology, Voice and Speech Research
March 30 – April 1, 2023
Mayo Clinic, Phoenix, Arizona

Voice efficiency for different voice qualities combining experimentally derived sound signals and numerical modeling of vocal tract models

Louisa Traser[1], Stefanie Rummel[2], Fiona Stritt[1], Johannes Fischer[3], Michael Bock[3], Matthias Echternach[4], Bernhard Richter[1], Mario Fleischer[5]

1. Institute of Musicians' Medicine, Medical Center University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany; 2. Institut Rummel, Frankfurt, Germany; 3. Dept. of Radiology, Medical Physics, Medical Center University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany; 4. Ludwig-Maximilians-University, Division of Phoniatrics and Pediatric Audiology, Department of Otorhinolaryngology, LMU, Munich, Germany; 5. Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Department of Audiology and Phoniatrics, Berlin, Germany

Keywords: voice efficiency, voice qualities, Estill Voice Training®, MRI, 3D vocal tract model, Finite-Element-Model, electroglottography

Introduction

From western classical singing to contemporary commercial music, only limited data about the voice efficiency of different singing styles are available.

Methods

In this study, we quantified the acoustic sound intensity within the human glottis for different vocal tract configurations and vocal fold vibration.

Combining Finite-Element-Models derived from 3D-MRI data, audio recordings, and electroglottography (EGG), we analyzed vocal tract transfer functions, particle velocity and acoustic pressure at the glottis, and EGG parameters to evaluate voice efficiency at the glottal level and resonance characteristics of different voice qualities according to Estill Voice Training®.

Results

Voice qualities Opera and Belting represent highly efficient vocal strategies for orchestral sounds. Twang and Belting use similar vowel qualities, but the Twang vocal tract configuration was associated with anti-resonances and a reduced vocal fold contact, but a comparable energy transfer from the glottis to the vocal tract. Speech has a highly efficient energy transfer from the glottis to the vocal tract. However, the absence of perceptual relevant energy portions in the higher frequency range makes it more susceptible to noise interference. Falsetto and Sobbing are less efficient. For Falsetto, it is mainly due to its voice source characteristics and for Sobbing due to energy loss in the vocal tract. Thus, sound amplification might be appropriate here.

Discussion and Conclusion

Differences exist between voice qualities regarding the sound intensity. Causes are seem by different vocal tract morphologies and oscillation characteristics of the vocal folds.

15th International Conference on Advances in
Quantitative Laryngology, Voice and Speech Research
March 30 – April 1, 2023
Mayo Clinic, Phoenix, Arizona

Contribution of resident commensal and pathogenic microbial communities on vocal fold mucosal integrity and function

VLASTA LUNGOVA, MADHU GOUDA, SUSAN L. THIBEAULT

Keywords: Vocal folds, benign lesions, epithelial cell junctions, pathogenic bacteria, commensals.

Introduction

Existing dilated intercellular spaces in the vocal fold (VF) epithelium were documented in VF benign lesions [1]. Mechanisms regarding the disruption of cell junctions are limited and do not address the contribution of resident microbial communities to this pathological phenomenon. Recent evidence suggests that the etiology and progression of VF benign diseases are likely associated with *Streptococcus pseudopneumoniae (SP)*, as a dominant bacterial pathogenic species present in every lesion at the expense of commensals, namely *Streptococcus salivarius (SS)* [2]. *SP* colonizes the host via adhesins and shedases, such as HtrA1 [3], that cleaves cell adhesion molecule E-cadherin [4]. In contrast, commensals have the probiotic effect, inhibit the action of pathogens, and stimulate the host's immune responses [5]. The main aim of this study was to determine the contribution of resident commensal and pathogenic bacteria on VF mucosal integrity and function. We hypothesize that the VF epithelial barrier will be compromised with exposure to pathogenic bacteria. These effects can be reversed in the presence of laryngeal commensals.

Methods

Engineered human VF mucosae were inoculated with 100µl of bacterial suspensions of (1) *SP* only, (2) *SS* only, (3) *SP and SS* co-cultures, and (4) plain culture medium for controls for 24 and 48 hours (h). We evaluated epithelial bacterial colonization using a Fluorescent In Situ Hybridization 16S probe, expression of HtrA1, E-cadherin, and inflammatory cytokines.

Results

Our data show that *SP* likely contributes to the pathological dilation of intercellular spaces via HtrA-mediated E-cadherin cleavage particularly detected 48h post-inoculation. *SS* reduces epithelial damage by preventing *SP* from binding to the epithelial surface and by stimulating IL6 and IL8 expression in VF mucosal cells.

Discussion and Conclusion

These findings extend our understanding of how pathogenic bacteria contribute to the VF benign disease etiology and progression and how probiotics may neutralize the effect of pathogens and/or prevent infection.

Acknowledgments

This work was funded the grants NIH-NIDCD R01 04336 and R01 012773.

References

1. Mostafa, B., Fawaz, S., Taha, M. & Salman, M. Ultrastructural Analysis of the Junctional Zone of Vocal Fold Polyps and Nodules. *J. Otolaryngol. Res.* **6**, (2017).
2. Hanshew, A. S., Jetté, M. E. & Thibeault, S. L. Characterization and comparison of bacterial communities in benign vocal fold lesions. *Microbiome* **2**, 1–9 (2014).
3. Garriss, G. *et al.* Genomic Characterization of the Emerging Pathogen Streptococcus pseudopneumoniae. *Am. Soc. Microbiol. mBio* **10**, e01286-19 (2019).
4. Bernegger, S. *et al.* E-Cadherin Orthologues as Substrates for the Serine Protease High Temperature Requirement A (HtrA). *Biomolecules* **12**, 1–13 (2022).
5. Manning, J. *et al.* Investigation of Streptococcus salivarius-mediated inhibition of pneumococcal adherence to pharyngeal epithelial cells. *BMC Microbiol.* **16**, 1–9 (2016).

15th International Conference on Advances in
Quantitative Laryngology, Voice and Speech Research
March 30 – April 1, 2023
Mayo Clinic, Phoenix, Arizona

Piezo1-expressing vocal fold epithelia modulate acute remodeling following Naphthalene injury

Foote AG, Lungova V, & Thibeault SL

Keywords: Epithelial mechanoreceptors; Mechanobiology; Regeneration; Vocal folds; Wound healing.

Introduction Mechanoreceptors are implicated as functional afferents within mucosa of the airways and the recent discovery of mechanosensitive channels *Piezo1* and *Piezo2* has proved essential for cells of various mechanically sensitive tissues.[1-3] However, the role for *Piezo1/2* in vocal fold (VF) mucosal epithelia, a cell that withstands excessive biomechanical insult, remains unknown. The purpose of this study was to test the hypothesis that *Piezo1* is required for VF mucosal repair pathways of epithelial cell injury.

Methods Utilizing a sonic hedgehog (shh) Cre line for epithelial-specific ablation of *Piezo1/2* mechanoreceptors, we investigated 6wk adult VF mucosa following naphthalene exposure for repair strategies at 1, 3, 7 and 14 days post-injury (dpi). Acute remodeling was assessed via histology, immunofluorescence and qPCR.

Results PIEZO1 localized to differentiated apical epithelia and was paramount for epithelial remodeling events. Injury to wildtype epithelium was most appreciated at 3 dpi. $Shh^{cre/+}$; $Piezo1^{loxP/loxP}$,$Piezo2^{loxP/+}$ mutant epithelium exhibited severe cell/nuclear defects compared to injured controls. Specifically, *Piezo1* was shown to modulate self-renewal via repressive effects on p63 expression and YAP subcellular localization. Conditional ablation of *Piezo1* and/or *Piezo2* to uninjured VF epithelium did not result in abnormal phenotypes across P0, P15 and 6wk postnatal stages compared to heterozygote and control tissue.

Discussion and Conclusion Results demonstrate a role for *Piezo1*-expressing VF epithelia in regulating self-renewal via effects on p63 transcription and YAP subcellular translocation-altering cytokeratin differentiation. Results from this investigation improve our understanding of acute VF wound healing in the context of *Piezo1* epithelial function and form the basis for an *in vivo* methodology to study VF responses to NAPH injury.

Acknowledgments We gratefully acknowledge Sierra Raglin for her expert assistance with tissue sample preparation, David Yang for help with histopathology, Julianna Tibbetts for assistance with genotyping and Xin Sun for helpful discussions.

References

1. Andreatta, R.D., et al., Mucosal afferents mediate laryngeal adductor responses in the cat. *J Appl Physiol* (1985), 2002. 93(5): p. 1622-9.
2. Coste, B., et al., Piezo1 and Piezo2 are essential components of distinct mechanically activated cation channels. *Science*, 2010. **330**(6000): p. 55-60.
3. Sinclair, C.F., M.J. Tellez, and S. Ulkatan, Human laryngeal sensory receptor mapping illuminates the mechanisms of laryngeal adductor reflex control. *Laryngoscope*, 2018. 128(11): p. E365-E370.

15th International Conference on Advances in
Quantitative Laryngology, Voice and Speech Research
March 30 – April 1, 2023
Mayo Clinic, Phoenix, Arizona

Shape design and functional evaluation scaffolds for reconstruction of partial laryngectomy defects

Juergen Neubauer, Michael McPhail, Stephanie Cotton Zacharias, David Lott

Keywords: Excised larynx, Generalized hemilarynx setup, Partial laryngectomy reconstruction, Optimizing laryngeal implant, Scaffold shape design, CT based design

Introduction

This presentation will discuss the innovative use of excised larynges in designing and evaluating scaffold shapes for partial laryngectomy reconstruction. To optimize all laryngeal functions for partially reconstructed larynges, breathing, swallowing, and phonation, we need to test the efficacy of different scaffold shape designs in a controlled and standardized environment.

Methods

We examined CT scans of excised larynges from dogs and pigs to investigate the relationship between laryngeal cartilage positioning and glottis configuration. Custom scaffold shapes were designed using this information and 3D-printed models were created. We performed partial laryngectomies on the excised larynges and inserted the scaffold shapes into the defects. To evaluate the efficacy of different scaffold designs, we conducted assessments of voice, breathing, and swallowing functions using a custom setup.

Results

We demonstrate how laryngeal cartilage positioning and glottis configuration can imitate laryngeal posturing for breathing, swallowing, and phonation. We present data on acoustic, high-speed imaging, pressure, and air flow to evaluate different scaffold shape designs for breathing, swallowing, and phonation.

Discussion and Conclusion

We show that we can explore the extremes of the interplay between cartilage posturing and glottis configuration to evaluate the functional performance of arbitrary 3D scaffold shapes in excised larynges. To optimize all laryngeal functions, breathing, swallowing, and phonation, a 3D implant needs to balance the competing and opposing needs of breathing, ideally having a wide-open glottis, swallowing, preventing aspiration with a tightly closed glottis, and normal phonation, requiring an intermediate glottal adduction and minimum glottal leakage for ease of phonation and sufficient loudness. We can find a compromising 3D implant shape by measuring the laryngeal functions in excised larynx setups.

Our approach extends the classical hemilarynx setup, which has traditionally been used with only a flat shape to study mechanisms of phonation, to a generalized setup for studying the interplay between cartilage posturing, glottis configuration and 3D implants for partial laryngectomy reconstruction. We demonstrate that this approach can predict and optimize 3D implant shapes for partial laryngectomy reconstruction by balancing the needs of breathing, swallowing, and phonation.

Acknowledgments

This work was funded by NIH grant 1R01DC019114.

References

15th International Conference on Advances in
Quantitative Laryngology, Voice and Speech Research
March 30 – April 1, 2023
Mayo Clinic, Phoenix, Arizona

FIBROUS BIOMIMETIC MEMBRANES FOR ENHANCED WOUND HEALING

Dina Gadalla, Ph.D.[1], Maeve Kennedy[1], Jamie Ganem[1], Mustafa Suppah, M.D.[1], David G. Lott, M.D.[1,2]

[1]Head and Neck Regenerative Medicine Laboratory
[2]Division of Laryngology
Mayo Clinic Arizona, Phoenix, AZ 85054, USA

Keywords: Murine wound model, angiogenesis, epithelialization, electrospun fibers

Introduction

Due to the dynamic environment of the airway system, methods to restore respiratory epithelium during airway reconstruction are not suitable for large defect areas. A lack of epithelium can lead to stenosis, collapse and infection that restrict airway patency. Consequently, engineered tissue capable of regenerating a functional epithelium is required. This can be achieved by providing basement membrane biomimetic surfaces favorable to the migration and polarization of epithelial cells. To this end, we envision basement membrane fiber surfaces – incorporated to the luminal side of tracheal replacements – could recruit host epithelial cells and stimulate epithelium formation at the replacements-environment interface.

Methods

We fabricated thin fiber layers that could adsorb a protein mixture resembling basement membrane properties. Fibers were prepared by electrospinning, and a protein mixture was adsorbed to the surface. Following *in vitro* migration, proliferation and differentiation studies, the response to the implant was studied in an *in vivo* murine model of excisional wound healing. Here, implants were placed after creating skin punches in the backs of mice. Wound photos were taken and the wound areas were measured. Additionally, immunofluorescent stainings to detect the presence of angiogenic and epithelial markers were carried out after 3 weeks.

Results & Discussion

Migration studies showed an increase in cell numbers migrating to fiber layers with proteins. This is vital since native basement membranes mainly act as substrates for the adhesion and migration of epithelial cells. Cell differentiation experiments showed that fibers enhanced the expression of several mucous secreting and ciliated cell genes. Wound areas decreased over the study time-period and an epithelium layer was regenerated over the surface of the implants. Furthermore, staining studies indicated the presence of CD31, α-SMA, pan-cytokeratin and E-cadherin in the wound area further confirming healing and wound area closure.

Discussion and Conclusion

The presentation of basement membrane proteins using fiber layers increases cell adherence and cell migration to fibers, stimulates the differentiation of basal cells into ciliated cells and promotes epithelialization in an *in vivo* dermal wound model by mimicking the composition of native membranes. Thus, the incorporation of such protein adsorbed fibers on the luminal side of tracheal replacements could ultimately facilitate the regeneration of epithelial tissue and is the current next step we are exploring.

Acknowledgments

We would like to thank Cathy Madsen, Cheryl Myers, Ph.D., and Dominique Hoelzinger, Ph.D. for their help with FACS staining and cytometry flow that were carried out to confirm the safety of the implants *in vivo*.

References

1. D. Gadalla, Y. Tchoukalova and D. G. Lott, “Regenerating Airway Epithelium using Fibrous Biomimetic Basement Membranes”, Journal of Biomedical Materials Research, vol. 110, no. 6, pp. 1251-1262, 2022.
2. R. D. Galiano, J. Michaels, M. Dobryansky, J. Levine and G. C. Gurtner, “Quantitative and reproducible murine model of excisional wound healing”, Wound Repair and Regeneration, vol. 12, pp. 485-492, 2004.